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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
05 March 2020 (05.03.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/047124 A1

(51) International Patent Classification:
*C12N 15/10* (2006.01) *C12N 9/12* (2006.01)

(21) International Application Number:
PCT/US2019/048607

(22) International Filing Date:
28 August 2019 (28.08.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 62/723,886 | 28 August 2018 (28.08.2018) | US |
| 62/725,778 | 31 August 2018 (31.08.2018) | US |
| 62/850,883 | 21 May 2019 (21.05.2019) | US |
| 62/864,924 | 21 June 2019 (21.06.2019) | US |


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

(72) Inventors: **RUBENS, Jacob, Rosenblum**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **VON MALTZAHN, Geoffrey, A.**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **CITORIK, Robert, James**; 55 Cambridge Parkway, 8th Floor, Sutie 800E, Cambridge, MA 02142 (US). **STEINBERG, Barrett, Ethan**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **LI, Donghui**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambaridge, MA 02142 (US).

(74) Agent: **KOYFMAN, Hannah, R.** et al.; Lando & Anastasi, LLP, Riverfront Office Park, One Main Street, Suite 1100, Cambridge, MA 02142 (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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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,

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

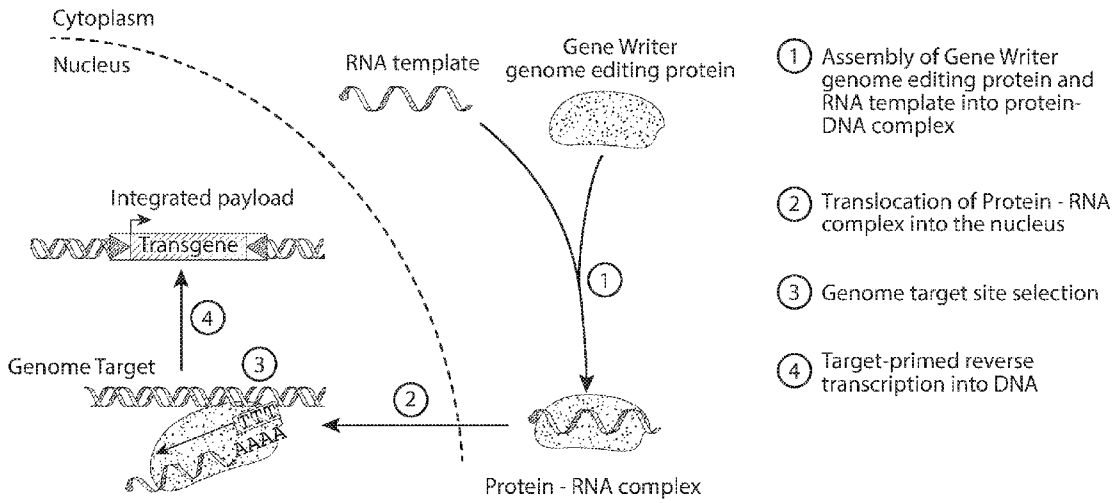



Figure 1

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

WO 2020/047124 A1

[Continued on next page]

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).

**Declarations under Rule 4.17:**

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

**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))*

— *with sequence listing part of description (Rule 5.2(a))*

# METHODS AND COMPOSITIONS FOR MODULATING A GENOME

This application claims priority to U.S. Serial No.: 62/723,886 filed August 28, 2018, U.S. Serial No.: 62/725,778 filed August 31, 2018, U.S. Serial No.: 62/850,883 filed May 21, 2019, and U.S. Serial No.: 62/864,924 filed June 21, 2019, the entire contents of each of which is incorporated herein by reference.

## SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on August 28, 2019, is named V2065-7000WO_SL.txt and is 4,004,548 bytes in size.

## 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 that encodes a therapeutic polypeptide or that encodes a mammalian (e.g., human) polypeptide, or a fragment or variant thereof.

2. 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 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.
vii. the heterologous object sequence is other than a T cell chimeric antigen receptor

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

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

4. 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 and (iii) 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.

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, wherein one or both of (i) or (ii) are derived from an avian retrotransposase, e.g., have a sequence of Table 2 or 3 or at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; 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.

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, 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; 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.

7. The system of embodiment 6, 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.

8. The system of embodiment 6, 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.

9. The system of embodiment 6, 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.

10. 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.

11. 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 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.

12. 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 5' untranslated sequence that binds the polypeptide, (ii) a 3' untranslated sequence that binds the polypeptide, (iii) a heterologous object sequence, and (iv) 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.

13. 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 5' untranslated sequence that binds the polypeptide, (ii) a 3' untranslated sequence that binds the polypeptide, and (iii) a heterologous object sequence, 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.

14. 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 at least one of (i) or (ii) is heterologous, 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.

15. 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, (i) a reverse transcriptase domain and (iii) an endonuclease domain, wherein at least one of (i), (ii) or (iii) is heterologous, 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.

16. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) 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 (ii) 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; 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.

17. A system for modifying DNA comprising:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) 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, (ii) 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, and (iii) a heterologous target DNA binding domain (e.g., a heterologous zinc-finger DNA binding 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.

18. The system of any of the preceding embodiments, wherein the template RNA comprises (iii) a promoter operably linked to the heterologous object sequence.

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

20. The system of embodiment 17, 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.

21. The system 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.

22. The system 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.

23. The system of embodiment 22, wherein the fusion nucleic acid comprises RNA.

24. The system of embodiment 22, wherein the fusion nucleic acid comprises DNA.

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

26. The system of embodiment 25, wherein the template RNA further comprises a nuclear localization signal.

27. The system of any of the preceding embodiments, wherein (a) comprises RNA encoding the polypeptide.

28. The system of embodiment 27, wherein the RNA of (a) and the RNA of (b) are separate RNA molecules.

29. The system of embodiment 28, wherein the RNA of (a) and the RNA of (b) are present at a ratio of between 10:1 and 5:1, 5:1 and 2:1, 2:1 and 1:1, 1:1 and 1:2, 1:2 and 1:5, or 1:5 and 1:10.

30. The system of embodiment 28, wherein the RNA of (a) does not comprise a nuclear localization signal.

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

32. The system 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.

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

34. The system of embodiment 33, wherein the RNA comprises a stem-loop sequence or a helix, 5' of the pseudoknot sequence.

35. The system of embodiment 33 or 34, 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.

36. The system of any of embodiments 33-35, wherein the template RNA comprising the pseudoknot has catalytic activity, e.g., RNA-cleaving activity, e.g, cis-RNA-cleaving activity.

37. The system 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.

38. Any above-numbered system, 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, or a reverse transcriptase domain or endonuclease domain thereof.

39. Any above-numbered system, 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 2-3 or a reverse transcriptase domain, endonuclease domain, or DNA binding domain thereof.

40. Any above-numbered system, 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.

41. Any above-numbered system, 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.

42. The system of embodiment 41, 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.

43. Any above-numbered system, wherein (a) comprises RNA and (b) comprises RNA.

44. Any above-numbered system, 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.

45. Any above-numbered system, which does not comprise DNA, or which does not comprise more than 10%, 5%, 4%, 3%, 2%, or 1% DNA by mass or by molar amount.

46. 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).

47. 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).

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

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

50. 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.

51. The system of embodiment 50, 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

52. 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 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 10 bases of 100% identity to a target DNA strand, 5' end of the template RNA.

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

54. The system of any of embodiments 1-52, wherein (a) and (b) are separate nucleic acids.

55. The system of any of the preceding embodiments, wherein the template RNA comprises at least 10 bases of 100% identity to a target DNA strand (e.g., wherein the target DNA strand is a human DNA sequence), at the 5' end of the template RNA.

56. The system of any of the preceding embodiments, wherein the template RNA comprises at least 10 bases of 100% identity to a target DNA strand (e.g., wherein the target DNA strand is a human DNA sequence), at the 3' end of the template RNA.

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

58. 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.

59. The method of embodiment 58, wherein the cell, tissue or subject is a mammalian (e.g., human) cell, tissue or subject.

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

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

62. The method of any of the preceeding embodimens, where in the cell is not immortalized.

63. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain, (ii) an endonuclease domain, and optionally (iii) a DNA-binding 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.

64. The method of embodiment 63, wherein the polypeptide does not comprise a target DNA binding domain.

65. The method of embodiment 63, wherein the polypeptide is derived from an APE-type transposon reverse transcriptase.

66. The method of embodiment 63, wherein the (i) a reverse transcriptase domain (ii) an endonuclease domain, or both of (i) and (ii), have a sequence of Table 1 or a sequence having at least 80%, 85%, 90%, 95%, 97%, 98%, 99%, 100% identity thereto.

67. The method of embodiment 63, wherein the polypeptide further comprises a target DNA binding domain.

68. A method of modifying the genome of a mammalian cell, comprising contacting the cell with:

(a) an RNA encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain, (ii) an endonuclease domain, and optionally (iii) a DNA-binding domain; and

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

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more 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.

69. The method of embodiment 68, 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 exogenous DNA sequence to the genome of the mammalian cell.

70. The method of embodiment 68 or 69, which results in the addition of a protein coding sequence to the genome of the mammalian cell.

71. A method of inserting DNA into the genome of a mammalian cell, comprising contacting the cell with an RNA composition, wherein the RNA composition comprises:

(a) a first RNA that directs insertion of a template RNA into the genome, and

(b) a template RNA comprising a heterologous sequence,

wherein the method does not comprise contacting the mammalian cell with DNA, or wherein the compositions of (a) and (b) do not comprise more 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,

wherein the method 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 (e.g., exogenous DNA) sequence to the genome of the mammalian cell.

72. The method of embodiment 71, wherein the first RNA encodes a polypeptide (e.g., a polypeptide of any of Tables 1, 2, or 3 herein), wherein the polypeptide directs insertion of the template RNA into the genome.

73. The method of embodiments 72, wherein the template RNA further comprises a sequence that binds the polypeptide.

74. 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.

75. 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, 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.

76. 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 delivering only RNA to the mammalian cell.

77. 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 delivering RNA and protein to the mammalian cell.

78. The method of any one of embodiments 68-77, wherein the template RNA serves as the template for insertion of the exogenous DNA.

79. The method of any one of embodiments 68-78, which does not comprise DNA-dependent RNA polymerization of exogenous DNA.

80. The method of any of embodiments 58-79, 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 mammalian cell.

81. The methods of any of embodiments 68-80, wherein the RNA of (a) and the RNA of (b) are covalently linked, e.g., are part of the same transcript.

82. The methods of any of embodiments 68-80, wherein the RNA of (a) and the RNA of (b) are separate RNAs.

83. The method of any of embodiments 58-82, which does not comprise contacting the mammalian cell with a template DNA.

84. A method of modifying the genome of a human cell, comprising contacting the cell with:

(a) a polypeptide or a nucleic acid encoding a polypeptide, wherein the polypeptide comprises (i) a reverse transcriptase domain, (ii) an endonuclease domain, and optionally (iii) a DNA-binding 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 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.

85. A method of adding an exogenous coding region to the genome of a cell (e.g., a mammalian cell), comprising contacting the cell with an RNA comprising the non-coding strand of the exogenous coding region, wherein optionally the RNA does not comprsise a coding strand of the exogenous coding region, wherein optionally the delivery comprises non-viral delivery.

86. A method of expressing a polypeptide in a cell (e.g., a mammalian cell), comprising comprising contacting the cell with an RNA, wherein the RNA comprises a non-coding strand that is the reverse complement of a sequence that would encoding the polypeptide, wherein optionally the RNA does not comprsise a coding strand encoding the polypeptide, wherein optionally the delivery comprises non-viral delivery.

87. The method of any of embodiments 58-86, wherein the sequence that is inserted into the mammalian genome is a sequence that is exogenous to the mammalian genome.

88. The method of any of embodiments 58-87, which operates independently of a DNA template.

89. The method of any of embodiments 58-88, wherein the cell is part of a tissue.

90. The method of any of embodiments 58-89, 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.

91. The method of any of embodiments 58-90, wherein the contacting comprises contacting the cell with a plasmid, virus, viral-like particle, virosome, liposome, vesicle, exosome, or lipid nanoparticle.

92. The method of any of embodiments 58-91, wherein the contacting comprises using non-viral delivery.

93. The method of any of embodiments 58-92, 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.

94. The method of any of embodiments 58-93, 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 comprsise a coding strand encoding the polypeptide, wherein optionally the delivery comprises non-viral delivery, thereby expressing the polypeptide in the cell.

95. The method of any of embodiments 63-94, wherein the contacting comprises administering (a) and (b) to a subject, e.g., intravenously.

96. The method of any of embodiments 63-95, wherein the contacting comprises administering a dose of (a) and (b) to a subject at least twice.

97. The method of any of embodiments 63-96, wherein the polypeptide reverse transcribes the template RNA sequence into the target DNA strand, thereby modifying the target DNA strand.

98. The method of any embodiments 63-97, wherein (a) and (b) are administered separately.

99. The method of any of embodiments 63-97, wherein (a) and (b) are administered together.

100. The method of any of embodiments 63-99, wherein the nucleic acid of (a) is not integrated into the genome of the host cell.

101. 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.

102. 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.

103. 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.

104. 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.

105. 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.

106. The method of embodiment 104 or 105, 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.

107. 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.

108. The method of embodiment 107, 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.

109. 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.

110. 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.

111. 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.

112. 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.

113. 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).

114. 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.

115. 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.

116. 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.

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

118. 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.

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

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

121. 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.

122. 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.

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

124. 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.

125. 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.

126. 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, or by the level of p53 phosphorylated at Ser15 and Ser20.

127. 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.

128. 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.

129. 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.

130. 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.

131. 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).

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

133. 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).

134. 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.

135. The method of embodiment 134, 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.

136. The method of embodiment 134 or 135, which further comprises contacting the template RNA with a cell.

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

138. 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.

139. 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.

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

141. 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.

142. 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.

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

144. 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.

145. 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 2 or 3 or at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

146. 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.

147. 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.

148. 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, or does not comprise a reverse transcriptase sequence.

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

150. 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.

151. 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.

152. 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.

153. 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.

154. 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.

155. 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.

156. A substantially pure polypeptide comprising (a) a reverse transcriptase domain and (b) a heterologous endonuclease domain.

157. A substantially pure polypeptide comprising (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.

158. A substantially pure polypeptide comprising (a) a reverse transcriptase domain, (b) an endonuclease domain, and (c) a heterologous target DNA binding domain.

159. A polypeptide or a nucleic acid encoding the polypeptide, wherein the polypeptide comprises (a) a reverse transcriptase domain and (b) an endonuclease domain, wherein at least one of (a) or (b) is heterologous to the other.

160. A polypeptide or a nucleic acid encoding the polypeptide, 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 to the other.

161. Any polypeptide of numbered embodiments 156-160, wherein the reverse transcriptase domain has at least 80% identity (e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity) to a reverse transcriptase domain of an APE-type or RLE-type non-LTR retrotransposon listed in any of Tables 1-3.

162. Any polypeptide of numbered embodiments 156-161, wherein the endonuclease domain has at least 80% identity e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity, to a endonuclease domain of an APE-type or RLE-type non-LTR retrotransposon listed in any of Tables 1-3.

163. Any polypeptide of numbered embodiments 156-162 or any preceding numbered method, wherein the DNA binding domain has at least 80% identity e.g., at least 85%, 90%, 95%, 97%, 98%, 99%, 100% identity, to a DNA binding domain of a sequence listed in Table 1, 2, or 3.

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

165. A vector comprising the nucleic acid of numbered embodiment 164.

166. A host cell comprising the nucleic acid of numbered embodiment 164.

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

168. A host cell comprising the vector of numbered embodiment 165.

169. A host cell (e.g., a human cell) comprising: (i) a heterologous object sequence (e.g., a sequence encoding a therapeutic polypeptide) at a target site in a chromosome, and (ii) one or both of an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of column 6 of Table 3) on one side (e.g., upstream) of the heterologous object sequence, and an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of column 7 of Table 3) on the other side (e.g., downstream) of the heterologous object sequence.

170. A host cell (e.g., a human cell) comprising: (i) a heterologous object sequence (e.g., a sequence encoding a therapeutic polypeptide) at a target site in a chromosome, wherein the target locus is a Natural Harbor™ site, e.g., a site of Table 4 herein.

171. The host cell of embodiment 170, which further comprises (ii) one or both of an untranslated region 5' of the heterologous object sequence, and an untranslated region 3' of the heterologous object sequence.

172. The host cell of embodiment 170, which further comprises (ii) one or both of an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of column 6 of Table 3) on one side (e.g., upstream) of the heterologous object sequence, and an untranslated region (e.g., a retrotransposon untranslated sequence, e.g., a sequence of column 7 of Table 3) on the other side (e.g., downstream) of the heterologous object sequence.

173. The host cell of any of embodiments 169-173, which comprises heterologous object sequence at only the target site.

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

175. The pharmaceutical composition of embodiment 174, 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.

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

177. 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, thereby modifying the target DNA strand.

178. 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 an amino acid sequence listed in Table 5 (e.g., any one of SEQ ID NOs: 1017-1022), or a functional fragment thereof.

179. 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 reverse transcriptase domain of an amino acid sequence listed in Table 5 (e.g., any one of SEQ ID NOs: 1017-1022), or a functional fragment thereof.

180. 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 an amino acid sequence listed in Table 5 (e.g., any one of SEQ ID NOs: 1017-1022), or a functional fragment thereof.

181. 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).

182. 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).

183. 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).

184. 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).

185. 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.

186. Numbered embodiment 185, 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 (e.g., as shown in any of Figures 17A-17F).

187. Numbered embodiment 185 or 186, 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.

188. Numbered embodiment 185 or 186, 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.

189. Numbered embodiment 185 or 186, 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.

190. Any one of numbered embodiments 185-189, 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).

191. 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.

192. 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.

193. 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.

194. 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.

195. 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.

196. 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).

197. 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, e.g., R2-1_GFo), or a functional fragment thereof.

198. 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, e.g., R2-1_GFo), or a functional fragment thereof.

199. 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, e.g., R2-1_GFo), or a functional fragment thereof.

200. Any one of numbered embodiments 197-199, 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.

201. 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 polypeptidefrom a large roundworm, e.g., *Ascaris lumbricoides* (e.g., as described herein, e.g., R4_AL), or a functional fragment thereof.

202. 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, e.g., R4_AL), or a functional fragment thereof.

203. 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 polypeptidefrom a large roundworm, e.g., *Ascaris lumbricoides* (e.g., as described herein, e.g., R4_AL), or a functional fragment thereof.

204. Any one of numbered embodiments 201-203, 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.

205. 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.

206. 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.

207. Numbered embodiment 205 or 206, wherein the p53 protein level is determined according to the method described in Example 30.

208. 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.

209. 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.

210. Numbered embodiment 205 or 209, wherein the p21 protein level is determined according to the method described in Example 30.

211. 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.

212. 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.

213. 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.

214. 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.

215. 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.

*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.

**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. 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.

**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.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic of the Gene Writing genome editing system.

Figure 2 is a schematic of the structure of the Gene Writer genome editor polypeptide.

Figure 3 is a schematic of a Gene Writer genome editor polypeptide comprising a heterologous DNA binding domain designed to target different sites of the genome.

Figure 4 is a schematic of the structure of Gene Writer genome editor template RNA.

Figure 5 is a schematic showing the Gene Writing genome editing system to add a gene expression unit into a safe harbor site in the genome.

Figure 6 is a schematic showing Gene Writing genome editing to add a new exon into an specific intron in the genome and replace downstream exons.

Figure 7 illustrates a schematic of Miseq library construction. Nested PCR was performed across the R2Tg-rDNA junction using (1) outer forward primer and tailed inner reverse primer followed by (2) tailed inner forward primer and tail reverse primer. The inner reverse primer contains a 1-4 base stagger, an 8-nucleotide randomized UMI, and a multiplexing barcode. The UMI allows for counting of individual amplification events to eliminate PCR bias.

Figures 8A-8B: Results of Miseq and Matlab analysis of DNA-mediated R2Tg integration into Hek293T cells. Each graph shows analysis of (Fig. 8A) the experimental R2Tg (Fig. 8B) and 1bp deletion negative control. The y-axis indicates aligned counts of unique sequences determined via unique UMIs found via Matlab. The X-axis indicates the sequence position of sequence coverage. The vertical gray line at the left of the graph indicates the position of the forward primer, while the vertical gray line at the right of the graph indicates the expected Tg-rDNA junction site. Bars at the right end of the graph indicate insertion without a truncation, and bars at the left end of the graph indicate truncation. Fig 8A shows that most sequences show high alignment to the expected integration product.

Figure 9 shows a ddPCR evaluation of copy number variation of the R2Tg-rDNA junction in human cells across transfection conditions. Forward primer and probe were expected to bind to the 3' UTR of the R2Tg, while reverse primer was targeted to the human rDNA. The resulting ddPCR signal was normalized to that of reference assay RPP30 to determine copy number. Significantly higher average copies per genome were found with the wildtype (WT, left set of bars) R2Tg as compared to genetic control altering translation with a 1-bp deletion (Frameshift mutant control, right set of bars).

Figure 10 illustrates the sequence alignment and coverage of TOPO cloning the nested PCR product from Example 7. The gray line at the right edge of the graph indicates the expected

transgene-rDNA junction. Most sequences showed high alignment to the expected integrated product.

Figure 11 is a schematic of an exemplary template RNA. It comprises a payload domain in the center (e.g., a heterologous object sequence, e.g., comprising a promoter and a protein-coding sequence). The payload domain is flanked by 5' and 3' protein interaction domains, e.g., sequences capable of binding the Gene Writer polypeptide, e.g., 5' and 3' UTR sequences shown in Table 3. Flanking the protein interaction domains are 5' and 3' homology domains, which have homology to the desired insertion region in the genome.

Figure 12 is a graph showing retrotransposition efficiency measured by ddPCR (digital droplet PCR) using different transfection conditions. Bars A-C represent samples that were transfected using 0.15µl Lipofectamine™ RNAiMAX with 100ng, 250ng, or 500ng respectively. Bars D-F represent samples that were transfected using 0.3µl Lipofectamine™ RNAiMAX with 100ng, 250ng, or 500ng respectively. Bars G-I represent samples that were transfected using 1µl TransIT®-mRNA transfection kit with 100ng, 250ng, or 500ng respectively.

Figure 13. Schematic of trans-transgene delivery machinery. This schematic illustrates a driver plasmid (left) with a pCEP4 backbone, which encodes the reverse transcriptase R2Tg, with a promoter and Kozak sequence upstream, and a polyadenylation signal downstream. The driver plasmid can drive expression of the GeneWriter protein. The transgene plasmid (right), with a pCDNA backbone, comprises (in order) a CMV promoter, an rDNA homology sequence, a 5' UTR, an antisense-orientation insert, a 3' UTR, a second rDNA homology sequence, a second polyadenylation signal, and a TK promoter driving a mKate2 marker. The antisense-orientation insert comprises an EF1α promoter, a coding region for EGFP that comprises an intron, and a polyadenylation signal. Use of the CMV promoter in the trangene plasmid drives

expression of a template RNA comprising the rDNA homology regions, the UTRs, and the antisense-orientation insert.

Figure 14 shows ddPCR evaluation of copy number variation of the transgene-rDNA junction in human cells across transfection conditions. Forward primer and probe were designed to bind to the 3' UTR of the R2Tg, while reverse primer was targeted to the human rDNA. The resulting ddPCR signal was normalized to that of reference assay RPP30 to determine copy number. Significantly higher average copies per genome were found with the wildtype (WT) R2Tg as compared to backbone construct with no R2Tg sequence involved. Condition 1 denotes a driver plasmid : transgene plasmid molar ratio of 9:1; condition 2 denotes the ratio is 4:1, condition 3 denotes the ratio is 1:1, condition 4 denotes the ratio is 1:4, and condition 5 denotes the ratio is 1:9.

Figures 15A and 15B. Fig. 15A: Hybrid capture of R2Tg identified on-target integrations in the human genome. The read coverage as aligned to the expected target integration in the R2 ribosomal site is indicated on the y-axis. The 5' junction between rDNA and R2Tg is indicated by the left vertical line, while the 3' junction is indicated by the right vertical line. Next-generation sequencing identifies reads spanning the expected junctions. Fig. 15B shows the number of reads from this experiment categorized as on-target integration or off-target integration at the 5' end and 3' end of the integrated sequence.

Figure 16. Sanger sequencing result of the 3' junction nested PCR. Lowercase nucleotides represent the designed SNP. Shaded uppercase nucleotides represent WT sequence. Figure 16 discloses SEQ ID NO: 1538.

Figures 17A-17F are schematic diagrams depicting various covalently dimerized Gene Writer protein configurations. 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.

Figures 18A-18F are a schematic diagam showing different modular components of a GeneWriter protein. The proteins depicted are: Fig. 18A: a wild-type full length enzyme. Fig. 18B: the DNA-binding domain of a GeneWriter may comprise zinc fingers, Cas9, or a transcription factor, or a fragment or variant of any of the forgoing. Fig. 18C: the reverse transcriptase domain and RNA-binding domain together may comprise a reverse transcriptase domain (e.g., from an R2 protein) that is heterologous to one or more other domains of the protein, and may optionally futher comprise one or more additional RNA binding domains, or a fragment or variant of any of the foregoing. Fig. 18D: the RNA binding domain may comprise, e.g., a B-box protein, an MS2 coat protein, a dCas protein, or a UTR binding protein, or a fragment or variant of any of the foregoing. Fig. 18E: the reverse transcriptase domain may comprise, e.g., a truncated reverse transcriptase domain, e.g., from an R2 protein; a reverse transcriptase domain from a virus (e.g., HIV), or a reverse transcriptase domain from AMV (avian myeloblastosis virus), or a fragment or variant of any of the foregoing. Fig. 18F: the endonuclease domain can comprise, e.g., a Cas9 nickase, a Cas ortholog, Fok I, or a restriction enzyme, or a fragment or variant of any of the foregoing. In some embodiments, a separate DNA binding domain can be attached to a polypeptide described herein (e.g., a DNA binding domain having stronger affinity for the target DNA sequence than an existing or prior DNA binding domain of the polypeptide, or a DNA binding sequence that binds to a different target DNA sequence than the existing or prior DNA binding domain of the polypeptide). In some embodiments, DNA binding domain mutants can be generated, e.g., having increased affinity to the target DNA sequence. In embodiments, the DNA binding domain comprises a zinc finger. In embodiments, the DNA binding domain is attached to the polypeptide (e.g., at the N-terminal or C-terminal ends) via a linker, e.g., as described herein. In embodiments, a zinc finger is

attached to a DNA binding domain mutant (e.g., as described herein), such that the polypeptide exhibits increased binding to the target DNA sequence (e.g., as dictated by the zinc finger) without competition with the rDNA.

Figure 19 is a graph showing linker mutant integration into the genome of HEK293T cells, assessed by a ddPCR assay evaluating copy number of R2Tg integration per genome. In v1 mutants, an insertion is located at the N-terminal side of an alpha helical region of R2Tg that preceded the predicted -1 RNA binding motif; in v2 mutants, an insertion is located at the C-terminal side of an alpha helical region of R2Tg that preceded the predicted -1 RNA binding motif; and in v3 mutants, an insertion is located C-terminal to a random coil region that came after the predicted c-myb DNA binding motif of R2Tg.

Figures 20A-20B are a series of graphs showing long-read sequencing confirming fidelity of R2Tg cis integration. Unique sequence coverage, as determined by UMI, is graphed across the expected reference sequence. The left vertical bar indicates expected 5' junction of the rDNA and R2Tg, while the right vertical bar indicates the 3' junction. Two separate amplicons spanning the 5' junction and 3' junction are shown.

Figures 21A-21B are a series of graphs showing long-read sequencing confirming fidelity of R2Tg cis integration. Unique sequence deletions (>3bp) as determined by UMI is graphed across the expected reference sequence. The left vertical bar indicates expected 5' junction of the rDNA and R2Tg, while the right vertical bar indicates the 3' junction. Two separate amplicons spanning the 5' junction and 3' junction are shown.

Figure 22 is a diagram showing exemplary plasmid map PLV033 for cis integration of R2Gfo.

Figure 23 is a graph showing integration of R2Gfo, R4Al, and R2Tg in cis in HEK293T cells. The mean of four replicates is shown; error bars indicate standard deviation.

Figure 24 is a graph showing that R2Tg integrates into human fibroblasts in cis. Integration efficiency of the wild-type (WT) and endonuclease (EN) control R2Tg were plotted over four replicate experiments as measured via ddPCR at the 3' junction of R2Tg and the rDNA target.

Figure 25 is a diagram showing Western Blot analysis for p53, p21, Actin, and Vinculin. U2OS cells were trested with the indicated compound or plasmid: GFP, R2Tg-WT (wild-type), or R2Tg-EN (endonuclease domain mutant). Plasmid transfections were performed with either lipofectamine 3000 (Lipo) or Fugene HD (Fug). Cells were analyzed 24 hours after treatment or transfection.

## 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). While not wishing to be bound by theory, in some embodiments, retrotransposases identified in homeothermic (warm blooded) species, like birds, may have improved thermostability relative to some other enzymes that evolved at lower

temperatures, and the thermostable retrotransposases may therefore be better suited for use in human cells. The disclosure also provides experimental evidence that several retrotransposases from different species, e.g., different species of animal and/or different species and clade of retrotransposon (e.g., as grouped by reverse transcriptase phylogeny, e.g., as described in Su et al. (2019) *RNA*; incorporated herein by reference in its entirety), can be used to catalyze DNA insertion into a target site in human cells (see Examples 7 and Example 28).

In some embodiments, systems described herein can have a number of advantages relative to various earlier systems. For instance, the disclosure describes retrotransposases capable of inserting long sequences (e.g., over 3000 nucleotides) of heterologous nucleic acid into a genome (see, e.g., Fig. 20A). In addition, retrotransposases described herein can insert heterologous nucleic acid in an endogenous site in the genome, such as the rDNA locus (see, e.g., Example 7). This is in contrast to Cre/loxP systems which require a first step of inserting an exogenous loxP site before a second step of inserting a sequence of interest into the loxP site.

*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.

The inventors have found that, surprisingly, 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 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.

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 retrotransposon listed in Table 1, Table 2, or Table 3. 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 retrotransposon whose DNA sequence is referenced in Table 1, Table 2, or Table 3. 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, reverse transcriptase domains are modified, for example by site-specific mutation. In embodiments, the reverse transcriptase domain is engineered to bind a heterologous template RNA.

*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. 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.

*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 replaces the endonuclease element of the polypeptide. 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 Table 1, Table 2, or Table 3. 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 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 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 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).

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."

**TABLE 1: APE-type non-LTR retrotransposon elements**

| Table 1 | | | | |
|---|---|---|---|---|
| **Family** | **Sequence Accession** | **Mobile Element** | **Name** | **Organism** |
| Dewa | AB097143 | ORF2 | Danio rerio retrotransposon DewaDr1 DNA, complete sequence | Danio rerio |
| HeT-A | KJ081250 | non-LTR retrotransposon:HeT-A | Drosophila melanogaster non-LTR retrotransposon HeT-A, partial sequence | Drosophila melanogaster |
| Keno | AB111948 | ORF2 | Tetraodon nigroviridis retrotransposon KenoTn1 DNA, partial sequence | Tetraodon nigroviridis |
| KenoDr1;Keno | AB097144 | ORF2 | Danio rerio retrotransposon KenoDr1 DNA, complete sequence | Danio rerio |
| KenoFr1;Keno | AB111947 | ORF2 | Takifugu rubripes retrotransposon KenoFr1 DNA, complete sequence | Takifugu rubripes |
| Kibi | AB097139 | ORF2 | Danio rerio retrotransposon KibiDr2 DNA, complete sequence | Danio rerio |
| Kibi | AB097138 | ORF2 | Danio rerio retrotransposon KibiDr1 DNA, complete sequence | Danio rerio |
| Kibi | AB097137 | ORF2 | Tetraodon nigroviridis retrotransposon KibiTn1 DNA, complete sequence | Tetraodon nigroviridis |
| Kibi | AB097136 | ORF2 | Takifugu rubripes retrotransposon KibiFr1 DNA, complete sequence | Takifugu rubripes |
| KoshiTn1 | AB097135 | ORF2 | Tetraodon nigroviridis retrotransposon KoshiTn1 DNA, complete sequence | Tetraodon nigroviridis |
| Mutsu | AB097142 | ORF2 | Danio rerio retrotransposon MutsuDr3 DNA, partial sequence | Danio rerio |
| Mutsu | AB097141 | ORF2 | Danio rerio retrotransposon MutsuDr2 DNA, partial sequence | Danio rerio |
| Mutsu | AB097140 | ORF2 | Danio rerio retrotransposon MutsuDr1 DNA, complete sequence | Danio rerio |
| R1 | HQ284568 | non-LTR retrotransposon:R1-like | Trilocha sp. GAS-2011 isolate TrilSp.6 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Trilocha sp. GAS-2011 |
| R1 | HQ284534 | non-LTR retrotransposon:R1-like | Scopula ornata isolate ScoOrn.6 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Scopula ornata |
| R1 | HQ284496 | non-LTR retrotransposon:R1-like | Perigonia ilus isolate PerIlus.31 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Perigonia ilus |
| R1 | HQ284489 | non-LTR retrotransposon:R1-like | Oxytenis modestia isolate OxyMod.2_3_4_7_9 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Oxytenis modestia |
| R1 | HQ284488 | non-LTR retrotransposon:R1-like | Oxytenis modestia isolate OxyMod.1 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Oxytenis modestia |

| Table 1 | | | | |
|---|---|---|---|---|
| **Family** | **Sequence Accession** | **Mobile Element** | **Name** | **Organism** |
| R1 | HQ284476 | non-LTR retrotransposon:R1-like | Oeneis magna dubia isolate OenMag.26 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Oeneis magna dubia |
| R1 | HQ284437 | non-LTR retrotransposon:R1-like | Lymantria dispar isolate LymDis.2 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Lymantria dispar |
| R1 | HQ284435 | non-LTR retrotransposon:R1-like | Lymantria dispar isolate LymDis.1 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Lymantria dispar |
| R1 | HQ284432 | non-LTR retrotransposon:R1-like | Janiodes laverna isolate JanLav.911 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Janiodes laverna |
| R1 | HQ284431 | non-LTR retrotransposon:R1-like | Janiodes laverna isolate JanLav.811 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Janiodes laverna |
| R1 | HQ284430 | non-LTR retrotransposon:R1-like | Janiodes laverna isolate JanLav.5 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Janiodes laverna |
| R1 | HQ284428 | non-LTR retrotransposon:R1-like | Janiodes laverna isolate JanLav.411 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Janiodes laverna |
| R1 | HQ284426 | non-LTR retrotransposon:R1-like | Janiodes laverna isolate JanLav.211 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Janiodes laverna |
| R1 | HQ284421 | non-LTR retrotransposon:R1-like | Heteropterus morpheus isolate HetMor.3 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Heteropterus morpheus |
| R1 | HQ284402 | non-LTR retrotransposon:R1-like | Erinnyis ello isolate EriEllo.22 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Erinnyis ello |
| R1 | HQ284399 | non-LTR retrotransposon:R1-like | Erebia theano isolate EreThe.29 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Erebia theano |
| R1 | HQ284398 | non-LTR retrotransposon:R1-like | Erebia theano isolate EreThe.28 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Erebia theano |

| Table 1 | | | | |
|---|---|---|---|---|
| **Family** | **Sequence Accession** | **Mobile Element** | **Name** | **Organism** |
| R1 | HQ284397 | non-LTR retrotransposon:R1-like | Erebia theano isolate EreThe.27 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Erebia theano |
| R1 | HQ284391 | non-LTR retrotransposon:R1-like | Emesis lucinda isolate EmeLuc.23 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Emesis lucinda |
| R1 | HQ284390 | non-LTR retrotransposon:R1-like | Emesis lucinda isolate EmeLuc.2 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Emesis lucinda |
| R1 | HQ284364 | non-LTR retrotransposon:R1-like | Coenonympha glycerion isolate CoeGly.9 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Coenonympha glycerion |
| R1 | HQ284363 | non-LTR retrotransposon:R1-like | Coenonympha glycerion isolate CoeGly.8 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Coenonympha glycerion |
| R1 | HQ284362 | non-LTR retrotransposon:R1-like | Coenonympha glycerion isolate CoeGly.7 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Coenonympha glycerion |
| R1 | HQ284361 | non-LTR retrotransposon:R1-like | Coenonympha glycerion isolate CoeGly.5 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Coenonympha glycerion |
| R1 | HQ284357 | non-LTR retrotransposon:R1-like | Coenonympha glycerion isolate CoeGly.13 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Coenonympha glycerion |
| R1 | HQ284356 | non-LTR retrotransposon:R1-like | Coenonympha glycerion isolate CoeGly.11 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Coenonympha glycerion |
| R1 | HQ284350 | non-LTR retrotransposon:R1-like | Catocyclotis adelina isolate CatAde.18 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Catocyclotis adelina |
| R1 | HQ284340 | non-LTR retrotransposon:R1-like | Caria rhacotis isolate CarRha.11 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Caria rhacotis |
| R1 | HQ284339 | non-LTR retrotransposon:R1-like | Caria rhacotis isolate CarRha.1 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Caria rhacotis |
| R1 | HQ284319 | non-LTR retrotransposon:R1-like | Archiearis parthenias isolate BrePar.1 non-LTR retrotransposon R1-like | Archiearis parthenias |

| Table 1 | | | | |
|---|---|---|---|---|
| **Family** | **Sequence Accession** | **Mobile Element** | **Name** | **Organism** |
| | | | reverse transcriptase gene, partial cds | |
| R1 | HQ284318 | non-LTR retrotransposon:R1-like | Brangas neora isolate BraNeo.32 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Brangas neora |
| R1 | HQ284292 | non-LTR retrotransposon:R1-like | Araschnia levana isolate AraLev.31 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Araschnia levana |
| R1 | HQ284286 | non-LTR retrotransposon:R1-like | Araschnia levana isolate AraLev.1 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Araschnia levana |
| R1 | HQ284280 | non-LTR retrotransposon:R1-like | Anteros formosus isolate AntForm.34 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Anteros formosus |
| R1 | HQ284279 | non-LTR retrotransposon:R1-like | Anteros formosus isolate AntForm.32 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Anteros formosus |
| R1 | HQ284278 | non-LTR retrotransposon:R1-like | Anteros formosus isolate AntForm.31 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Anteros formosus |
| R1 | HQ284270 | non-LTR retrotransposon:R1-like | Agrotis exclamationis isolate AgrExcl.27 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Agrotis exclamationis |
| R1 | HQ284267 | non-LTR retrotransposon:R1-like | Agrius cingulata isolate AgrCing.36_39 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Agrius cingulata |
| R1 | HQ284266 | non-LTR retrotransposon:R1-like | Agrius cingulata isolate AgrCing.3 non-LTR retrotransposon R1-like reverse transcriptase-like gene, partial sequence | Agrius cingulata |
| R1 | HQ284263 | non-LTR retrotransposon:R1-like | Aglia tau isolate AglTau.8 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Aglia tau |
| R1 | HQ284262 | non-LTR retrotransposon:R1-like | Aglia tau isolate AglTau.7 non-LTR retrotransposon R1-like reverse transcriptase gene, partial cds | Aglia tau |
| R1 | DQ836362 | MalR1 | Maculinea alcon R1-like non-LTR retrotransposon R1 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris alcon |

| Table 1 | | | | |
|---|---|---|---|---|
| **Family** | **Sequence Accession** | **Mobile Element** | **Name** | **Organism** |
| R1 | DQ836391 | MnaR1 | Maculinea nausithous R1-like non-LTR retrotransposon R1 reverse transcriptase (RT) gene, partial cds | Phengaris nausithous |
| R1 | KU543683 | non-LTR retrotransposon and non-LTR retrovirus reverse transcriptase; Region: RT_nLTR | Bactrocera tryoni clone Btry_5404 non-LTR retrotransposon R1, complete sequence | Bactrocera tryoni |
| R1 | KU543682 | non-LTR retrotransposon and non-LTR retrovirus reverse transcriptase; Region: RT_nLTR | Bactrocera tryoni clone Btry_5167 non-LTR retrotransposon R1, complete sequence | Bactrocera tryoni |
| R1 | KU543679 | non-LTR retrotransposon and non-LTR retrovirus reverse transcriptase; Region: RT_nLTR | Bactrocera tryoni clone Btry_4956 non-LTR retrotransposon R1, complete sequence | Bactrocera tryoni |
| R1 | KU543678 | non-LTR retrotransposon and non-LTR retrovirus reverse transcriptase; Region: RT_nLTR | Bactrocera tryoni clone Btry_5979 non-LTR retrotransposon R1, complete sequence | Bactrocera tryoni |
| R1 | AB078933 | ORF1 | Papilio xuthus non-LTR retrotransposon gene for gag-like protein, partial cds, clone: SARTPx2-2 | Papilio xuthus |
| R1 | AB078932 | ORF1 | Papilio xuthus non-LTR retrotransposon gene for gag-like protein, partial cds, clone: SARTPx2-1 | Papilio xuthus |
| R1 | AB078936 | ORF2 | Papilio xuthus non-LTR retrotransposon genes for gag-like protein, reverse transcriptase, partial cds, clone: SARTPx4-N18 | Papilio xuthus |
| R1 | AB078935 | ORF2 | Papilio xuthus non-LTR retrotransposon genes for gag-like protein, reverse transcriptase, partial and complete cds, clone: SARTPx3-N7 | Papilio xuthus |
| R1 | AB078934 | ORF2 | Papilio xuthus non-LTR retrotransposon genes for gag-like protein, reverse transcriptase, partial cds, clone: SARTPx3-N3 | Papilio xuthus |
| R1 | AB078931 | ORF2 | Papilio xuthus non-LTR retrotransposon genes for gag-like protein, reverse transcriptase, partial and complete cds, clone: SARTPx1-N14 | Papilio xuthus |
| R1 | AB078930 | ORF2 | Papilio xuthus non-LTR retrotransposon genes for gag-like | Papilio xuthus |

| Table 1 | | | | |
|---|---|---|---|---|
| **Family** | **Sequence Accession** | **Mobile Element** | **Name** | **Organism** |
| | | | protein, reverse transcriptase, partial and complete cds, clone: SARTPx1-N5 | |
| R1 | AB078929 | ORF2 | Papilio xuthus non-LTR retrotransposon genes for gag-like protein, reverse transcrpitase, partial and complete cds, clone: SARTPx1-N4 | Papilio xuthus |
| R1 | AB078928 | ORF2 | Papilio xuthus non-LTR retrotransposon gene for gag-like protein, reverse transcrpitase, complete and partial cds, clone: SARTPx1-3 | Papilio xuthus |
| R1 | KP771712 | ORF2; contains endonuclease, reverse transcriptase and RNaseH | Blattella germanica non-LTR retrotransposon TRAS-like 2, complete sequence | Blattella germanica |
| R1 | KP771711 | ORF2; contains endonuclease, reverse transcriptase and RNaseH | Blattella germanica non-LTR retrotransposon TRAS-like 1, complete sequence | Blattella germanica |
| R1 | AF015813 | R1 ORF | Dugesiella sp. retrotransposon R1 reverse transcriptase gene, partial cds | Aphonopelma sp. WDB-1998 |
| R1 | AF015489 | R1 ORF | Dugesiella sp. retrotransposon R1 reverse transcriptase gene, partial cds | Aphonopelma sp. WDB-1998 |
| R1Bm | AB182560 | non-LTR retrotransposon R1Bmks ORF2 | Bombyx mori genes for non-LTR retrotransposon R1Bmks ORF1 protein, non-LTR retrotransposon R1Bmks ORF2 protein, complete cds | Bombyx mori |
| R6 | AB090819 | ORF2 | Anopheles gambiae retrotransposon R6Ag3 DNA, complete sequence | Anopheles gambiae |
| R6 | AB090818 | ORF2 | Anopheles gambiae retrotransposon R6Ag2 DNA, complete sequence | Anopheles gambiae |
| R6 | AB090817 | ORF2 | Anopheles gambiae retrotransposon R6Ag1 DNA, complete sequence | Anopheles gambiae |
| R6 | KJ958615 | R2 | Bacillus rossius non-LTR retrotransposon reVIR6, partial sequence | Bacillus rossius |
| R6 | KJ958596 | R2 | Bacillus rossius non-LTR retrotransposon reBER6, partial sequence | Bacillus rossius |
| R6 | AF352480 | transposon:NLRCth1-like non-LTR retrotransposon | Chironomus circumdatus clone cir6 transposon NLRCth1-like non-LTR retrotransposon reverse transcriptase gene, partial cds | Chironomus circumdatus |
| R6 | AF373367 | transposon:non-LTR retrotransposon LINE2 | Clelia rustica clone CR6 non-LTR retrotransposon LINE2 reverse | Paraphimophis rusticus |

| Table 1 | | | | |
|---|---|---|---|---|
| **Family** | **Sequence Accession** | **Mobile Element** | **Name** | **Organism** |
| | | | transcriptase pseudogene, partial sequence | |
| R7 | AB090820 | ORF2 | Anopheles gambiae retrotransposon R7Ag1 DNA, complete sequence | Anopheles gambiae |
| R7 | AB090821 | ORF2 | Anopheles gambiae retrotransposon R7Ag2 DNA, complete sequence | Anopheles gambiae |
| R7 | KJ958622 | R2 | Bacillus rossius non-LTR retrotransposon trKOR7, partial sequence | Bacillus rossius |
| R7 | KJ958616 | R2 | Bacillus rossius non-LTR retrotransposon reVIR7, partial sequence | Bacillus rossius |
| R7 | KJ958597 | R2 | Bacillus rossius non-LTR retrotransposon reBER7, partial sequence | Bacillus rossius |
| R7 | AF352514 | transposon:NLRCth1-like non-LTR retrotransposon | Chironomus circumdatus clone cir7 transposon NLRCth1-like non-LTR retrotransposon reverse transcriptase pseudogene, partial sequence | Chironomus circumdatus |
| Rt2 | AY379084 | truncated; similar to reverse transcriptase | Leptocheirus plumulosus retrotransposon LpRt2, partial sequence | Leptocheirus plumulosus |
| Rt2 | MSQRT2RET | | Anopheles gambiae retrotransposon RT2, complete sequence | Anopheles gambiae |
| RTAg4 | AB090813 | ORF2 | Anopheles gambiae retrotransposon RTAg4 DNA, complete sequence | Anopheles gambiae |
| TRAS1 | BMOTRAS1 | DNA binding domain at AA1103-1120. | Bombyx mori gene, complete sequence of retrotransposon TRAS1 | Bombyx mori |
| TRAS3 | JX875955 | similar to reverse transcriptases | Acyrthosiphon pisum clone LSR1 non-LTR retrotransposon TRAS3, complete sequence | Acyrthosiphon pisum |
| Tx1 | AJ621359 | transposon:non-LTR retrotransposon TX1-1_Tet | Tetraodon nigroviridis non-LTR retrotransposon TX1-1_Tet, complete sequence | Tetraodon nigroviridis |
| Tx1 | AJ621360 | transposon:non-LTR retrotransposon TX1-2_Tet | Tetraodon nigroviridis partial non-LTR retrotransposon TX1-2_Tet | Tetraodon nigroviridis |
| Tx1 | AJ621361 | transposon:non-LTR retrotransposon TX1-3_Tet | Tetraodon nigroviridis partial non-LTR retrotransposon TX1-3_Tet | Tetraodon nigroviridis |
| Tx1 | AJ621362 | transposon:non-LTR retrotransposon TX1-4_Tet | Tetraodon nigroviridis partial non-LTR retrotransposon TX1-4_Tet | Tetraodon nigroviridis |
| Tx1 | DQ118004 | transposon:Tx1-like retrotransposon Tx1Aru | Acipenser ruthenus clone dg194 transposon Tx1-like retrotransposon Tx1Aru reverse transcriptase-like gene, partial sequence | Acipenser ruthenus |

| Table 1 | | | | |
|---|---|---|---|---|
| **Family** | **Sequence Accession** | **Mobile Element** | **Name** | **Organism** |
| Tx1 | AB097134 | ORF2 | Takifugu rubripes retrotransposon KoshiFr1 DNA, complete sequence | Takifugu rubripes |
| Tx1 | AB090816 | ORF2 | Anopheles gambiae retrotransposon MinoAg1 DNA, complete sequence | Anopheles gambiae |
| Tx1 | AB090812 | ORF2 | Anopheles gambiae retrotransposon RTAg3 DNA, complete sequence | Anopheles gambiae |
| Waldo | AH009917 | non-LTR retrotransposon:Waldo-A | Drosophila melanogaster Waldo-A non-LTR retrotransposon, 5' sequence | Drosophila melanogaster |
| Waldo | AH009916 | non-LTR retrotransposon:Waldo-A | Drosophila melanogaster clone CBE9 Waldo-A non-LTR retrotransposon, 5' sequence | Drosophila melanogaster |
| Waldo | AH009915 | non-LTR retrotransposon:Waldo-A | Drosophila melanogaster Waldo-A non-LTR retrotransposon, 5' sequence | Drosophila melanogaster |
| Waldo | AH009914 | non-LTR retrotransposon:Waldo-A | Drosophila melanogaster Waldo-A non-LTR retrotransposon | Drosophila melanogaster |
| Waldo | AH009920 | non-LTR retrotransposon:Waldo-B | Drosophila melanogaster Waldo-B non-LTR retrotransposon, 5' sequence | Drosophila melanogaster |
| Waldo | AH009919 | non-LTR retrotransposon:Waldo-B | | Drosophila melanogaster |
| Waldo | AH009918 | non-LTR retrotransposon:Waldo-B | | Drosophila melanogaster |
| Waldo | AB090815 | ORF2 | Anopheles gambiae retrotransposon WaldoAg2 DNA, complete sequence | Anopheles gambiae |
| Waldo | AB090814 | ORF2 | Anopheles gambiae retrotransposon WaldoAg1 DNA, complete sequence | Anopheles gambiae |
| Waldo | AB078939 | ORF2 | Forficula scudderi non-LTR retrotransposon pseudogene for reverse transcriptase, clone: WaldoFs1-26 | Forficula scudderi |
| Waldo | AB078938 | ORF2 | Forficula scudderi non-LTR retrotransposon pseudogene for reverse transcriptase, clone: WaldoFs1-2 | Forficula scudderi |
| Waldo | AB078937 | ORF2 | Forficula scudderi non-LTR retrotransposon pseudogene for reverse transcriptase, clone: WaldoFs1-1 | Forficula scudderi |

## TABLE 2: RLE-type non-LTR retrotransposon elements

| Table 2 | | | | |
|---|---|---|---|---|
| Family | Accession | Mobile Element | Name/Description | Organism |
| CRE | EF067892 | | Colletotrichum cereale clone 9F8-1558 Ccret3 non-LTR retrotransposon, partial sequence | Colletotrichum cereale |
| CRE | EF067894 | | Colletotrichum cereale clone 9F8-2137 Ccret3 non-LTR retrotransposon, partial sequence | Colletotrichum cereale |
| CRE | MG028000 | non-LTR retrotransposon:Rex3 | Characidium gomesi voucher MNRJ20998 non-LTR retrotransposon Rex3, partial sequence | Characidium gomesi |
| CRE | KY566213 | non-LTR retrotransposon:Rex3 | Characidium gomesi non-LTR retrotransposon Rex3, partial sequence | Characidium gomesi |
| CRE | GU949558 | | Kalotermes flavicollis isolate Crete non-LTR retrotransposon R2, complete sequence; and R2 protein gene, complete cds | Kalotermes flavicollis |
| CRE;CRE2 | CFU19151 | poly dA tracts in 5' and 3' UTRs | Crithidia fasciculata retrotransposon CRE2 in mini-exon gene, putative reverse transcriptase gene, complete cds | Crithidia fasciculata |
| CZAR | BR000987 | pol | TPA_inf: Capsaspora owczarzaki DNA, non-LTR retrotransposon CoL4, complete sequence, strain: ATCC 30864 | Capsaspora owczarzaki |
| CZAR | BR000986 | pol | TPA_inf: Capsaspora owczarzaki DNA, non-LTR retrotransposon CoL3, complete sequence, strain: ATCC 30864 | Capsaspora owczarzaki |
| CZAR | BR000985 | pol | TPA_inf: Capsaspora owczarzaki DNA, non-LTR retrotransposon CoL2, complete sequence, strain: ATCC 30864 | Capsaspora owczarzaki |
| CZAR | BR000984 | pol | TPA_inf: Capsaspora owczarzaki DNA, non-LTR retrotransposon CoL1, complete sequence, strain: ATCC 30864 | Capsaspora owczarzaki |
| DongAG;Dong | AB097127 | rt | Anopheles gambiae retrotransposon DongAg DNA, partial sequence | Anopheles gambiae |
| EhRLE2 | AB097128 | rt | Entamoeba histolytica retrotransposon EhRLE2 DNA, complete sequence | Entamoeba histolytica |
| EhRLE3 | AB097129 | rt | Entamoeba histolytica retrotransposon EhRLE3 DNA, complete sequence | Entamoeba histolytica |
| Genie | AF440196 | endonuclease | Giardia intestinalis non-LTR retrotransposon GENIE 1 pol polyprotein gene, complete cds | Giardia intestinalis |
| Genie | BK000097 | endonuclease | TPA_exp: Giardia intestinalis non-LTR retrotransposon Genie 1A gene, partial sequence | Giardia intestinalis |
| Genie | BK000095 | endonuclease | TPA_exp: Giardia intestinalis non-LTR retrotransposon Genie 1 gene, partial sequence | Giardia intestinalis |
| Genie | BK000096 | insertion site for non-LTR retrotransposon Genie 1 | TPA_exp: Giardia intestinalis non-LTR retrotransposon Genie 1 target site sequence | Giardia intestinalis |

| Table 2 | | | | |
|---|---|---|---|---|
| Genie | AY216701 | non-experimental evidence, no additional details recorded | Girardia tigrina GENIE retrotransposon, complete sequence | Girardia tigrina |
| Genie | BK000098 | similar to endonuclease | TPA_exp: Giardia intestinalis non-LTR retrotransposon Genie 2 gene, complete sequence | Giardia intestinalis |
| GilD | AF433877 | (tca)n (SEQ ID NO: 1532) or (tga)n (SEQ ID NO: 1533), n=2-4 | Giardia intestinalis inactive non-LTR retrotransposon GilD, consensus sequence | Giardia intestinalis |
| GilM | AF433875 | poly(dA) tract | Giardia intestinalis non-LTR LINE-like retrotransposon GilM, complete sequence | Giardia intestinalis |
| Hero | AB097132 | rt | Danio rerio retrotransposon HERODr DNA, complete sequence | Danio rerio |
| Hero | AB097130 | rt | Takifugu rubripes retrotransposon HEROFr DNA, complete sequence | Takifugu rubripes |
| HEROTn | AB097131 | rt | Tetraodon nigroviridis retrotransposon HEROTn DNA, complete sequence | Tetraodon nigroviridis |
| NeSL_3_135_68117 | FJ905846 | non-LTR retrotransposon:NeSL_3_135_68117 | Daphnia pulex non-LTR retrotransposon NeSL_3_135_68117, complete sequence | Daphnia pulex |
| NeSL; NeSL-1 | DQ099731 | target site duplication | Caenorhabditis briggsae transposon NeSl-1-like non-LTR retrotransposon NeSL-1Cb reverse transcriptase (pol) gene, complete cds | Caenorhabditis briggsae |
| PERERE-9 | BN000800 | | TPA_exp: Schistosoma mansoni Perere-9 non-LTR retrotransposon | Schistosoma mansoni |
| R2 | AF015814 | R2 | Limulus polyphemus retrotransposon R2, complete sequence | Limulus polyphemus |
| R2 | AF090145 | R2 | Nasonia vitripennis R2 non-LTR retrotransposable element reverse transcriptase gene, partial cds | Nasonia vitripennis |
| R2 | AF015818 | R2 | Porcellio scaber retrotransposon R2, complete sequence | Porcellio scaber |
| R2 | AF015815 | R2 | Anurida maritima retrotransposon R2, complete sequence | Anurida maritima |
| R2 | M16558 | R2 | Bombyx mori rDNA insertion element R2 (typeII), complete cds | Bombyx mori |
| R2 | AF015819 | R2 | Forficula auricularia retrotransposon R2, complete sequence. | Forficula auricularia |
| R2 | EU854578 | R2 | Triops cancriformis non-LTR retrotransposon R2 reverse transcriptase gene, complete cds | Triops cancriformis |
| R2 | GU949555 | R2 | Reticulitermes lucifugus non-LTR retrotransposon R2, complete sequence; and R2 protein gene, complete cds | Reticulitermes lucifugus |
| R2 | AB097123 | rt | Ciona intestinalis retrotransposon R2Ci-C DNA, partial sequence | Ciona intestinalis |
| R2 | AB097124 | rt | Ciona intestinalis retrotransposon R2Ci-D DNA, partial sequence | Ciona intestinalis |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | FJ461304 | R2 | Rhynchosciara americana non-LTR retrotransposon RaR2 reverse transcriptase gene, complete cds | Rhynchosciara americana |
| R2 | AB097121 | rt | Ciona intestinalis retrotransposon R2Ci-A DNA, complete sequence | Ciona intestinalis |
| R2 | KP657892 | R2 | Bacillus rossius isolate roCAP(full).9 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657890 | R2 | Bacillus rossius isolate roCAP(full).7 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657888 | R2 | Bacillus rossius isolate roCAP(full).5 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657870 | R2 | Bacillus rossius isolate roCAP(full).1 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657833 | R2 | Bacillus rossius isolate roANZ(full).13 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657832 | R2 | Bacillus rossius isolate roANZ(full).12 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657830 | R2 | Bacillus rossius isolate roANZ(full).10 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657807 | R2 | Bacillus rossius isolate roANZ(-101).8 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657806 | R2 | Bacillus rossius isolate roANZ(-101).7 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657805 | R2 | Bacillus rossius isolate roANZ(-101).6 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657802 | R2 | Bacillus rossius isolate roANZ(-101).3 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657799 | R2 | Bacillus rossius isolate roANZ(-101).1 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | FJ461304 | R2 | Rhynchosciara americana non-LTR retrotransposon RaR2 reverse transcriptase gene, complete cds | Rhynchosciara americana |
| R2 | JQ082370 | polyA_signal_sequence | Eyprepocnemis plorans non-LTR retrotransposon R2 R2 protein gene, complete cds | Eyprepocnemis plorans |
| R2 | KJ958672 | R2 | Bacillus rossius non-LTR retrotransposon reCUR4_deg, partial sequence | Bacillus rossius |
| R2 | KJ958671 | R2 | Bacillus rossius non-LTR retrotransposon reCUR3_deg, partial sequence | Bacillus rossius |
| R2 | KJ958670 | R2 | Bacillus rossius non-LTR retrotransposon reCUR2_deg, partial sequence | Bacillus rossius |
| R2 | KJ958669 | R2 | Bacillus rossius non-LTR retrotransposon reCUR1_deg, partial sequence | Bacillus rossius |
| R2 | KJ958668 | R2 | Bacillus rossius non-LTR retrotransposon reCDF6_deg, partial sequence | Bacillus rossius |
| R2 | KJ958667 | R2 | Bacillus rossius non-LTR retrotransposon reCDF5_deg, partial sequence | Bacillus rossius |
| R2 | KJ958666 | R2 | Bacillus rossius non-LTR retrotransposon reCDF4_deg, partial sequence | Bacillus rossius |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | KJ958665 | R2 | Bacillus rossius non-LTR retrotransposon reCDF3_deg, partial sequence | Bacillus rossius |
| R2 | KJ958664 | R2 | Bacillus rossius non-LTR retrotransposon reCDF2_deg, partial sequence | Bacillus rossius |
| R2 | KJ958663 | R2 | Bacillus rossius non-LTR retrotransposon reCDF1_deg, partial sequence | Bacillus rossius |
| R2 | KJ958662 | R2 | Bacillus rossius non-LTR retrotransposon reCOM10_deg, partial sequence | Bacillus rossius |
| R2 | KJ958661 | R2 | Bacillus rossius non-LTR retrotransposon reCOM5_deg, partial sequence | Bacillus rossius |
| R2 | KJ958660 | R2 | Bacillus rossius non-LTR retrotransposon reCOM7_deg, partial sequence | Bacillus rossius |
| R2 | KJ958659 | R2 | Bacillus rossius non-LTR retrotransposon reCOM3_deg, partial sequence | Bacillus rossius |
| R2 | KJ958658 | R2 | Bacillus rossius non-LTR retrotransposon reCOM2_deg, partial sequence | Bacillus rossius |
| R2 | KJ958657 | R2 | Bacillus rossius non-LTR retrotransposon reMSN6_deg, partial sequence | Bacillus rossius |
| R2 | KJ958656 | R2 | Bacillus rossius non-LTR retrotransposon reMSN5_deg, partial sequence | Bacillus rossius |
| R2 | KJ958655 | R2 | Bacillus rossius non-LTR retrotransposon reMSN4_deg, partial sequence | Bacillus rossius |
| R2 | KJ958654 | R2 | Bacillus rossius non-LTR retrotransposon reMSN3_deg, partial sequence | Bacillus rossius |
| R2 | KJ958653 | R2 | Bacillus rossius non-LTR retrotransposon reMSN2_deg, partial sequence | Bacillus rossius |
| R2 | KJ958652 | R2 | Bacillus rossius non-LTR retrotransposon reMSN1_deg, partial sequence | Bacillus rossius |
| R2 | KJ958651 | R2 | Bacillus rossius non-LTR retrotransposon rePAT8_deg, partial sequence | Bacillus rossius |
| R2 | KJ958650 | R2 | Bacillus rossius non-LTR retrotransposon rePAT7_deg, partial sequence | Bacillus rossius |
| R2 | KJ958649 | R2 | Bacillus rossius non-LTR retrotransposon rePAT6_deg, partial sequence | Bacillus rossius |
| R2 | KJ958648 | R2 | Bacillus rossius non-LTR retrotransposon rePAT5_deg, partial sequence | Bacillus rossius |
| R2 | KJ958647 | R2 | Bacillus rossius non-LTR retrotransposon rePAT4_deg, partial sequence | Bacillus rossius |
| R2 | KJ958646 | R2 | Bacillus rossius non-LTR retrotransposon rePAT3_deg, partial sequence | Bacillus rossius |
| R2 | KJ958645 | R2 | Bacillus rossius non-LTR retrotransposon rePAT2_deg, partial sequence | Bacillus rossius |
| R2 | KJ958644 | R2 | Bacillus rossius non-LTR retrotransposon rePAT1_deg, partial sequence | Bacillus rossius |
| R2 | KJ958643 | R2 | Bacillus rossius non-LTR retrotransposon rePAT9_deg, partial sequence | Bacillus rossius |
| R2 | KJ958642 | R2 | Bacillus rossius non-LTR retrotransposon trKOR2_deg, partial sequence | Bacillus rossius |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | KJ958641 | R2 | Bacillus rossius non-LTR retrotransposon reGAB9_deg, partial sequence | Bacillus rossius |
| R2 | KJ958640 | R2 | Bacillus rossius non-LTR retrotransposon reGAB8_deg, partial sequence | Bacillus rossius |
| R2 | KJ958639 | R2 | Bacillus rossius non-LTR retrotransposon reGAB7_deg, partial sequence | Bacillus rossius |
| R2 | KJ958638 | R2 | Bacillus rossius non-LTR retrotransposon reGAB6_deg, partial sequence | Bacillus rossius |
| R2 | KJ958637 | R2 | Bacillus rossius non-LTR retrotransposon reGAB5_deg, partial sequence | Bacillus rossius |
| R2 | KJ958636 | R2 | Bacillus rossius non-LTR retrotransposon reGAB4_deg, partial sequence | Bacillus rossius |
| R2 | KJ958635 | R2 | Bacillus rossius non-LTR retrotransposon reGAB3_deg, partial sequence | Bacillus rossius |
| R2 | KJ958634 | R2 | Bacillus rossius non-LTR retrotransposon reGAB2_deg, partial sequence | Bacillus rossius |
| R2 | KJ958633 | R2 | Bacillus rossius non-LTR retrotransposon reGAB1_deg, partial sequence | Bacillus rossius |
| R2 | KJ958632 | R2 | Bacillus rossius non-LTR retrotransposon reTDS1_deg, partial sequence | Bacillus rossius |
| R2 | KJ958631 | R2 | Bacillus rossius non-LTR retrotransposon reTDS11_deg, partial sequence | Bacillus rossius |
| R2 | KJ958629 | R2 | Bacillus rossius non-LTR retrotransposon reCOM8, partial sequence | Bacillus rossius |
| R2 | KJ958628 | R2 | Bacillus rossius non-LTR retrotransposon reCOM6, partial sequence | Bacillus rossius |
| R2 | KJ958627 | R2 | Bacillus rossius non-LTR retrotransposon reCOM4, partial sequence | Bacillus rossius |
| R2 | KJ958626 | R2 | Bacillus rossius non-LTR retrotransposon reCOM1, partial sequence | Bacillus rossius |
| R2 | KJ958624 | R2 | Bacillus rossius non-LTR retrotransposon trKOR10, partial sequence | Bacillus rossius |
| R2 | KJ958623 | R2 | Bacillus rossius non-LTR retrotransposon reGAB10, partial sequence | Bacillus rossius |
| R2 | KJ958619 | R2 | Bacillus rossius non-LTR retrotransposon trKOR4, partial sequence | Bacillus rossius |
| R2 | KJ958618 | R2 | Bacillus rossius non-LTR retrotransposon trKOR3, partial sequence | Bacillus rossius |
| R2 | KJ958617 | R2 | Bacillus rossius non-LTR retrotransposon trKOR1, partial sequence | Bacillus rossius |
| R2 | KJ958613 | R2 | Bacillus rossius non-LTR retrotransposon reVIR4, partial sequence | Bacillus rossius |
| R2 | KJ958612 | R2 | Bacillus rossius non-LTR retrotransposon reVIR3, partial sequence | Bacillus rossius |
| R2 | KJ958611 | R2 | Bacillus rossius non-LTR retrotransposon reVIR2, partial sequence | Bacillus rossius |
| R2 | KJ958610 | R2 | Bacillus rossius non-LTR retrotransposon reVIR1, partial sequence | Bacillus rossius |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | KJ958609 | R2 | Bacillus rossius non-LTR retrotransposon reTDS17, partial sequence | Bacillus rossius |
| R2 | KJ958608 | R2 | Bacillus rossius non-LTR retrotransposon reTDS16, partial sequence | Bacillus rossius |
| R2 | KJ958607 | R2 | Bacillus rossius non-LTR retrotransposon reTDS15, partial sequence | Bacillus rossius |
| R2 | KJ958606 | R2 | Bacillus rossius non-LTR retrotransposon reTDS14, partial sequence | Bacillus rossius |
| R2 | KJ958605 | R2 | Bacillus rossius non-LTR retrotransposon reTDS13, partial sequence | Bacillus rossius |
| R2 | KJ958604 | R2 | Bacillus rossius non-LTR retrotransposon reTDS12, partial sequence | Bacillus rossius |
| R2 | KJ958603 | R2 | Bacillus rossius non-LTR retrotransposon reTDS7, partial sequence | Bacillus rossius |
| R2 | KJ958602 | R2 | Bacillus rossius non-LTR retrotransposon reTDS6, partial sequence | Bacillus rossius |
| R2 | KJ958601 | R2 | Bacillus rossius non-LTR retrotransposon reTDS5, partial sequence | Bacillus rossius |
| R2 | KJ958600 | R2 | Bacillus rossius non-LTR retrotransposon reTDS4, partial sequence | Bacillus rossius |
| R2 | KJ958599 | R2 | Bacillus rossius non-LTR retrotransposon reTDS3, partial sequence | Bacillus rossius |
| R2 | KJ958598 | R2 | Bacillus rossius non-LTR retrotransposon reTDS2, partial sequence | Bacillus rossius |
| R2 | KJ958594 | R2 | Bacillus rossius non-LTR retrotransposon reBER4, partial sequence | Bacillus rossius |
| R2 | KJ958593 | R2 | Bacillus rossius non-LTR retrotransposon reBER2, partial sequence | Bacillus rossius |
| R2 | KJ958592 | R2 | Bacillus rossius non-LTR retrotransposon reBER1, partial sequence | Bacillus rossius |
| R2 | KJ958591 | R2 | Bacillus rossius non-LTR retrotransposon roFOL7, partial sequence | Bacillus rossius |
| R2 | KJ958590 | R2 | Bacillus rossius non-LTR retrotransposon roFOL6, partial sequence | Bacillus rossius |
| R2 | KJ958589 | R2 | Bacillus rossius non-LTR retrotransposon roFOL5, partial sequence | Bacillus rossius |
| R2 | KJ958588 | R2 | Bacillus rossius non-LTR retrotransposon roFOL4, partial sequence | Bacillus rossius |
| R2 | KJ958587 | R2 | Bacillus rossius non-LTR retrotransposon roFOL3, partial sequence | Bacillus rossius |
| R2 | KJ958586 | R2 | Bacillus rossius non-LTR retrotransposon roFOL2, partial sequence | Bacillus rossius |
| R2 | KJ958585 | R2 | Bacillus rossius non-LTR retrotransposon roFOL1, partial sequence | Bacillus rossius |
| R2 | KJ958584 | R2 | Bacillus rossius non-LTR retrotransposon roANZ10, partial sequence | Bacillus rossius |
| R2 | KJ958583 | R2 | Bacillus rossius non-LTR retrotransposon roANZ9, partial sequence | Bacillus rossius |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | KJ958582 | R2 | Bacillus rossius non-LTR retrotransposon roANZ8, partial sequence | Bacillus rossius |
| R2 | KJ958581 | R2 | Bacillus rossius non-LTR retrotransposon roANZ7, partial sequence | Bacillus rossius |
| R2 | KJ958580 | R2 | Bacillus rossius non-LTR retrotransposon roANZ6, partial sequence | Bacillus rossius |
| R2 | KJ958579 | R2 | Bacillus rossius non-LTR retrotransposon roANZ5, partial sequence | Bacillus rossius |
| R2 | KJ958578 | R2 | Bacillus rossius non-LTR retrotransposon roANZ4, partial sequence | Bacillus rossius |
| R2 | KJ958577 | R2 | Bacillus rossius non-LTR retrotransposon roANZ3, partial sequence | Bacillus rossius |
| R2 | KJ958576 | R2 | Bacillus rossius non-LTR retrotransposon roANZ2, partial sequence | Bacillus rossius |
| R2 | KJ958575 | R2 | Bacillus rossius non-LTR retrotransposon roANZ1, partial sequence | Bacillus rossius |
| R2 | KJ958574 | R2 | Bacillus rossius non-LTR retrotransposon roCAP9, partial sequence | Bacillus rossius |
| R2 | KJ958573 | R2 | Bacillus rossius non-LTR retrotransposon roCAP8, partial sequence | Bacillus rossius |
| R2 | KJ958572 | R2 | Bacillus rossius non-LTR retrotransposon roCAP7, partial sequence | Bacillus rossius |
| R2 | KJ958571 | R2 | Bacillus rossius non-LTR retrotransposon roCAP6, partial sequence | Bacillus rossius |
| R2 | KJ958570 | R2 | Bacillus rossius non-LTR retrotransposon roCAP5, partial sequence | Bacillus rossius |
| R2 | KJ958569 | R2 | Bacillus rossius non-LTR retrotransposon roCAP4, partial sequence | Bacillus rossius |
| R2 | KJ958568 | R2 | Bacillus rossius non-LTR retrotransposon roCAP3, partial sequence | Bacillus rossius |
| R2 | KJ958567 | R2 | Bacillus rossius non-LTR retrotransposon roCAP2, partial sequence | Bacillus rossius |
| R2 | KJ958566 | R2 | Bacillus rossius non-LTR retrotransposon roCAP1, partial sequence | Bacillus rossius |
| R2 | KJ958565 | R2 | Bacillus rossius non-LTR retrotransposon roCAP10, partial sequence | Bacillus rossius |
| R2 | JN937654 | R2 | Lepidurus apus lubbocki isolate lu8a 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937653 | R2 | Lepidurus apus lubbocki isolate lu7a 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937652 | R2 | Lepidurus apus lubbocki isolate lu2a 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937651 | R2 | Lepidurus apus lubbocki isolate b7c7 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | JN937650 | R2 | Lepidurus apus lubbocki isolate b6c4 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937649 | R2 | Lepidurus apus lubbocki isolate lu5a 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937648 | R2 | Lepidurus apus lubbocki isolate lu1a 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937647 | R2 | Lepidurus apus lubbocki isolate b6c5 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937646 | R2 | Lepidurus apus lubbocki isolate b6c6 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937645 | R2 | Lepidurus apus lubbocki isolate lu4a 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937644 | R2 | Lepidurus apus lubbocki isolate lu3a 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937643 | R2 | Lepidurus apus lubbocki isolate b6c3 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937642 | R2 | Lepidurus apus lubbocki isolate lu6a 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon psR2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937641 | R2 | Lepidurus apus lubbocki isolate LM5h2 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937640 | R2 | Lepidurus apus lubbocki isolate LM2h5 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937639 | R2 | Lepidurus apus lubbocki isolate LM2h4 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937638 | R2 | Lepidurus apus lubbocki isolate LM5h5 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937637 | R2 | Lepidurus apus lubbocki isolate LM5h4 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937636 | R2 | Lepidurus apus lubbocki isolate LM5h3 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937635 | R2 | Lepidurus apus lubbocki isolate LM5h1 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | JN937634 | R2 | Lepidurus apus lubbocki isolate LM2h3 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937633 | R2 | Lepidurus apus lubbocki isolate LM2h2 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937632 | R2 | Lepidurus apus lubbocki isolate LM2h1 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2Ll, complete sequence | Lepidurus apus lubbocki |
| R2 | JN937631 | R2 | Lepidurus arcticus isolate T6 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937630 | R2 | Lepidurus arcticus isolate T5 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937629 | R2 | Lepidurus arcticus isolate T4 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937628 | R2 | Lepidurus arcticus isolate T3 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937627 | R2 | Lepidurus arcticus isolate T2 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937626 | R2 | Lepidurus arcticus isolate T1 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937625 | R2 | Lepidurus arcticus isolate V4 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937624 | R2 | Lepidurus arcticus isolate V3 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937623 | R2 | Lepidurus arcticus isolate V2 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937622 | R2 | Lepidurus arcticus isolate V1 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2La, complete sequence | Lepidurus arcticus |
| R2 | JN937615 | R2 | Lepidurus couesii isolate D3a7f 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937614 | R2 | Lepidurus couesii isolate D3a5f 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937613 | R2 | Lepidurus couesii isolate D3a4f 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | JN937612 | R2 | Lepidurus couesii isolate D3a3f 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937611 | R2 | Lepidurus couesii isolate D3a2f 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937610 | R2 | Lepidurus couesii isolate D3_8 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937609 | R2 | Lepidurus couesii isolate D3_7 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937608 | R2 | Lepidurus couesii isolate D3_6 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937607 | R2 | Lepidurus couesii isolate D3_5 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937606 | R2 | Lepidurus couesii isolate D3_4 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937605 | R2 | Lepidurus couesii isolate D3_3 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937604 | R2 | Lepidurus couesii isolate D3_2 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937603 | R2 | Lepidurus couesii isolate D3_1 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcB, complete sequence | Lepidurus couesii |
| R2 | JN937602 | R2 | Lepidurus couesii isolate C2_5 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | JN937601 | R2 | Lepidurus couesii isolate LcoC2r1_5 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | JN937600 | R2 | Lepidurus couesii isolate LcoC2r1_6 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | JN937599 | R2 | Lepidurus couesii isolate C2_8 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | JN937598 | R2 | Lepidurus couesii isolate C2_4 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | JN937597 | R2 | Lepidurus couesii isolate C2_9 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | JN937596 | R2 | Lepidurus couesii isolate C2_7 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | JN937595 | R2 | Lepidurus couesii isolate C2_6 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | JN937594 | R2 | Lepidurus couesii isolate C2_3 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | JN937593 | R2 | Lepidurus couesii isolate C2_2 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | JN937592 | R2 | Lepidurus couesii isolate C2_1 28S ribosomal RNA gene, partial sequence; and non-LTR retrotransposon R2LcA, complete sequence | Lepidurus couesii |
| R2 | AF015822 | R2 ORF | Tenebrio molitor retrotransposon R2 reverse transcriptase gene, partial cds | Tenebrio molitor |
| R2 | AF015817 | R2 ORF | Tenebrio molitor retrotransposon R2 reverse transcriptase gene, partial cds | Tenebrio molitor |
| R2 | AF015816 | R2 ORF | Hippodamia convergens retrotransposon R2 reverse transcriptase gene, partial cds | Hippodamia convergens |
| R2 | KP657866 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-714).5 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657865 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-714).4 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657863 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-714).2 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657862 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-714).1 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657861 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-1297).9 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657860 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-1297).8 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657859 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-1297).7 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657858 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-1297).6 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657857 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-1297).5 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657856 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-1297).4 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657855 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-1297).3 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657854 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-1297).2 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP657853 | retrotransposon:R2Br | Bacillus rossius isolate roCAP(-1297).10 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | KP65 7852 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1297).1 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7851 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).9 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7850 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).8 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7849 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).7 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7848 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).6 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7847 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).5 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7846 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).4 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7845 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).3 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7844 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).2 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7843 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).10 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7842 | retrotransposon:R2 Br | Bacillus rossius isolate roCAP(-1172).1 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7824 | retrotransposon:R2 Br | Bacillus rossius isolate roANZ(-1062).5 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7823 | retrotransposon:R2 Br | Bacillus rossius isolate roANZ(-1062).4 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7822 | retrotransposon:R2 Br | Bacillus rossius isolate roANZ(-1062).3 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7820 | retrotransposon:R2 Br | Bacillus rossius isolate roANZ(-1062).2 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7816 | retrotransposon:R2 Br | Bacillus rossius isolate roANZ(-1062).16 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7814 | retrotransposon:R2 Br | Bacillus rossius isolate roANZ(-1062).14 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7810 | retrotransposon:R2 Br | Bacillus rossius isolate roANZ(-1062).10 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7809 | retrotransposon:R2 Br | Bacillus rossius isolate roANZ(-1062).1 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7759 | retrotransposon:R2 Br | Bacillus rossius isolate rePAT(-1297).8 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7757 | retrotransposon:R2 Br | Bacillus rossius isolate rePAT(-1297).6 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |
| R2 | KP65 7751 | retrotransposon:R2 Br | Bacillus rossius isolate rePAT(-1297).1 retrotransposon R2Br reverse transcriptase gene, partial cds | Bacillus rossius |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | AB097125 | rt | Ciona savignyi retrotransposon R2Cs-D DNA, partial sequence | Ciona savignyi |
| R2 | AB097124 | rt | Ciona intestinalis retrotransposon R2Ci-D DNA, partial sequence | Ciona intestinalis |
| R2 | AB097123 | rt | Ciona intestinalis retrotransposon R2Ci-C DNA, partial sequence | Ciona intestinalis |
| R2 | AB097121 | rt | Ciona intestinalis retrotransposon R2Ci-A DNA, complete sequence | Ciona intestinalis |
| R2 | AB201417 | rt | Triops longicaudatus non-LTR retrotransposon R2Tl gene for reverse transcriptase, partial cds | Triops longicaudatus |
| R2 | AB201416 | rt | Procambarus clarkii non-LTR retrotransposon R2Pc gene for reverse transcriptase, partial cds | Procambarus clarkii |
| R2 | AB201415 | rt | Hasarius adansoni non-LTR retrotransposon R2Ha gene for reverse transcriptase, partial cds | Hasarius adansoni |
| R2 | AB201414 | rt | Metacrinus rotundus non-LTR retrotransposon R2Mr gene for reverse transcriptase, partial cds | Metacrinus rotundus |
| R2 | AB201413 | rt | Mauremys reevesii non-LTR retrotransposon R2Cr-B2 gene for reverse transcriptase, partial cds | Mauremys reevesii |
| R2 | AB201412 | rt | Mauremys reevesii non-LTR retrotransposon R2Cr-B1 gene for reverse transcriptase, partial cds | Mauremys reevesii |
| R2 | AB201411 | rt | Mauremys reevesii non-LTR retrotransposon R2Cr-A gene for reverse transcriptase, partial cds | Mauremys reevesii |
| R2 | AB201410 | rt | Oryzias latipes non-LTR retrotransposon R2Ol-A gene for reverse transcriptase, partial cds | Oryzias latipes |
| R2 | AB201409 | rt | Tanichthys albonubes non-LTR retrotransposon R2Ta gene for reverse transcriptase, partial cds | Tanichthys albonubes |
| R2 | AB201408 | rt | Eptatretus burgeri non-LTR retrotransposon R2Eb gene for reverse transcriptase, partial cds | Eptatretus burgeri |
| R2 | DQ099732 | transposon:R2-like non-LTR retrotransposon R2Ag | Aedes aegypti transposon R2-like non-LTR retrotransposon R2Ag reverse transcriptase (pol) gene, partial cds | Aedes aegypti |
| R2 | DQ099728 | transposon:R2-like non-LTR retrotransposon R2Ag_B | Aedes aegypti transposon R2-like non-LTR retrotransposon R2Ag_B reverse transcriptase (pol) gene, partial cds | Aedes aegypti |
| R2 | GU949559 | | Kalotermes flavicollis isolate Livorno non-LTR retrotransposon R2, complete sequence; and R2 protein gene, complete cds | Kalotermes flavicollis |
| R2 | GU949557 | | Reticulitermes balkanensis non-LTR retrotransposon R2, partial sequence; and R2 protein gene, partial cds | Reticulitermes balkanensis |
| R2 | GU949556 | | Reticulitermes grassei non-LTR retrotransposon R2, partial sequence; and R2 protein gene, partial cds | Reticulitermes grassei |
| R2 | GU949554 | | Reticulitermes urbis non-LTR retrotransposon R2, complete sequence; and R2 protein gene, complete cds | Reticulitermes urbis |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | AF412214 | | Schistosoma japonicum clone S10A non-LTR retrotransposon SjR2-like, partial sequence | Schistosoma japonicum |
| R2 | AF015685 | | Drosophila mercatorum R2 retrotransposon reverse transcriptase domain protein gene, complete cds | Drosophila mercatorum |
| R2 | KJ958674 | | Bacillus rossius retrotransposon R2Br, complete sequence | Bacillus rossius |
| R2 | AF015814 | R2 | Limulus polyphemus retrotransposon R2, complete sequence | Limulus polyphemus |
| R2 | M16558 | R2 | Bombyx mori rDNA insertion element R2 (typeII), complete cds. | Bombyx mori |
| R2 | GQ398057 | R9Av | Adineta vaga copy 1 non-LTR retrotransposon R9, complete sequence; and disrupted 28S ribosomal RNA gene, partial sequence | Adineta vaga |
| R2Bm | AB076841 | R2 | Bombyx mori non-LTR retrotransposon R2Bm gene for reverse transcriptase, complete cds and 28S rRNA | Bombyx mori |
| R2Ci-B | AB097122 | rt | Ciona intestinalis retrotransposon R2Ci-B DNA, complete sequence | Ciona intestinalis |
| R2Dr | AB097126 | rt | Danio rerio retrotransposon R2Dr DNA, complete sequence | Danio rerio |
| R4 | AH003588 | | Parascaris equorum transposon non-LTR retrotransposable element R4 reverse transcriptase gene, partial cds | Parascaris equorum |
| R4 | ALU29445 | R4 | Ascaris lumbricoides site-specific non-LTR retrotransposable element R4 in 26S rDNA, complete sequence | Ascaris lumbricoides |
| R4 | L08889 | R4 Dong | Bombyx mori reverse transcriptase gene, complete cds | Bombyx mori |
| R4 | DQ836390 | MalR4-5 | Maculinea alcon R4-like non-LTR retrotransposon R4-5 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris alcon |
| R4 | DQ836385 | MnaR4-3 | Maculinea nausithous R4-like non-LTR retrotransposon R4-3 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris nausithous |
| R4 | DQ836386 | MnaR4-4 | Maculinea nausithous R4-like non-LTR retrotransposon R4-4 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris nausithous |
| R4 | DQ836387 | MnaR4-7 | Maculinea nausithous R4-like non-LTR retrotransposon R4-7 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris nausithous |
| R4 | DQ836388 | MnaR4-8 | Maculinea nausithous R4-like non-LTR retrotransposon R4-8 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris nausithous |
| R4 | DQ836389 | MnaR4-9 | Maculinea nausithous R4-like non-LTR retrotransposon R4-9 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris nausithous |

| Table 2 | | | | |
|---|---|---|---|---|
| R4 | DQ836379 | MteR4-1 | Maculinea teleius R4-like non-LTR retrotransposon R4-1 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris teleius |
| R4 | DQ836384 | MteR4-10 | Maculinea teleius R4-like non-LTR retrotransposon R4-10 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris teleius |
| R4 | DQ836367 | MteR4-2 | Maculinea teleius R4-like non-LTR retrotransposon R4-2 reverse transcriptase (RT) gene, partial cds | Phengaris teleius |
| R4 | DQ836380 | MteR4-3 | Maculinea teleius R4-like non-LTR retrotransposon R4-3 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris teleius |
| R4 | DQ836381 | MteR4-4 | Maculinea teleius R4-like non-LTR retrotransposon R4-4 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris teleius |
| R4 | DQ836382 | MteR4-6 | Maculinea teleius R4-like non-LTR retrotransposon R4-6 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris teleius |
| R4 | DQ836383 | MteR4-8 | Maculinea teleius R4-like non-LTR retrotransposon R4-8 reverse transcriptase (RT) pseudogene, partial sequence | Phengaris teleius |
| R4 | DQ836374 | transposon:R4-like non-LTR retrotransposon R4-1 | Maculinea alcon R4-like non-LTR retrotransposon R4-1 reverse transcriptase (RT) gene, partial cds | Phengaris alcon |
| R4 | DQ836373 | transposon:R4-like non-LTR retrotransposon R4-1 | Maculinea nausithous R4-like non-LTR retrotransposon R4-1 reverse transcriptase (RT) gene, partial cds | Phengaris nausithous |
| R4 | DQ836375 | transposon:R4-like non-LTR retrotransposon R4-2 | Maculinea alcon R4-like non-LTR retrotransposon R4-2 reverse transcriptase (RT) gene, partial cds | Phengaris alcon |
| R4 | DQ836376 | transposon:R4-like non-LTR retrotransposon R4-3 | Maculinea alcon R4-like non-LTR retrotransposon R4-3 reverse transcriptase (RT) gene, partial cds | Phengaris alcon |
| R4 | DQ836377 | transposon:R4-like non-LTR retrotransposon R4-4 | Maculinea alcon R4-like non-LTR retrotransposon R4-4 reverse transcriptase (RT) gene, partial cds | Phengaris alcon |
| R4 | DQ836371 | transposon:R4-like non-LTR retrotransposon R4-5 | Maculinea nausithous R4-like non-LTR retrotransposon R4-5 reverse transcriptase (RT) gene, partial cds | Phengaris nausithous |
| R4 | DQ836368 | transposon:R4-like non-LTR retrotransposon R4-5 | Maculinea teleius R4-like non-LTR retrotransposon R4-5 reverse transcriptase (RT) gene, partial cds | Phengaris teleius |
| R4 | DQ836378 | transposon:R4-like non-LTR | Maculinea alcon R4-like non-LTR retrotransposon R4-6 reverse transcriptase (RT) gene, partial cds | Phengaris alcon |

| Table 2 | | | | |
|---|---|---|---|---|
| | | retrotransposon R4-6 | | |
| R4 | DQ836372 | transposon:R4-like non-LTR retrotransposon R4-6 | Maculinea nausithous R4-like non-LTR retrotransposon R4-6 reverse transcriptase (RT) gene, partial cds | Phengaris nausithous |
| R4 | DQ836369 | transposon:R4-like non-LTR retrotransposon R4-7 | Maculinea teleius R4-like non-LTR retrotransposon R4-7 reverse transcriptase (RT) gene, partial cds | Phengaris teleius |
| R4 | DQ836370 | transposon:R4-like non-LTR retrotransposon R4-9 | Maculinea teleius R4-like non-LTR retrotransposon R4-9 reverse transcriptase (RT) gene, partial cds | Phengaris teleius |
| R4 | AF286191 | transposon:retrotransposon Rex6 | Xiphophorus maculatus retrotransposon Rex6 reverse transcriptase pseudogene, partial sequence | Xiphophorus maculatus |
| R5 | KJ958673 | R2 | Bacillus rossius non-LTR retrotransposon reCUR5_deg, partial sequence | Bacillus rossius |
| R5 | KJ958620 | R2 | Bacillus rossius non-LTR retrotransposon trKOR5, partial sequence | Bacillus rossius |
| R5 | KJ958614 | R2 | Bacillus rossius non-LTR retrotransposon reVIR5, partial sequence | Bacillus rossius |
| R5 | KJ958595 | R2 | Bacillus rossius non-LTR retrotransposon reBER5, partial sequence | Bacillus rossius |
| R5 | AJ006560 | transposon:Amer5 non-LTR retrotransposon | Anopheles merus Amer5 non-LTR retrotransposon encoding reverse transcriptase, partial | Anopheles merus |
| R5 | AF352479 | transposon:NLRCth1-like non-LTR retrotransposon | Chironomus circumdatus clone cir5 transposon NLRCth1-like non-LTR retrotransposon reverse transcriptase gene, partial cds | Chironomus circumdatus |
| R5 | AF352454 | transposon:NLRCth1-like non-LTR retrotransposon | Chironomus alpestris clone dor50 transposon NLRCth1-like non-LTR retrotransposon reverse transcriptase gene, partial cds | Chironomus alpestris |
| R5 | AF352404 | transposon:NLRCth1-like non-LTR retrotransposon | Chironomus luridus clone lur5 transposon NLRCth1-like non-LTR retrotransposon reverse transcriptase gene, partial cds | Chironomus luridus |
| R8 | FR852798 | poly(A) tail | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline2_3 | Beta vulgaris subsp. vulgaris |
| R8 | KJ958630 | R2 | Bacillus rossius non-LTR retrotransposon reVIR8_deg, partial sequence | Bacillus rossius |
| R8 | KJ958621 | R2 | Bacillus rossius non-LTR retrotransposon trKOR8, partial sequence | Bacillus rossius |
| R8 | KP001560 | rex3-RT pseudogene_ContigILU_TR8 | Iberochondrostoma lusitanicum clone tr8a non-LTR retrotransposon Rex3, complete sequence | Iberochondrostoma lusitanicum |
| R8 | FR852885 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR114 (Belline1_114) | Beta vulgaris |

| Table 2 | | | | |
|---|---|---|---|---|
| | | | | subsp. vulgaris |
| R8 | FR85 2856 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR45 (Belline1_45) | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2844 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR22 (Belline1_22) | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2836 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline17_6 | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2834 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline17_4 | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2831 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline17_1 | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2829 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline16_2 | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2827 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline15_3 | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2819 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline12_2 | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2813 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline9_5 | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2807 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline8_1 | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2806 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline7_18 | Beta vulgaris subsp. vulgaris |
| R8 | FR85 2799 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon Belline2_4 | Beta vulgaris subsp. vulgaris |

| Table 2 | | | | |
|---|---|---|---|---|
| R8 | FR852795 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR19 (Belline1_19) | Beta vulgaris subsp. vulgaris |
| R8 | AF352481 | transposon:NLRCth1-like non-LTR retrotransposon | Chironomus circumdatus clone cir8 transposon NLRCth1-like non-LTR retrotransposon reverse transcriptase gene, partial cds | Chironomus circumdatus |
| R8;R5 | FR852861 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR59 (Belline1_59) | Beta vulgaris subsp. vulgaris |
| R8;R5 | FR852857 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR51 (Belline1_51) | Beta vulgaris subsp. vulgaris |
| R8;R7 | FR852866 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR76 (Belline1_76) | Beta vulgaris subsp. vulgaris |
| R8;R7 | FR852838 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR7 (Belline1_7) | Beta vulgaris subsp. vulgaris |
| R8;R7 | AF352455 | transposon:NLRCth1-like non-LTR retrotransposon | Chironomus alpestris clone dor70 note identical sequence found in dor80 transposon NLRCth1-like non-LTR retrotransposon reverse transcriptase gene, partial cds | Chironomus alpestris |
| R8;R9 | FR852878 | right terminal repeat | Beta vulgaris subsp. vulgaris LINE-type retrotransposon BNR96 (Belline1_96) | Beta vulgaris subsp. vulgaris |
| R9 | KJ958625 | R2 | Bacillus rossius non-LTR retrotransposon trKOR9, partial sequence | Bacillus rossius |
| Rex6 | AJ293547 | en | Oreochromis niloticus Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Oni-3 | Oreochromis niloticus |
| Rex6 | AJ293546 | en | Oreochromis niloticus Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Oni-2 | Oreochromis niloticus |
| Rex6 | AJ293545 | en | Oreochromis niloticus Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Oni-1 | Oreochromis niloticus |
| Rex6 | AJ293517 | en | Xiphophorus maculatus Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Xma-6 | Xiphophorus maculatus |
| Rex6 | AJ293516 | en | Xiphophorus maculatus Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Xma-5 | Xiphophorus maculatus |
| Rex6 | AJ293515 | en | Xiphophorus maculatus Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Xma-4 | Xiphophorus maculatus |
| Rex6 | AJ293514 | en | Xiphophorus maculatus Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Xma-3 | Xiphophorus maculatus |
| Rex6 | AJ293513 | en | Xiphophorus maculatus Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Xma-2 | Xiphophorus maculatus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | AJ293512 | en | Xiphophorus maculatus Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Xma-1 | Xiphophorus maculatus |
| Rex6 | AJ293538 | en | Poecilia formosa Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pfo-6 | Poecilia formosa |
| Rex6 | AJ293537 | en | Poecilia formosa Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pfo-5 | Poecilia formosa |
| Rex6 | AJ293536 | en | Poecilia formosa Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pfo-4 | Poecilia formosa |
| Rex6 | AJ293535 | en | Poecilia formosa Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pfo-3 | Poecilia formosa |
| Rex6 | AJ293534 | en | Poecilia formosa Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pfo-2 | Poecilia formosa |
| Rex6 | AJ293533 | en | Poecilia formosa Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pfo-1 | Poecilia formosa |
| Rex6 | AJ293526 | en | Poeciliopsis gracilis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pgr-4 | Poeciliopsis gracilis |
| Rex6 | AJ293525 | en | Poeciliopsis gracilis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pgr-3 | Poeciliopsis gracilis |
| Rex6 | AJ293524 | en | Poeciliopsis gracilis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pgr-2 | Poeciliopsis gracilis |
| Rex6 | AJ293523 | en | Poeciliopsis gracilis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Pgr-1 | Poeciliopsis gracilis |
| Rex6 | AJ293522 | en | Oryzias latipes Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Ola-5 | Oryzias latipes |
| Rex6 | AJ293521 | en | Oryzias latipes Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Ola-4 | Oryzias latipes |
| Rex6 | AJ293520 | en | Oryzias latipes Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Ola-3 | Oryzias latipes |
| Rex6 | AJ293519 | en | Oryzias latipes Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Ola-2 | Oryzias latipes |
| Rex6 | AJ293518 | en | Oryzias latipes Rex6 retrotransposon partial en pseudogene for endonuclease, clone Rex6-Ola-1 | Oryzias latipes |
| Rex6 | AJ293549 | en | Cichlasoma labridens Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Cla-2 | Herichthys labridens |
| Rex6 | AJ293548 | en | Cichlasoma labridens Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Cla-1 | Herichthys labridens |
| Rex6 | AJ293544 | en | Heterandria bimaculata Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Hbi-6 | Pseudoxiphophorus bimaculatus |
| Rex6 | AJ293543 | en | Heterandria bimaculata Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Hbi-5 | Pseudoxiphophorus bimaculatus |
| Rex6 | AJ293542 | en | Heterandria bimaculata Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Hbi-4 | Pseudoxiphophorus bimaculatus |
| Rex6 | AJ293541 | en | Heterandria bimaculata Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Hbi-3 | Pseudoxiphophorus bimaculatus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | AJ293540 | en | Heterandria bimaculata Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Hbi-2 | Pseudoxiphophorus bimaculatus |
| Rex6 | AJ293539 | en | Heterandria bimaculata Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Hbi-1 | Pseudoxiphophorus bimaculatus |
| Rex6 | AJ293532 | en | Gambusia affinis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Gaf-5 | Gambusia affinis |
| Rex6 | AJ293531 | en | Gambusia affinis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Gaf-5 | Gambusia affinis |
| Rex6 | AJ293530 | en | Gambusia affinis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Gaf-4 | Gambusia affinis |
| Rex6 | AJ293529 | en | Gambusia affinis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Gaf-3 | Gambusia affinis |
| Rex6 | AJ293528 | en | Gambusia affinis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Gaf-2 | Gambusia affinis |
| Rex6 | AJ293527 | en | Gambusia affinis Rex6 retrotransposon partial en pseudogene for endonuclease, clone rex6-Gaf-1 | Gambusia affinis |
| Rex6 | JX576459 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate i non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576458 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate h non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576457 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate g non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576456 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate f non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576455 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate e non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576454 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate d non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576453 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate c non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576452 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate b non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576451 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate a non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576450 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate z8 non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | JX576449 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate z7 non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576448 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate z6 non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576447 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate z5 non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576446 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate z4 non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576445 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate z3 non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576444 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate z2 non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576443 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate z1 non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576442 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate z non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576441 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate x non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576440 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate v non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576439 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate u non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576438 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate t non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576437 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate s non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576436 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate r non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576435 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate q non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576434 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate p non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | JX576433 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate o non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576432 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate n non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576431 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate m non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576430 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate l non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576429 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate k non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576428 | non-LTR retrotransposon:Rex6 | Symphysodon discus isolate j non-LTR retrotransposon Rex6, partial sequence | Symphysodon discus |
| Rex6 | JX576427 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone e non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | JX576426 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone d non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | JX576425 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone c non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | JX576424 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone b non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | JX576423 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone a non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | JX576422 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone g non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | JX576421 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone f non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | JX576420 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone e non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | JX576419 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone d non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | JX576418 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone c non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | JX576417 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone b non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | JX576416 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone a non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | JX576415 | non-LTR retrotransposon:Rex6 | Astronotus ocellatus clone g non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | JX576414 | non-LTR retrotransposon:Rex6 | Astronotus ocellatus clone f non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | JX576413 | non-LTR retrotransposon:Rex6 | Astronotus ocellatus clone e non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | JX576412 | non-LTR retrotransposon:Rex6 | Astronotus ocellatus clone d non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | JX576411 | non-LTR retrotransposon:Rex6 | Astronotus ocellatus clone c non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | JX576410 | non-LTR retrotransposon:Rex6 | Astronotus ocellatus clone b non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | JX576409 | non-LTR retrotransposon:Rex6 | Astronotus ocellatus clone a non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | JX576408 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone h non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | JX576407 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone g non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | JX576406 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone f non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | JX576405 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone e non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | JX576404 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone d non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | JX576403 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone c non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | JX576402 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone b non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | JX576401 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone a non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131853 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone z7 non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131852 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone z6 non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131851 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone z5 non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131850 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone z4 non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131849 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone z3 non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131848 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone z2 non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131847 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone z1 non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131846 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone z non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131845 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone x non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131844 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone v non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131843 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone u non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131842 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone t non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131841 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone s non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131840 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone r non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |
| Rex6 | KF131839 | non-LTR retrotransposon:Rex6 | Pterophyllum scalare clone q non-LTR retrotransposon Rex6, partial sequence | Pterophyllum scalare |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | KF13 1838 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone p non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1837 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone n non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1836 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone m non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1835 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone l non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1834 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone k non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1833 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone j non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1832 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone i non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1831 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone h non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1830 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone g non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1829 | non-LTR retrotransposon:Rex 6 | Pterophyllum scalare clone f non-LTR retrotransposon Rex6, partial sequence | Pterophyllu m scalare |
| Rex6 | KF13 1828 | non-LTR retrotransposon:Rex 6 | Geophagus proximus clone z6 non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF13 1827 | non-LTR retrotransposon:Rex 6 | Geophagus proximus clone z5 non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF13 1826 | non-LTR retrotransposon:Rex 6 | Geophagus proximus clone z4 non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF13 1825 | non-LTR retrotransposon:Rex 6 | Geophagus proximus clone z3 non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF13 1824 | non-LTR retrotransposon:Rex 6 | Geophagus proximus clone z2 non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF13 1823 | non-LTR retrotransposon:Rex 6 | Geophagus proximus clone z1 non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | KF131822 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone z non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131821 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone x non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131820 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone v non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131819 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone u non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131818 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone t non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131817 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone s non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131816 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone r non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131815 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone q non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131814 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone p non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131813 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone o non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131812 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone n non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131811 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone m non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131810 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone l non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131809 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone k non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131808 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone j non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF131807 | non-LTR retrotransposon:Rex6 | Geophagus proximus clone i non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | KF13 1806 | non-LTR retrotransposon:Rex 6 | Geophagus proximus clone h non-LTR retrotransposon Rex6, partial sequence | Geophagus proximus |
| Rex6 | KF13 1805 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z10 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1804 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z9 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1803 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z8 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1802 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z7 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1801 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z6 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1800 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z5 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1799 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z4 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1798 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z3 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1797 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z2 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1796 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z1 non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1795 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone z non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1794 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone x non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1793 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone v non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1792 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone u non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1791 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone t non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | KF13 1790 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone s non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1789 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone r non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1788 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone q non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1787 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone p non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1786 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone o non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1785 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone n non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1784 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone m non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1783 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone l non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1782 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone k non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1781 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone j non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1780 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone i non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1779 | non-LTR retrotransposon:Rex 6 | Astronotus ocellatus clone h non-LTR retrotransposon Rex6, partial sequence | Astronotus ocellatus |
| Rex6 | KF13 1778 | non-LTR retrotransposon:Rex 6 | Cichla monoculus clone z6 non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF13 1777 | non-LTR retrotransposon:Rex 6 | Cichla monoculus clone z5 non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF13 1776 | non-LTR retrotransposon:Rex 6 | Cichla monoculus clone z4 non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF13 1775 | non-LTR retrotransposon:Rex 6 | Cichla monoculus clone z3 non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | KF131774 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone z2 non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131773 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone z1 non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131772 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone z non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131771 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone x non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131770 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone v non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131769 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone u non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131768 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone t non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131767 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone s non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131766 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone r non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131765 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone q non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131764 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone p non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131763 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone o non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131762 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone n non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131761 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone m non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131760 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone l non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF131759 | non-LTR retrotransposon:Rex6 | Cichla monoculus clone k non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |

| Table 2 | | | | |
|---|---|---|---|---|
| Rex6 | KF13 1758 | non-LTR retrotransposon:Rex 6 | Cichla monoculus clone j non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| Rex6 | KF13 1757 | non-LTR retrotransposon:Rex 6 | Cichla monoculus clone i non-LTR retrotransposon Rex6, partial sequence | Cichla monoculus |
| SLACS | JN60 8782 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-46 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8781 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-45 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8780 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-41 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8779 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8778 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-38 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8777 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-37 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8776 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-36 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8775 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8774 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-34 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8773 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-33 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8772 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8771 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-30 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8770 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-29 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN60 8769 | non-LTR retrotransposon:SLA CS-like | Silene latifolia isolate Y-28 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608768 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-27 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608767 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-26 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608766 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-25 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608765 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-24 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608764 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608763 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-22 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608762 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-20 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608761 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-19 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608760 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-18 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608759 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-17 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608758 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-16 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608757 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608756 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-12 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608755 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608754 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-10 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608753 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-08 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608752 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-07 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608751 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-06 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608750 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608749 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-04 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608748 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608747 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate Y-01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608746 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-83 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608745 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-81 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608744 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-80 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608743 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-79 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608742 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-78 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608741 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-76 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608740 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-75 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608739 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-74 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608738 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-66 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608737 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-65 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608736 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-62 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608735 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-61 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608734 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-60 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608733 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-59 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608732 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-57 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608731 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-49 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608730 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-41 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608729 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-28 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608728 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-27 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608727 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608726 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-21 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608725 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-14 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608724 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608723 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-07 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608722 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-06 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608721 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608720 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608719 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate X-01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608718 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608717 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG39 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608716 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG38 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608715 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG37 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608714 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG36 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608713 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608712 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG34 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608711 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG33 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608710 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608709 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG31 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608708 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG30 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608707 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG29 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608706 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG28 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608705 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG27 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608704 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG26 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608703 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG25 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608702 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG24 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608701 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608700 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG22 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608699 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG21 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608698 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG20 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608697 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG19 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608696 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG18 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608695 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG17 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608694 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG16 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608693 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG14 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608692 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608691 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG12 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608690 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608689 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG10 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608688 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG09 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608687 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG08 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608686 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG07 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608685 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG06 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608684 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608683 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG04 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608682 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608681 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG02 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608680 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mG01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608679 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608678 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG39 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608677 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG38 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608676 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG37 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608675 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG36 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608674 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608673 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG34 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608672 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG33 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608671 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608670 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG31 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608669 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG30 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608668 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG29 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608667 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG28 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608666 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG27 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608665 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG26 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608664 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG25 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608663 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG24 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608662 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608661 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG22 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608660 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG21 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608659 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG20 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608658 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG19 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608657 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG18 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608656 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG17 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608655 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG16 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608654 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG15 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608653 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG14 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608652 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608651 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608650 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG10 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608649 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG09 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608648 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG08 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608647 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG07 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608646 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG06 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608645 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608644 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608643 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG02 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608642 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fG01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608641 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-47 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608640 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-46 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608639 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-45 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608638 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-44 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608637 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-42 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608636 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-41 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608635 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608634 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-37 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608633 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-36 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608632 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608631 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-34 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608630 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-33 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608629 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608628 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-31 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608627 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-26 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608626 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-24 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608625 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608624 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-22 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608623 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-21 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608622 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-20 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608621 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-19 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608620 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-18 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608619 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-17 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608618 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-16 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608617 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-15 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608616 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-14 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608615 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608614 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-12 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608613 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608612 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-10 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608611 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-09 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608610 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-08 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608609 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-06 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608608 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608607 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608606 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate A-01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608236 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608235 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR39 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608234 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR38 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608233 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608232 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608231 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR29 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608230 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR26 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608229 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR25 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608228 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR18 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608227 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR17 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608226 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR14 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608225 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608224 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR12 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608223 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608222 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR09 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608221 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR08 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608220 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR07 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608219 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR06 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608218 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608217 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR04 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608216 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608215 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR02 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608214 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mR01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608213 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608212 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL39 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608211 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL37 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608210 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL36 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608209 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608208 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL34 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608207 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL33 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608206 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608205 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL31 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608204 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL28 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608203 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL27 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608202 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL25 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608201 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL24 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608200 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608199 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL19 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608198 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL16 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608197 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608196 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL12 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608195 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608194 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL10 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608193 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL07 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608192 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL04 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608191 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608190 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL02 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608189 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mL01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608188 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608187 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF39 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608186 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF38 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608185 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF37 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608184 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF36 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608183 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608182 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF33 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608181 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608180 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF31 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608179 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF30 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608178 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF29 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608177 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF28 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608176 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF27 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608175 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF26 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608174 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF25 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608173 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608172 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF22 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608171 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF21 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608170 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF20 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608169 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF19 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608168 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF18 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608167 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF17 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608166 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF16 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608165 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF15 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608164 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF14 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608163 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608162 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608161 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF10 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608160 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF09 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608159 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF08 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608158 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF07 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608157 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF06 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608156 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608155 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF04 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608154 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608153 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF02 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608152 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate mF01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608151 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608150 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR39 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608149 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR38 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608148 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR37 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608147 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR36 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608146 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608145 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR34 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608144 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR33 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608143 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608142 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR31 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608141 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR30 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608140 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR29 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608139 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR28 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608138 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR27 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608137 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR26 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608136 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR25 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608135 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR24 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608134 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608133 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR22 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608132 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR21 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608131 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR20 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608130 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR19 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608129 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR18 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608128 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR17 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608127 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR16 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608126 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR15 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608125 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR14 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608124 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608123 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR12 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608122 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608121 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR09 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608120 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR08 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608119 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR07 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608118 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608117 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR04 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608116 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608115 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR02 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608114 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fR01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608113 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608112 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL39 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608111 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL38 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608110 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL37 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608109 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL36 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608108 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608107 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL34 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608106 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL33 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608105 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608104 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL31 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608103 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL30 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608102 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL29 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608101 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL28 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608100 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL27 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608099 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL26 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608098 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL25 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608097 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL24 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608096 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608095 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL22 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608094 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL20 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608093 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL19 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608092 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL18 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608091 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL17 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608090 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL16 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608089 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL15 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608088 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL14 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608087 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608086 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL12 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608085 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608084 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL10 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608083 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL09 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608082 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL07 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608081 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL06 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608080 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608079 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL04 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608078 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608077 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fL01 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608076 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF40 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608075 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF39 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608074 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF38 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608073 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF37 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608072 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF36 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608071 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF35 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608070 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF34 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608069 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF33 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608068 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF32 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608067 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF31 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608066 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF30 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608065 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF29 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608064 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF28 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608063 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF27 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608062 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF26 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608061 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF24 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608060 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF23 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608059 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF21 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608058 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF20 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608057 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF18 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608056 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF17 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608055 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF16 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608054 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF15 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608053 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF14 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608052 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF13 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608051 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF12 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608050 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF11 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608049 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF10 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608048 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF09 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |

| Table 2 | | | | |
|---|---|---|---|---|
| SLACS | JN608047 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF08 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608046 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF06 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608045 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF05 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608044 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF04 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608043 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF03 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| SLACS | JN608042 | non-LTR retrotransposon:SLACS-like | Silene latifolia isolate fF02 non-LTR retrotransposon SLACS-like, partial sequence | Silene latifolia |
| YURECi | AB097133 | rt | Ciona intestinalis retrotransposon YURECi DNA, complete sequence | Ciona intestinalis |
| CRE | . | Cnl1 | C. neoformans non-LTR retrotransposon - consensus. | Cryptococcus neoformans |
| CRE | . | CRE-1_ACas | CRE non-LTR retrotransposon: consensus. | Acanthamoeba castellanii |
| CRE | . | Cre-1_BM | Cre-1_BM non-LTR retrotransposon - consensus. | Bombyx mori |
| CRE | . | CRE-1_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | Cre-1_FCy | Cre-1_FCy non-LTR retrotransposon - conceptual consensus. | Fragilariopsis cylindrus |
| CRE | . | Cre-1_HM | Cre-1_HM non-LTR retrotransposon - consensus. | Hydra vulgaris |
| CRE | . | CRE-1_HRo | Cre-like non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| CRE | . | CRE-1_LSa | CRE non-LTR retrotransposon: consensus. | Lactuca sativa |
| CRE | . | Cre-1_MB | Cre-1_MB non-LTR retrotransposon - consensus. | Monosiga brevicollis |
| CRE | . | Cre-1_NV | Cre-1_NV non-LTR retrotransposon - consensus. | Nematostella vectensis |
| CRE | . | CRE-1_PXu | Non-LTR retrotransposon from Papilio xuthus: consensus. | Papilio xuthus |
| CRE | . | CRE-10_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-11_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |

| Table 2 | | | | |
|---|---|---|---|---|
| CRE | . | CRE-12_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-13_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-14_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-15_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-16_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-17_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | Cre-2_BM | Cre-2_BM non-LTR retrotransposon - consensus. | Bombyx mori |
| CRE | . | CRE-2_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-2_HMa | CRE non-LTR retrotransposon: consensus. | Hydra vulgaris |
| CRE | . | CRE-2_HRo | Cre-like non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| CRE | . | CRE-2_NV | CRE non-LTR retrotransposon: consensus. | Nematostella vectensis |
| CRE | . | CRE-2_PXu | Non-LTR retrotransposon from Papilio xuthus: consensus. | Papilio xuthus |
| CRE | . | CRE-3_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-3_HRo | CRE-like non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| CRE | . | CRE-3_NV | CRE non-LTR retrotransposon: consensus. | Nematostella vectensis |
| CRE | . | CRE-4_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-4_HRo | CRE-like non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| CRE | . | CRE-5_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-5_HRo | CRE-like non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| CRE | . | CRE-6_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-6_HRo | CRE-like non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| CRE | . | CRE-7_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-8_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |
| CRE | . | CRE-9_CCri | Non-LTR retrotransposon from the red seaweed: consensus. | Chondrus crispus |

| Table 2 | | | | |
|---|---|---|---|---|
| CRE | M33009 | CRE1 | C.fasciculata retrotransposable element (CRE1). | Crithidia fasciculata |
| CRE | U19151 | CRE2 | C.fasciculata retrotransposable element (CRE2). | Crithidia fasciculata |
| CRE | M62862 | CZAR | T. cruzi SL-RNA-associated non-LTR retrotransposon. | Trypanosoma cruzi |
| R4 | . | Dong | Bombyx mori non-LTR retrotransposable element. | Bombyx mori |
| R4 | . | DONG_FR2 | Non-LTR retrotransposon; site-specific LINE; R4/Dong superfamily; DONG_FR2. | Takifugu rubripes |
| R4 | . | Dong-1_AFC | Dong/R4-type non-LTR retrotransposon - consensus. | Cichlidae |
| R4 | . | Dong-1_HMM | Non-LTR retrotransposon family from Heliconius melpomene melpomene. | Heliconius melpomene melpomene |
| R4 | . | Dong-1_NVe | A Dong non-LTR retrotransposon family from Nematostella vectensis. | Nematostella vectensis |
| R4 | . | Dong-1_PPo | Non-LTR retrotransposon from Papilio polytes: consensus. | Papilio polytes |
| R4 | . | Dong-1_PXu | Non-LTR retrotransposon from Papilio xuthus: consensus. | Papilio xuthus |
| R4 | . | Dong-2_BM | Non-LTR retrotransposon - a consensus. | Bombyx mori |
| R4 | . | Dong-2_HMM | Non-LTR retrotransposon family from Heliconius melpomene melpomene. | Heliconius melpomene melpomene |
| R4 | . | Dong-2_Lch | Dong-like non-LTR retrotransposon - consensus. | Latimeria chalumnae |
| R4 | . | Dong-2_PPo | Non-LTR retrotransposon from Papilio polytes: consensus. | Papilio polytes |
| R4 | . | DongAa | A Dong non-LTR retrotransposon family from Aedes aegypti. | Aedes aegypti |
| R4 | AB097127 | DongAG | Anopheles gambiae non-LTR retrotransposon DongAg - a partial sequence. | Anopheles gambiae |
| R4 | AB097128 | EhRLE2 | Entamoeba histolytica retrotransposon EhRLE2, complete sequence. | Entamoeba histolytica |
| R4 | AB097129 | EhRLE3 | Entamoeba histolytica retrotransposon EhRLE3, complete sequence. | Entamoeba histolytica |
| HERO | . | HERO-1_AFC | Hero-type non-LTR retrotransposon - consensus. | Cichlidae |
| HERO | . | HERO-1_BF | Amphioxus HERO-1_BF autonomous non-LTR Retrotransposon - consensus. | Branchiostoma floridae |
| HERO | . | HERO-1_HR | A family of HERO non-LTR retrotransposons - a consensus sequence. | Helobdella robusta |
| HERO | . | HERO-1_PP | A family of HERO non-LTR retrotransposons - a consensus sequence. | Physarum polycephalum |
| HERO | AAGJ02121261 | HERO-1_SP | Sea urchin HERO-1_SP autonomous non-LTR Retrotransposon - consensus. | Strongylocentrotus purpuratus |

| Table 2 | | | | |
|---|---|---|---|---|
| HERO | . | HERO-2_BF | Amphioxus HERO-2_BF autonomous non-LTR Retrotransposon - consensus. | Branchiostoma floridae |
| HERO | . | HERO-2_DR | HERO-2_DR is a family of HERO non-LTR retrotransposons - a consensus. | Danio rerio |
| HERO | . | HERO-2_HR | Non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| HERO | 048B 05 | Hero-2_SPur | HERO-type non-ltr retrotransposon from sea urchin. | Strongylocentrotus purpuratus |
| HERO | . | HERO-3_BF | HERO-3_BF is a family of HERO non-LTR retrotransposons - a consensus. | Branchiostoma floridae |
| HERO | . | HERO-3_DR | HERO-3_DR is a family of HERO non-LTR retrotransposons - a consensus. | Danio rerio |
| HERO | . | HERO-3_HR | Non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| HERO | . | Hero-3_SPur | HERO-type non-LTR retrotransposon from sea urchin. | Strongylocentrotus purpuratus |
| HERO | . | HERO-4_DR | HERO-4_DR is a family of HERO non-LTR retrotransposons - a consensus. | Danio rerio |
| HERO | . | HERO-4_HR | Non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| HERO | . | HERO-5_HR | Non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| HERO | . | HERO-6_HR | Non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| HERO | . | HERO-7_HR | Non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| HERO | . | HERO-8_HR | Non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| HERO | . | HERO-9_HR | Non-LTR retrotransposon: consensus sequence. | Helobdella robusta |
| HERO | . | HERODr | HERODr is a family of HERO non-LTR retrotransposons - a consensus. | Danio rerio |
| HERO | . | HEROFr | A HERO clade non-LTR Retrotransposon family - consensus. | Takifugu rubripes |
| HERO | . | HEROTn | HEROTn or Zebulon non-LTR retrotransposon - a consensus sequence. | Tetraodon nigroviridis |
| NeSL | . | LIN10B_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN11_SM | Non-LTR retrotransposon: consensus. | Schmidtea mediterranea |
| NeSL | . | LIN13_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |

| Table 2 | | | | |
|---|---|---|---|---|
| NeSL | . | LIN14_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN15_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN2_SM | Non-LTR retrotransposon (consensus). | Schmidtea mediterranea |
| NeSL | . | LIN21_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN23_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN24_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN24B_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN25_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN26_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN3_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN4_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN4b_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | LIN5_SM | Non-LTR retrotransposon from Schmidtea mediterranea: consensus. | Schmidtea mediterranea |
| NeSL | . | LIN6_SM | Non-LTR retrotransposon from Schmidtea mediterranea: consensus. | Schmidtea mediterranea |
| NeSL | . | LIN7_SM | Non-LTR retrotransposon from Schmidtea mediterranea: consensus. | Schmidtea mediterranea |
| NeSL | . | LIN7B_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |

| Table 2 | | | | |
|---|---|---|---|---|
| NeSL | . | LIN9_SM | Non-LTR retrotransposon: consensus. | Schmidtea mediterranea |
| CRE | JQ747487 | MoTeR1 | Telomere-specific non-LTR retrotransposon MoTeR1 from Magnaporthe oryzae. | Magnaporthe oryzae |
| CRE | JQ747488 | MoTeR2 | Telomere-specific non-LTR retrotransposon MoTeR2 from Magnaporthe oryzae. | Magnaporthe oryzae |
| NeSL | Z82058 | NeSL-1 | NeSL-1 is a non-LTR retrotransposon, complete sequence. | Caenorhabditis elegans |
| NeSL | . | NeSL-1_C11 | A family of NeSL non-LTR retrotransposons. | Caenorhabditis tropicalis |
| NeSL | . | NeSL-1_CA | A family of NeSL non-LTR retrotransposons. | Caenorhabditis angaria |
| NeSL | . | NeSL-1_CBre | A family of NeSL non-LTR retrotransposons - consensus. | Caenorhabditis brenneri |
| NeSL | . | NeSL-1_CBri | A family of NeSL non-LTR retrotransposons. | Caenorhabditis briggsae |
| NeSL | . | NeSL-1_CJap | A family of NeSL non-LTR retrotransposons - consensus. | Caenorhabditis japonica |
| NeSL | . | NeSL-1_CRem | A family of NeSL non-LTR retrotransposons - consensus. | Caenorhabditis remanei |
| NeSL | . | NeSL-1_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | NeSL-1_TV | A family of NeSL non-LTR retrotransposons - consensus. | Trichomonas vaginalis |
| NeSL | . | NeSL-2_CBre | A family of NeSL non-LTR retrotransposons - consensus. | Caenorhabditis brenneri |
| NeSL | . | NeSL-2_CRem | A family of NeSL non-LTR retrotransposons - consensus. | Caenorhabditis remanei |
| NeSL | . | NeSL-2_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| NeSL | . | NeSL-3_CBre | A family of NeSL non-LTR retrotransposons - consensus. | Caenorhabditis brenneri |
| NeSL | . | NeSL-3_CRem | A family of NeSL non-LTR retrotransposons - consensus. | Caenorhabditis remanei |
| NeSL | chrUn | NeSL-4_CRem | A family of NeSL non-LTR retrotransposons. | Caenorhabditis remanei |
| NeSL | . | NeSL-4_SM | Non-LTR retrotransposon; consensus. | Schmidtea mediterranea |
| R2 | BN000800 | PERERE-9 | Schistosoma mansoni Perere-9 non-LTR retrotransposon (EST). | Schistosoma mansoni |
| R4 | . | Plat_R4 | R4 Non-LTR Retrotransposon from Ornithorhynchus. | Ornithorhynchus |
| R2 | AF015815 | R2_AM | Anurida maritima retrotransposon R2, complete sequence. | Anurida maritima |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | M16558 | R2_BM | Bombyx mori rDNA insertion element R2 (type II), complete cds. | Bombyx mori |
| R2 | AB097121 | R2_CI | R2-type LINE. | Ciona intestinalis |
| R2 | . | R2_CPB | Non-LTR retrotransposon: consensus. | Chrysemys picta bellii |
| R2 | . | R2_DAn | 28S rDNA-specific non-LTR retrotransposon R2 in Drosophila ananassae. | Drosophila ananassae |
| R2 | X51967 | R2_DM | LINE-like retrotransposable element R2DM. | Drosophila melanogaster |
| R2 | . | R2_DPe | 28S rDNA-specific non-LTR retrotransposon R2 in Drosophila persimilis. | Drosophila persimilis |
| R2 | . | R2_DPs | 28S rDNA-specific non-LTR retrotransposon R2 in Drosophila pseudoobscura. | Drosophila pseudoobscura |
| R2 | . | R2_DSe | 28S rDNA-specific non-LTR retrotransposon R2 in Drosophila sechellia. | Drosophila sechellia |
| R2 | . | R2_DSi | 28S rDNA-specific non-LTR retrotransposon R2 in Drosophila simulans. | Drosophila simulans |
| R2 | . | R2_DYa | 28S rDNA-specific non-LTR retrotransposon R2 in Drosophila yakuba. | Drosophila yakuba |
| R2 | AF015819 | R2_FA | Forficula auricularia retrotransposon R2, complete sequence. | Forficula auricularia |
| R2 | AF015816 | R2_HC | Hippodamia convergens retrotransposon R2 reverse transcriptase gene, partial cds. | Hippodamia convergens |
| R2 | GU949558 | R2_KF | 28S rDNA-specific non-LTR retrotransposon R2 from Kalotermes flavicollis. | Kalotermes flavicollis |
| R2 | AF015814 | R2_LP | Limulus polyphemus retrotransposon R2, complete sequence. | Limulus polyphemus |
| R2 | AF015818 | R2_PS | Porcellio scaber retrotransposon R2, complete sequence. | Porcellio scaber |
| R2 | GU949555 | R2_RL | 28S rDNA-specific non-LTR retrotransposon R2 from Reticulitermes lucifugus. | Reticulitermes lucifugus |
| R2 | GU949554 | R2_RU | 28S rDNA-specific non-LTR retrotransposon R2 from Reticulitermes urbis. | Reticulitermes urbis |
| R2 | . | R2-1_AAm | R2 non-LTR retrotransposon from lone star tick. | Amblyomma americanum |
| R2 | . | R2-1_ACC | R2 non-LTR retrotransposon from golden eagle. | Aquila chrysaetos canadensis |
| R2 | . | R2-1_ACh | R2 non-LTR retrotransposon from rifleman. | Acanthisitta chloris |
| R2 | . | R2-1_AFo | R2 non-LTR retrotransposon from emperor penguin. | Aptenodytes forsteri |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | . | R2-1_AMi | R2-type non-LTR retrotransposon. | Alligator mississippiensis |
| R2 | . | R2-1_AOM | R2 non-LTR retrotransposon from kiwi. | Apteryx spp. |
| R2 | . | R2-1_ApA | R2 non-LTR retrotransposon from north island brown kiwi. | Apteryx australis mantelli |
| R2 | . | R2-1_APi | R2 non-LTR retrotransposon from pea aphid. | Acyrthosiphon pisum |
| R2 | . | R2-1_BRG | R2 non-LTR retrotransposon from East African grey crowned crane. | Balearica regulorum gibbericeps |
| R2 | . | R2-1_BTe | R2 non-LTR retrotransposon from buff-tailed bumblebee. | Bombus terrestris |
| R2 | . | R2-1_CAnn | R2 non-LTR retrotransposon from Anna's hummingbird. | Calypte anna |
| R2 | . | R2-1_CAu | R2 non-LTR retrotransposon from turkey vulture. | Cathartes aura |
| R2 | . | R2-1_CBr | R2 non-LTR retrotransposon from American crow. | Corvus brachyrhynchos |
| R2 | . | R2-1_CCa | R2 non-LTR retrotransposon from chuck-will's-widow. | Antrostomus carolinensis |
| R2 | . | R2-1_CCan | R2 non-LTR retrotransposon from common cuckoo. | Cuculus canorus |
| R2 | . | R2-1_CPu | R2 non-LTR retrotransposon from ruff. | Calidris pugnax |
| R2 | . | R2-1_Crp | Non-LTR retrotransposon. | Crocodylus porosus |
| R2 | . | R2-1_CSt | R2 non-LTR retrotransposon from speckled mousebird. | Colius striatus |
| R2 | . | R2-1_CU | R2 non-LTR retrotransposon from MacQueen's bustard. | Chlamydotis macqueenii |
| R2 | . | R2-1_CVo | R2 non-LTR retrotransposon from killdeer. | Charadrius vociferus |
| R2 | . | R2-1_DWi | 28S rDNA-specific non-LTR retrotransposon R2 in Drosophila willistoni. | Drosophila willistoni |
| R2 | . | R2-1_EGa | R2 non-LTR retrotransposon from little egret. | Egretta garzetta |
| R2 | . | R2-1_FAl | R2 non-LTR retrotransposon from collared flycatcher. | Ficedula albicollis |
| R2 | . | R2-1_FCh | R2 non-LTR retrotransposon from Saker falcon. | Falco cherrug |
| R2 | . | R2-1_FPe | R2 non-LTR retrotransposon from peregrine falcon. | Falco peregrinus |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | . | R2-1_GA | R2 non-LTR retrotransposon from three-spined stickleback. | Gasterosteus aculeatus |
| R2 | . | R2-1_Gav | Non-LTR retrotransposon. | Gavialis gangeticus |
| R2 | . | R2-1_GFo | R2 non-LTR retrotransposon from medium ground finch. | Geospiza fortis |
| R2 | . | R2-1_GSt | R2 non-LTR retrotransposon from red-throated loon. | Gavia stellata |
| R2 | . | R2-1_HAl | R2 non-LTR retrotransposon from white-tailed eagle. | Haliaeetus albicilla |
| R2 | . | R2-1_IS | R2 non-LTR retrotransposon from deer tick. | Ixodes scapularis |
| R2 | . | R2-1_LCh | R2-type non-LTR retrotransposon - consensus. | Latimeria chalumnae |
| R2 | . | R2-1_LDi | R2 non-LTR retrotransposon from cuckoo roller. | Leptosomus discolor |
| R2 | . | R2-1_LSal | non-LTR retrotransposon, consensus. | Lepeophtheirus salmonis |
| R2 | . | R2-1_LV | R2 non-LTR retrotransposon from green sea urchin. | Lytechinus variegatus |
| R2 | . | R2-1_MDe | R2 non-LTR retrotransposon from Hessian fly. | Mayetiola destructor |
| R2 | . | R2-1_MLe | R2 non-LTR retrotransposon - consensus. | Mnemiopsis leidyi |
| R2 | . | R2-1_MR | R2 non-LTR retrotransposon from alfalfa leafcutter bee. | Megachile rotundata |
| R2 | . | R2-1_MUn | R2 non-LTR retrotransposon fragment from budgerigar. | Melopsittacus undulatus |
| R2 | . | R2-1_MUni | R2 non-LTR retrotransposon from brown mesite. | Mesitornis unicolor |
| R2 | . | R2-1_MVi | R2 non-LTR retrotransposon from golden-collared manakin. | Manacus vitellinus |
| R2 | . | R2-1_NNi | R2 non-LTR retrotransposon from created ibis. | Nipponia nippon |
| R2 | . | R2-1_NV | Starlet sea anemone R2-1_NV autonomous Non-LTR Retrotransposon - consensus. | Nematostella vectensis |
| R2 | . | R2-1_OHo | R2 non-LTR retrotransposon from hoatzin. | Opisthocomus hoazin |
| R2 | . | R2-1_PAd | R2 non-LTR retrotransposon from Adelie penguin. | Pygoscelis adeliae |
| R2 | . | R2-1_PBa | R2 non-LTR retrotransposon from red harvester ant. | Pogonomyrmex barbatus |
| R2 | . | R2-1_PCar | R2 non-LTR retrotransposon from great cormorant. | Phalacrocorax carbo |
| R2 | . | R2-1_PCau | R2 non-LTR retrotransposon sequence. | Priapulus caudatus |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | . | R2-1_PCr | R2 non-LTR retrotransposon from great crested grebe. | Podiceps cristatus |
| R2 | . | R2-1_PCri | R2 non-LTR retrotransposon from Dalmatian pelican. | Pelecanus crispus |
| R2 | . | R2-1_PGu | R2 non-LTR retrotransposon from sandgrouse. | Pterocles gutturalis |
| R2 | . | R2-1_PLe | R2 non-LTR retrotransposon from tropicbird. | Phaethon lepturus |
| R2 | . | R2-1_PM | R2-1_PM is a family of R2 non-LTR retrotransposons - consensus. | Petromyzon marinus |
| R2 | . | R2-1_PPap | R2 non-LTR retrotransposon from sand fly. | Phlebotomus papatasi |
| R2 | . | R2-1_PPu | R2 non-LTR retrotransposon from downy woodpecker. | Picoides pubescens |
| R2 | . | R2-1_PRR | R2 non-LTR retrotransposon from American flamingo. | Phoenicopterus ruber ruber |
| R2 | . | R2-1_PSi | R2 non-LTR retrotransposon from Chinese soft-shelled turtle. | Pelodiscus sinensis |
| R2 | . | R2-1_RMi | R2 non-LTR retrotransposon from brown tick. | Rhipicephalus microplus |
| R2 | . | R2-1_RPr | R2 non-LTR retrotransposon sequence. | Rhodnius prolixus |
| R2 | . | R2-1_RPu | R2 non-LTR retrotransposon cDNA sequence from brown tick. | Rhipicephalus pulchellus |
| R2 | . | R2-1_SCa | R2 non-LTR retrotransposon from Atlantic canary. | Serinus canaria |
| R2 | . | R2-1_SK | R2 non-LTR retrotransposon from acorn worm. | Saccoglossus kowalevskii |
| R2 | . | R2-1_SM | R2-type retrotransposon from Schmidtea mediterranea: consensus. | Schmidtea mediterranea |
| R2 | . | R2-1_SP | R2 non-LTR retrotransposon from purple sea urchin. | Strongylocentrotus purpuratus |
| R2 | AGKD01072455 | R2-1_SSa | R2-type non-LTR retrotransposon. | Salmo salar |
| R2 | . | R2-1_StC | R2 non-LTR retrotransposon from ostrich. | Struthio camelus australis |
| R2 | . | R2-1_TAl | R2 non-LTR retrotransposon from barn owl. | Tyto alba |
| R2 | . | R2-1_TCas | R2 non-LTR retrotransposon from red flour beetle - consensus -. | Tribolium castaneum |
| R2 | . | R2-1_TG | A family of R2 non-LTR retrotransposons - consensus sequence. | Taeniopygia guttata |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | . | R2-1_TGut | R2 non-LTR retrotransposon from white-throated tinamou. | Tinamus guttatus |
| R2 | . | R2-1_TSP | A family of R2 non-LTR retrotransposons in the Trichinella spiralis genome - a consensus. | Trichinella spiralis |
| R2 | scaffold_6 | R2-1_TUr | R2 non-LTR retrotransposon from twospotted spider mite. | Tetranychus urticae |
| R2 | . | R2-1_XM | R2 non-LTR retrotransposon fragment from Southern platyfish. | Xiphophorus maculatus |
| R2 | . | R2-1_ZA | R2 non-LTR retrotransposon from white-throated sparrow. | Zonotrichia albicollis |
| R2 | . | R2-1_ZLM | R2 non-LTR retrotransposon from silvereye. | Zosterops lateralis melanops |
| R2 | . | R2-2_APi | R2 non-LTR retrotransposon from pea aphid. | Acyrthosiphon pisum |
| R2 | . | R2-2_CCan | R2 non-LTR retrotransposon from common cuckoo. | Cuculus canorus |
| R2 | . | R2-2_CMa | R2 non-LTR retrotransposon from Macqueen's bustard. | Chlamydotis macqueenii |
| R2 | . | R2-2_DWi | 28S rDNA-specific non-LTR retrotransposon R2 in Drosophila willistoni. | Drosophila willistoni |
| R2 | . | R2-2_HAl | R2 non-LTR retrotransposon from white-tailed eagle. | Haliaeetus albicilla |
| R2 | . | R2-2_IS | R2 non-LTR retrotransposon from deer tick. | Ixodes scapularis |
| R2 | . | R2-2_MR | R2 non-LTR retrotransposon from alfalfa leafcutter bee. | Megachile rotundata |
| R2 | . | R2-2_MUn | R2 non-LTR retrotransposon fragment from budgerigar. | Melopsittacus undulatus |
| R2 | . | R2-2_MUni | R2 non-LTR retrotransposon from brown mesite. | Mesitornis unicolor |
| R2 | . | R2-2_NNi | R2 non-LTR retrotransposon from created ibis. | Nipponia nippon |
| R2 | . | R2-2_NV | Starlet sea anemone R2-2_NV autonomous Non-LTR Retrotransposon - consensus. | Nematostella vectensis |
| R2 | . | R2-2_PBa | R2 non-LTR retrotransposon from red harvester ant. | Pogonomyrmex barbatus |
| R2 | . | R2-2_PM | R2-2_PM is a family of R2 non-LTR retrotransposons - a consensus. | Petromyzon marinus |
| R2 | . | R2-2_RPr | R2 non-LTR retrotransposon sequence. | Rhodnius prolixus |
| R2 | . | R2-2_SMed | R2 non-LTR retrotransposon from Schmidtea mediterranea: consensus. | Schmidtea mediterranea |
| R2 | . | R2-2_TCas | R2 non-LTR retrotransposon from red flour beetle. | Tribolium castaneum |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | scaffold_37 | R2-2_TUr | R2 non-LTR retrotransposon from twospotted spider mite. | Tetranychus urticae |
| R2 | ABJB010555169 | R2-3_IS | R2 non-LTR retrotransposon from deer tick. | Ixodes scapularis |
| R2 | . | R2-3_MR | R2 non-LTR retrotransposon from alfalfa leafcutter bee. | Megachile rotundata |
| R2 | . | R2-4_MR | R2 non-LTR retrotransposon from alfalfa leafcutter bee. | Megachile rotundata |
| R2 | . | R2-5_MR | R2 non-LTR retrotransposon from alfalfa leafcutter bee. | Megachile rotundata |
| R2 | . | R2-6_MR | R2 non-LTR retrotransposon from alfalfa leafcutter bee. | Megachile rotundata |
| R2 | . | R2-7_MR | R2 non-LTR retrotransposon from alfalfa leafcutter bee. | Megachile rotundata |
| R2 | . | R2-8_MR | R2 non-LTR retrotransposon from alfalfa leafcutter bee. | Megachile rotundata |
| R2 | . | R2-N1_Gav | Non-LTR retrotransposon. | Gavialis gangeticus |
| R2 | . | R2-N2_Gav | Non-LTR retrotransposon. | Gavialis gangeticus |
| R2 | . | R2-N2B_Gav | Non-LTR retrotransposon. | Gavialis gangeticus |
| R2 | . | R2A_NVi | 28S rDNA-specific non-LTR retrotransposon R2 in Nasonia vitripennis. | Nasonia vitripennis |
| R2 | AF015817 | R2A_TM | Tenebrio molitor retrotransposon R2 reverse transcriptase gene, partial cds. | Tenebrio molitor |
| R2 | . | R2Amel | R2Amel - R2 non-LTR retrotransposon from the honeybee Apis mellifera. | Apis mellifera |
| R2 | AF015685 | R2B_DM | Drosophila mercatorum R2 retrotransposon reverse transcriptase domain protein gene, complete cds. | Drosophila mercatorum |
| R2 | . | R2B_NVi | 28S rDNA-specific non-LTR retrotransposon R2 in Nasonia vitripennis. | Nasonia vitripennis |
| R2 | AF015822 | R2B_TM | Tenebrio molitor retrotransposon R2 reverse transcriptase gene, partial cds. | Tenebrio molitor |
| R2 | . | R2C_NGi | 28S rDNA-specific non-LTR retrotransposon R2 in Nasonia giraulti. | Nasonia giraulti |
| R2 | AB097122 | R2Ci-B | Ciona intestinalis retrotransposon R2Ci-B, complete sequence. | Ciona intestinalis |
| R2 | . | R2Ci-D | Ciona intestinalis retrotransposon R2CiD, complete sequence. | Ciona intestinalis |
| R2 | AB097121 | R2CIA_CI | Ciona intestinalis retrotransposon R2Ci-A, complete sequence. | Ciona intestinalis |
| R2 | AB097125 | R2Cs-D | Ciona intestinalis retrotransposon R2CsD, partial sequence. | Ciona savignyi |
| R2 | . | R2D_NGi | 28S rDNA-specific non-LTR retrotransposon R2 in Nasonia giraulti. | Nasonia giraulti |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | NM_001003009 7 | R2Dr | R2 non-LTR retrotransposon in the Danio rerio genome - a single copy. | Danio rerio |
| R2 | . | R2E_NLo | 28S rDNA-specific non-LTR retrotransposon R2 in Nasonia longicornisi. | Nasonia longicornis |
| R2 | AB201408 | R2Eb | R2 non-LTR retrotransposon from Eptatretus burgeri. | Eptatretus burgeri |
| R2 | AB201415 | R2Ha | R2 non-LTR retrotransposon from Hasarius adansoni. | Hasarius adansoni |
| R2 | JN937617 | R2La | R2-type non-LTR retrotransposon. | Lepidurus arcticus |
| R2 | . | R2LcA | R2-type non-LTR retrotransposon. | Lepidurus couesii |
| R2 | JN937619 | R2LcB | R2-type non-LTR retrotransposon. | Lepidurus couesii |
| R2 | . | R2LcC | R2-type non-LTR retrotransposon. | Lepidurus couesii |
| R2 | JN937616 | R2Ll | R2-type non-LTR retrotransposon. | Lepidurus apus lubbocki |
| R2 | AB201414 | R2Mr | R2 non-LTR retrotransposon from Metacrinus rotundus. | Metacrinus rotundus |
| R2 | . | R2NS-1_CGi | R2-type retrotransposon from Crassostrea gigas. | Crassostrea gigas |
| R2 | . | R2NS-1_CSi | R2-type retrotransposon from Clonorchis sinensis: consensus. | Clonorchis sinensis |
| R2 | . | R2NS-1_PMi | R2-like non-LTR retrotransposon from bat star. | Patiria miniata |
| R2 | . | R2NS-1_SMed | R2-type retrotransposon from Schmidtea mediterranea: consensus. | Schmidtea mediterranea |
| R2 | . | R2Nvec-A | R2Nvec-A - R2 non-LTR retrotransposon from the starlet sea anemone Nematostella vectensis. | Nematostella vectensis |
| R2 | . | R2Ol-A | R2 non-LTR retrotransposon from the medaka Oryzias latipes - consensus. | Oryzias latipes |
| R2 | AB201416 | R2Pc | R2 non-LTR retrotransposon from Procambarus clarkii. | Procambarus clarkii |
| R2 | . | R2Sm-A | R2Sm-A - R2 non-LTR retrotransposon from the bloodfluke Schistosoma mansoni. | Schistosoma mansoni |
| R2 | AB201409 | R2Ta | R2 non-LTR retrotransposon from Tanichthys albonubes. | Tanichthys albonubes |
| R2 | EU854578 | R2Tc | R2-type non-LTR retrotransposon. | Triops cancriformis |
| R2 | JN937621 | R2Tc_it | R2-type non-LTR retrotransposon. | Triops cancriformis |

| Table 2 | | | | |
|---|---|---|---|---|
| R2 | AB201417 | R2Tl | R2 non-LTR retrotransposon from Triops longicaudatus. | Triops longicaudatus |
| R4 | U29445 | R4_AL | Ascaris lumbricoides site-specific non-LTR retrotransposable element R4 in 26S rDNA, complete sequence. | Ascaris lumbricoides |
| R4 | U29590 | R4_HC | Haemonchus contortus non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. | Haemonchus contortus |
| R4 | . | R4_Hmel | a R4 element from Heliconius melpomene. | Heliconius melpomene |
| R4 | . | R4-1_AC | A family of R4 non-LTR retrotransposons - consensus sequence. | Anolis carolinensis |
| R4 | . | R4-1_ADi | R4-type retrotransposon: consensus. | Acropora digitifera |
| R4 | . | R4-1_BM | Non-LTR retrotransposon - a consensus. | Bombyx mori |
| R4 | CADV01008175 | R4-1_BX | An R4 non-LTR retrotransposon family from Bursaphelenchus xylophilus. | Bursaphelenchus xylophilus |
| R4 | ABLE03011482 | R4-1_CJap | An R4 non-LTR retrotransposon family from Caenorhabditis japonica. | Caenorhabditis japonica |
| R4 | . | R4-1_CM | Non-LTR retrotransposon from the elephant shark - consensus. | Callorhinchus milii |
| R4 | . | R4-1_CPB | Non-LTR retrotransposon: consensus. | Chrysemys picta bellii |
| R4 | . | R4-1_ED | Autonomous non-LTR retrotransposon from the R4 clade - a consensus sequence. | Entamoeba dispar |
| R4 | . | R4-1_HG | An R4 non-LTR retrotransposon family from Heterodera glycines. | Heterodera glycines |
| R4 | . | R4-1_HMe | Non-LTR retrotransposon family from Heliconius melpomene melpomene. | Heliconius melpomene melpomene |
| R4 | CABB01003843 | R4-1_MI | An R4 non-LTR retrotransposon family from Meloidogyne incognita. | Meloidogyne incognita |
| R4 | . | R4-1_PH | Non-LTR Retrotransposon, consensus. | Parhyale hawaiensis |
| R4 | CACX01002001 | R4-1_SRa | An R4 non-LTR retrotransposon family from Strongyloides ratti. | Strongyloides ratti |
| R4 | . | R4-1_TCa | R4-type retrotransposon: consensus. | Tribolium castaneum |
| R4 | . | R4-1B_AC | Dong-type non-LTR retrotransposons - a consensus sequence. | Anolis carolinensis |
| R4 | . | R4-2_AS | An R4 non-LTR retrotransposon family from Ascaris suum. | Ascaris suum |

| Table 2 | | | | |
|---|---|---|---|---|
| R4 | CADV01009048 | R4-2_BX | An R4 non-LTR retrotransposon family from Bursaphelenchus xylophilus. | Bursaphelenchus xylophilus |
| R4 | ABLA01000389 | R4-2_HG | An R4 non-LTR retrotransposon family from Heterodera glycines. | Heterodera glycines |
| R4 | CACX01002006 | R4-2_SRa | An R4 non-LTR retrotransposon family from Strongyloides ratti. | Strongyloides ratti |
| R4 | CADV01008832 | R4-3_BX | An R4 non-LTR retrotransposon family from Bursaphelenchus xylophilus. | Bursaphelenchus xylophilus |
| R4 | . | R4-3_SRa | An R4 non-LTR retrotransposon family from Strongyloides ratti. | Strongyloides ratti |
| R4 | . | R4-4_BX | An R4 non-LTR retrotransposon family from Bursaphelenchus xylophilus. | Bursaphelenchus xylophilus |
| R4 | . | R4-4_SRa | An R4 non-LTR retrotransposon family from Strongyloides ratti. | Strongyloides ratti |
| R4 | . | R4-5_BX | An R4 non-LTR retrotransposon family from Bursaphelenchus xylophilus. | Bursaphelenchus xylophilus |
| NeSL | AY216701 | R5 | Girardia tigrina R5 retrotransposon, complete sequence. | Girardia tigrina |
| NeSL | . | R5-1_SM | A family of planarian NeSL non-LTR retrotransposons - consensus. | Schmidtea mediterranea |
| NeSL | . | R5-2_SM | A family of planarian NeSL non-LTR retrotransposons - consensus. | Schmidtea mediterranea |
| R2 | . | R8Hm-A | R8Hm-A - 18S rDNA-specific non-LTR retrotransposon from Hydra magnipapillata. | Hydra vulgaris |
| R2 | . | R8Hm-B | R8Hm-B - 18S rDNA-specific non-LTR retrotransposon from Hydra magnipapillata. | Hydra vulgaris |
| R2 | . | R9Av | R9Av, an rDNA-specific non-LTR retrotransposon family from rotifer. | Adineta vaga |
| R2 | FJ461304 | RaR2 | 28S rDNA-specific non-LTR retrotransposon R2 from Rhynchosciara americana. | Rhynchosciara americana |
| R4 | . | Rex6 | Non-LTR retrotransposon; site-specific LINE; R4/Dong superfamily; REX6; DONG_FR. | Takifugu rubripes |
| R4 | . | Rex6-1_OL | A Rex6 non-LTR retrotransposon family from Olyzias latipes. | Oryzias latipes |
| CRE | X17078 | SLACS | Trypanosoma brucei DNA for retrotransposable element SLACS. | Trypanosoma brucei |
| NeSL | . | Utopia-1_ACa | Utopia-1_ACa is a protozoan Utopia non-LTR retrotransposon - a complete sequence. | Acanthamoeba castellanii |

| Table 2 | | | | |
|---|---|---|---|---|
| NeSL | scaffold_474 | Utopia-1_ACar | A family of NeSL non-LTR retrotransposons. | Anolis carolinensis |
| NeSL | . | Utopia-1_AEc | A family of Utopia non-LTR retrotransposons - consensus. | Acromyrmex echinatior |
| NeSL | . | Utopia-1_AMi | A family of NeSL non-LTR retrotransposons - consensus. | Alligator mississippiensis |
| NeSL | . | Utopia-1_APi | A family of Utopia non-LTR retrotransposons - consensus. | Acyrthosiphon pisum |
| NeSL | . | Utopia-1_APl | A family of Utopia non-LTR retrotransposons. | Agrilus planipennis |
| NeSL | . | Utopia-1_CFl | A family of Utopia non-LTR retrotransposons - consensus. | Camponotus floridanus |
| NeSL | . | Utopia-1_CMy | A family of Utopia non-LTR retrotransposons - consensus. | Chelonia mydas |
| NeSL | . | Utopia-1_CPB | A family of Utopia non-LTR retrotransposons - consensus. | Chrysemys picta bellii |
| NeSL | . | Utopia-1_Crp | Non-LTR retrotransposon. | Crocodylus porosus |
| NeSL | . | Utopia-1_DPo | A family of Utopia non-LTR retrotransposons. | Dendroctonus ponderosae |
| NeSL | . | Utopia-1_DPu | A family of Utopia non-LTR retrotransposons - consensus. | Daphnia pulex |
| NeSL | . | Utopia-1_DYak | A family of Utopia non-LTR retrotransposons - consensus. | Drosophila yakuba |
| NeSL | . | Utopia-1_EBr | A family of Utopia non-LTR retrotransposons - consensus. | Eimeria brunetti |
| NeSL | . | Utopia-1_EMi | A family of Utopia non-LTR retrotransposons - consensus. | Eimeria mitis |
| NeSL | . | Utopia-1_ENe | A family of Utopia non-LTR retrotransposons - consensus. | Eimeria necatrix |
| NeSL | . | Utopia-1_Gav | Non-LTR retrotransposon. | Gavialis gangeticus |
| NeSL | . | Utopia-1_GG1 | A family of Utopia non-LTR retrotransposons. | Ganaspis |
| NeSL | . | Utopia-1_HAra | A family of Utopia non-LTR retrotransposons. | Hyaloperonospora arabidopsidis |
| NeSL | . | Utopia-1_HG | A family of Utopia non-LTR retrotransposons. | Heterodera glycines |
| NeSL | . | Utopia-1_HMM | A family of Utopia non-LTR retrotransposons. | Heliconius melpomene melpomene |
| NeSL | . | Utopia-1_HSal | A family of Utopia non-LTR retrotransposons - consensus. | Harpegnathos saltator |
| NeSL | . | Utopia-1_IS | A family of Utopia non-LTR retrotransposons - consensus. | Ixodes scapularis |

| Table 2 | | | | |
|---|---|---|---|---|
| NeSL | . | Utopia-1_LAl | A family of Utopia non-LTR retrotransposons. | Lasioglossum albipes |
| NeSL | . | Utopia-1_LFu | A family of Utopia non-LTR retrotransposons. | Ladona fulva |
| NeSL | AGCV01358106 | Utopia-1_LV | A family of Utopia non-LTR retrotransposons. | Lytechinus variegatus |
| NeSL | . | Utopia-1_MRo | A family of Utopia non-LTR retrotransposons - consensus. | Megachile rotundata |
| NeSL | . | Utopia-1_NVit | A family of Utopia non-LTR retrotransposons - consensus. | Nasonia vitripennis |
| NeSL | . | Utopia-1_PAlni | NeSL non-LTR retrotransposon from Phytophthora alni. | Phytophthora alni |
| NeSL | . | Utopia-1_PArrh | NeSL non-LTR retrotransposon from Pythium arrhenomanes. | Pythium arrhenomanes |
| NeSL | . | Utopia-1_PBa | A family of Utopia non-LTR retrotransposons - consensus. | Pogonomyrmex barbatus |
| NeSL | . | Utopia-1_PCa | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora capsici |
| NeSL | . | Utopia-1_PCinn | NeSL non-LTR retrotransposon from Phytophthora cinnamomi. | Phytophthora cinnamomi |
| NeSL | AHJF01004292 | Utopia-1_PCu | A family of Utopia non-LTR retrotransposons - consensus. | Pseudoperonospora cubensis |
| NeSL | . | Utopia-1_PI | A family of NeSL non-LTR retrotransposons - consensus. | Phytophthora infestans |
| NeSL | . | Utopia-1_PInsi | NeSL non-LTR retrotransposon from Pythium insidiosum. | Pythium insidiosum |
| NeSL | . | Utopia-1_PKern | NeSL non-LTR retrotransposon from Phytophthora kernoviae. | Phytophthora kernoviae |
| NeSL | . | Utopia-1_PLate | NeSL non-LTR retrotransposon from Phytophthora lateralis. | Phytophthora lateralis |
| NeSL | . | Utopia-1_PMi | A family of Utopia non-LTR retrotransposons. | Patiria miniata |
| NeSL | . | Utopia-1_PPac | A family of Utopia non-LTR retrotransposons. | Pristionchus pacificus |
| NeSL | . | Utopia-1_PPini | NeSL non-LTR retrotransposon from Phytophthora pinifolia. | Phytophthora pinifolia |
| NeSL | . | Utopia-1_PR | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora ramorum |
| NeSL | . | Utopia-1_PRe | A family of Utopia non-LTR retrotransposons - consensus. | Panagrellus redivivus |
| NeSL | . | Utopia-1_PS | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora sojae |

| Table 2 | | | | |
|---|---|---|---|---|
| NeSL | . | Utopia-1_PSi | A family of Utopia non-LTR retrotransposons - consensus. | Pelodiscus sinensis |
| NeSL | . | Utopia-1_PT | A family of Utopia non-LTR retrotransposons. | Parasteatoda tepidariorum |
| NeSL | ADOS01001321 | Utopia-1_PU | A family of Utopia non-LTR retrotransposons. | Pythium ultimum |
| NeSL | . | Utopia-1_PVexa | NeSL non-LTR retrotransposon from Phytopythium vexans. | Phytopythium aff. vexans |
| NeSL | . | Utopia-1_SaPa | A family of Utopia non-LTR retrotransposons. | Saprolegnia parasitica |
| NeSL | . | Utopia-1_SDicl | NeSL non-LTR retrotransposon from Saprolegnia diclina. | Saprolegnia diclina |
| NeSL | . | Utopia-1_SM | A family of Utopia non-LTR retrotransposons. | Strigamia maritima |
| NeSL | AAGJ02140537 | Utopia-1_SP | A family of Utopia non-LTR retrotransposons. | Strongylocentrotus purpuratus |
| NeSL | . | Utopia-1_TSP | A family of Utopia non-LTR retrotransposons. | Trichinella spiralis |
| NeSL | . | Utopia-1B_CPB | A family of Utopia non-LTR retrotransposons - consensus. | Chrysemys picta bellii |
| NeSL | . | Utopia-2_APi | A family of Utopia non-LTR retrotransposons. | Acyrthosiphon pisum |
| NeSL | . | Utopia-2_CMy | A family of Utopia non-LTR retrotransposons - consensus. | Chelonia mydas |
| NeSL | . | Utopia-2_CPB | A family of Utopia non-LTR retrotransposons - consensus. | Chrysemys picta bellii |
| NeSL | . | Utopia-2_DPu | A family of Utopia non-LTR retrotransposons. | Daphnia pulex |
| NeSL | . | Utopia-2_LFu | A family of Utopia non-LTR retrotransposons. | Ladona fulva |
| NeSL | . | Utopia-2_PCa | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora capsici |
| NeSL | . | Utopia-2_PI | A family of NeSL non-LTR retrotransposons - consensus. | Phytophthora infestans |
| NeSL | . | Utopia-2_PR | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora ramorum |
| NeSL | . | Utopia-2_PS | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora sojae |
| NeSL | . | Utopia-2_PU | A family of Utopia non-LTR retrotransposons. | Pythium ultimum |
| NeSL | . | Utopia-3_CPB | A family of Utopia non-LTR retrotransposons - consensus. | Chrysemys picta bellii |
| NeSL | . | Utopia-3_DPu | A family of Utopia non-LTR retrotransposons. | Daphnia pulex |

| Table 2 | | | | |
|---|---|---|---|---|
| NeSL | . | Utopia-3_LFu | A family of Utopia non-LTR retrotransposons. | Ladona fulva |
| NeSL | . | Utopia-3_PCa | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora capsici |
| NeSL | . | Utopia-3_PI | A family of NeSL non-LTR retrotransposons - consensus. | Phytophthora infestans |
| NeSL | . | Utopia-3_PR | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora ramorum |
| NeSL | . | Utopia-4_LFu | A family of Utopia non-LTR retrotransposons. | Ladona fulva |
| NeSL | AATU01001281.1 | Utopia-4_PI | A family of NeSL non-LTR retrotransposons - a copy. | Phytophthora infestans |
| NeSL | . | Utopia-4_PR | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora ramorum |
| NeSL | . | Utopia-5_LFu | A family of Utopia non-LTR retrotransposons. | Ladona fulva |
| NeSL | . | Utopia-5_PI | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora infestans |
| NeSL | . | Utopia-5_PR | A family of Utopia non-LTR retrotransposons - consensus. | Phytophthora ramorum |
| NeSL | . | Utopia-6_LFu | A family of Utopia non-LTR retrotransposons. | Ladona fulva |
| R4 | . | X4_LINE | Conserved LINE element reconstructed from the human genome - consensus. | Vertebrata |
| NeSL | . | YURE_CSa | A NeSL non-LTR retrotransposon from Ciona savignyi. | Ciona savignyi |
| R2 | . | YURE-2_Cis | YURE non-LTR retrotransposon from Ciona savignyi. | Ciona savignyi |
| NeSL | . | YURECi | Ciona intestinalis retrotransposon YURECi. | Ciona intestinalis |

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.

| 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 |

| | | | | | |
|---|---|---|---|---|---|
| | | | | ACAGAC | MDEIRRHPDKGNGQ |
| | | | | TTTTACT | QRPTKQKSEEQLQAY |
| | | | | ATTTCAT | YKKTLEERLSAGALNT |
| | | | | AACGCG | FPRAFKQVMEGRDIK |
| | | | | TCAATTA | LVINQTAQDCFGCLE |
| | | | | CCACCT | SISQIRTATRDKKDTV |
| | | | | GATTTG | TREKHPKKPFQKWM |
| | | | | GACCAA | KDRAIKKGNYLRFQR |
| | | | | TTCACG | LFYLDRGKLAKIILDDI |
| | | | | GGATTT | ECLSCDIPLSEIYSVFK |
| | | | | GTCCAA | TRWETTGSFKSLGDF |
| | | | | GGTGGA | KTYGKADNTAFRELIT |
| | | | | CGGGCC | AKEIEKNVQEMSKGS |
| | | | | ACCTTTA | APGPDGITLGDVVK |
| | | | | CTTAACC | MDPEFSRTMEIFNL |
| | | | | CGGAAA | WLTTGKIPDMVRGC |
| | | | | AGGAAC | RTVLIPKSSKPPDRLKDI |
| | | | | ATATATA | NNWRPITIGSILLRLF |
| | | | | ATTTATG | SRIVTARLSKACPLNP |
| | | | | TGTGTTC | RQRGFIRAAGCSENL |
| | | | | GATAAA | KLLQTIIWSAKREHRP |
| | | | | (SEQ ID | LGVVFVDIAKAFDTV |
| | | | | NO: | SHQHIIHALQQREVD |
| | | | | 1263) | PHIVGLVSNMYENIS |
| | | | | | TYITTKRNTHTDKIQI |
| | | | | | RVGVKQGDPMSPLL |
| | | | | | FNLAMDPLLCKLEES |
| | | | | | GKGYHRGQSSITAM |
| | | | | | AFADDLVLLSDSWEN |
| | | | | | MNTNISILETFCNLTG |
| | | | | | LKTQGQKCHGFYIKP |
| | | | | | TKDSYTINDCAAWTI |
| | | | | | NGTPLNMIDPGESEK |
| | | | | | YLGLQFDPWIGIARS |
| | | | | | GLSTKLDFWLQRIDQ |
| | | | | | APLKPLQKTDILKTYTI |
| | | | | | PRLIYIADHSEVKTALL |
| | | | | | ETLDQKIRTAVKEWL |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GGTTTGGGCATCACCAAATTGGCAGGACTGATCCCCAGCGTGCAGGCCC | | | | HLPPCTCDAILYSSTR |
| | | | GTAGACTGCATCGGATCGCACAGTCATCTGACGATACGATGAAATGCTT | | | | DGGLGITKLAGLIPSV |
| | | | CATGGAAAAAGAGAAAATGGAACAGCTGCATAAGAAATTGTGGATTCA | | | | QARRLHRIAQSSDDT |
| | | | AGCTGGAGGGGACAGAGAGAACATACCCTCGATTTGGGAAGCACCACC | | | | MKCFMEKEKMEQL |
| | | | GTCGAGTGAACCACCAACAACGTGAGCACAAATTCGGAATGGGAAGC | | | | 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 |
| | | | TGCAAAGGACGTTACTTTTGTGGGCTTTCCCCTAGGAGCGCGGGGGAA | | | | VDVTVRYEAAKSSLE |
| | | | ATGGCACCAAGATAACTTTAAACTTTGACTGAGCTTGGCCTCTCCAAAT | | | | EAAAEKVRKYKHLET |
| | | | CGAGGCAAGTGAAAATGGCAGAGACTTTTTCCACAGTAGCGCTCTTTTC | | | | EVRHLTNAKDVTFVG |
| | | | ATCTGTGGACATTGTACATATGTTTGCCAGTAGGGCCAGAAAATCTATG | | | | FPLGARGKWHQDNF |
| | | | GTTATGTAATTCAGGTTATTTAGATGCTTAGTTTTTGTACCTTTCTTGTTT | | | | KLLTELGLSKSRQVK |
| | | | TGTTTAGGATTTTGATAGTGTTAGTATTTTTATATTTTTGTACGATTGCAT | | | | MAETFSTVALFSSVDI |
| | | | AATGTTCTTTTTTATACAGTTCTGTTTTAATAAAATAGACGATAGCTAGA | | | | VHMFASRARKSMV |
| | | | GACGTTAGGGCAGCCACAAGCCAGTTAGGTAGCGGATAGTAGGTAGGA | | | | M (SEQ ID NO: 1016) |
| | | | ACAGACTTTTACTATTTCATAACGCGTCAATTACCACCTGATTTGGACCA | | | | |
| | | | ATTCACGGGATTTGTCCAAGGTGGACGGGCCACCTTTACTTAACCCGGA | | | | |
| | | | AAAGGAACATATATAATTTATGTGTGTTCGATAAA (SEQ ID NO: 1539) | AGACTTAA | GGTAGA | | VGLCPSPGVDGTHQ |
| R2 | R2-1_Gfo | . | AGACTTAAGTGAGTTTGGTTACAACTGGGCATAGCTGCAGAGACCGCG | GTGAGTTT | TAATCTT | | PNDSFQNFGETNFSV |
| | | Geospiza fortis | CCTCCTCGCGGCCCCGCTGGTAAGCCCTTAACAGGGTGACTAAGTCGGT | GGTTACAA | TGTATA | | QVARLVTRNLAPRSV |
| | | | CTCTGCCCCAGTCCGGGAGTCGATGGGACTCACCAGCCCAACGATTCCT | CTGGGCAT | GTGGGG | | RGNGFGSGMATHPV |
| | | | TCCAAAATTTCGGTGAAACAAATTTCTCGGTGCAAGTCGCAAGGCTTGT | AGCTGCAG | GGGGAT | | PADESGHESDPFLVG |
| | | | CACCCGAAACCTAGCCCCCCGGTCGGTCAGGGGCAACGGGTTCGGAAG | AGACCGCG | CTCATGT | | RSCGQPARLTRQSVG |
| | | | TGGGATGGCCACCCACCCCGTTCCCGCAGACGAATCCGGCCATGAATCT | CCTCCTCG | ACCGGG | | TQTSRDDILPSKTTKL |
| | | | GATCCATTCCTTGTAGGGAGGAGCTGCGGACAACCGGCACGCCTTACTA | CGGCCCCG | TTTCTTT | | TENELDLLVNFSLELY |
| | | | GGCAATCGGTTGGCACCCAGACCTCCCGAGATGATATTTTACCATCTAA | CTGGTAAG | TATTTGA | | RSDLQGFVQEGIHFS |
| | | | AACCACCAAATTGACAGAGAATGAATTGGACTTGCTGGTGAACTTTTCT | CCCTTAAC | TTTTCAA | | VNREVLEGFPEVYEQ |
| | | | TTAGAATTGTATAGGTCAGATCTGCAGGGATTTGTGCAGGAGGGGGATTC | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | ATTTTTCTGTGAATAGGGAGGTGTTAGAGGGGTTCCTGAGGTGTATGA | AGGGTGA | TAAAAC | PAPQPAVGDDLNTSL |
| | ACAACCTGCACCACAACCGGCAGTAGGGGACGATTTAAACACCAGTCTC | CTAA (SEQ | AGACGG | PPDNNICVLEKGSSE |
| | CCACCGGACAATAATATATGCGTACTTGAGAAGGGTAGCAGTGAAGCA | ID NO: | TAGCTA | AVEDGTPEVAHPVPE |
| | GTGGAGGATGGCACACCGGAGGTAGCGCACCCCGTGCCTGAAACCCAG | 1141) | 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 |
| | GAGATGTTGGACAAAAGAGGAGGAGATGCTGATAAAGTTGGAGG | | GACAGT | TKEEEEMLIKLEVQFE |
| | TACAGTTCGAGGGACACAGAAACATCAATAAGCTTATCGCGGAACACTT | | GCCTTTT | GHRNINKLIAEHLTTK |
| | AACAACTAAAACATCCAAACAGATTAGTGATAAAAGGAGACTATTACCC | | GATTCA | TSKQISDKRRLLPRKQ |
| | AGAAAACAATTAACAGATCTAAGTAAGGGAGTGGCTGGACAGAAGGGTG | | CAACGC | LTDLSKGVAGQKVLD |
| | CTGGACCCAGGACTGAGTCATCAACCCCAGCTGGGGGTAGTTGACAAT | | GTCAAT | PGLSHQPQLGVVDN |
| | GGACTTGGTGGGGGTCATCTGCCAGGGGGGCCAGCTGCTGAAGGAAG | | ACCATCT | GLGGGHLPGGPAAE |
| | AACAATAGAGCCATTAGGACACCACCTTGATAAGGATAACGGTCACCGG | | GACACG | GRTIEPLGHHLDKDN |
| | GAAATCGCTGACCAGCACACAAGGCAGGGAGGCTGCAGGCCCATTACCGA | | GATACC | GHREIADQHKAGRL |
| | AAGAAGATAAGGAAGCGCCTTTCAGAAGGGATGATTAGCAACTTCCCC | | CTTACCG | QAHYRKKIRKRLSEG |
| | GAAGTATTTGAACAACTACTGGACTGCCAGGAAGCACAACCATTGATCA | | GACTTG | MISNFPEVFEQLLDC |
| | ATCAAGCAGCGCAGGATTGCTTTGGATGCCTGGATTCAGCAAGCCAGAT | | TCATGAT | QEAQPLINQAAQDC |
| | AAGGAAGGCGCTCCGAAAACAGAACACACAGAAAGACCAGGGGGATC | | CTCCCA | FGCLDSASQIRKALRK |
| | AACCCAAAAGACCAGCTCAGAAGTGGATGAAAAAAAGAGCAGTTAAGA | | GACTTG | QNTQKDQGGDQPKR |
| | GGGGTCACTTCCTCCGCTTTCAGAAATTATTTCATCTTGACAGGGGGAA | | TCCAAG | PAQKWMKKRAVKR |
| | ATTGGCAAAGATTATTTTGGACGACGTAGAGTGTTTGTCCTGTGATATA | | GTGGAC | GHFLRFQKLFHLDRG |
| | CCACCCAGTGAAATTTATTCGGTATTCAAAGCCCGATGGGAAACACCTG | | GGGCCA | KLAKILDDVECLSCDI |
| | 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 | | | ACPLNPRQRSFISAA |

| | | | | |
|---|---|---|---|---|
| | | GGGTGTTGTATTCGTGGACATCGCCAAGGCCTTTGACACCGTGAGCCAC | NO: 1264) | GCSENLKLLQTIIRTA |
| | | CAACACATCATACATGTATTGCAAAGGAGAGGAGTGGACCCCCACATCA | | 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 |
| | | TGACCTACGTAGCTGACCACTCAGAGATGAAAGCGGGGGGCCCTTGAAG | | GIAKTSLPEKLDFWLE |
| | | CCCTTGACCGGACAATTCGATCGGCGGTCAAGGACTGGCTGCACCTACC | | RIDRAPLKPFQKLDIL |
| | | TTCGAGCACCTGTGATGCCATCTTGTACACGAGCATGAAGGACGGTGGT | | KTYTIPRLTYVADHSE |
| | | TTGGGAGTGACCAAATTGGTGGGACTGATTCCGAGTGTACAAGCCCGG | | MKAGALEALDRTIRS |
| | | AGGCTGCACAGGATTGCGCAGTCACCGGAGGAGACGATGAAAGACTTC | | AVKDWLHLPSSTCDA |
| | | CTGGAAAAGGCCCAGATGGAGAAGATGTACGAGAAAATTGTGGGTCCAA | | ILYTSMKDGGLGVTK |
| | | GCTGGAGGGAAAGAAAGAGGATGCCGTCAATTTGGGAAGCGCTCCC | | LVGLIPSVQARRLHRI |
| | | GGAGGTTGTACCATCCATAGACACAGCCACAACTTCGGAGTGGGAAGC | | AQSPEETMKDFLEKA |
| | | ACCGAACCCTAAAAGTAAGTACCCTAGACCTTGTAATTGGCGCAGAAAA | | QMEKMYEKLWVQA |
| | | GAATTTAAAAAGTGGACTAAATTAATAGCCCAGGGCTGGGGAATTAGG | | GGKRKRMPSIWEAL |
| | | TGTTTTAAGGGGGACAAAATTAGTAACAATTGGATTCGACATTATAGAT | | PEVVPSIDTATTSEW |
| | | ACATACCTCACAGGAAACTTCTCACTGCCATACAGCTCCGGGCCAGTGT | | EAPNPKSKYPRPCN |
| | | GTACCCCACAAGGGAATTTCTCGCGCGGGGAGGGAAGATAACTGTGT | | WRRKEFKKWTKLIA |
| | | TAAGTCTTGTAGGCACTGTGAGGCGGCAGAGGAGTCCTGTGCCCACATC | | QGWGIRCFKGDKISN |
| | | ATCGGCATGTGTCCAGTCGTGAGGGATGCCCGAATCAAGAGGCACAAT | | NWIRHYRYIPHRKLLT |
| | | CGCATTTGCGAGAGGCTGATGGAGGAGGCGGGGAAGAGGGACTGGAC | | AIQLRASVYPTREFLA |
| | | GGTGTTTCAGGAGCCGCACATAAGGGACGTCACCAAGGAACTGTACAA | | RGREDNCVKSCRHCE |
| | | ACCGGACTTGATATTCGTGAAAGAAGGCCTTGCACTTGTTGTGGATGTT | | AAEESCAHIIGMCPV |
| | | ACAATACGGTTCGAGTCAACCAAGACAACGTTGGAGGAGGCTGCTGCA | | VRDARIKRHNRICERL |
| | | GAGAAGGTGAACAAGTACAAACATCTGGAGACCGAAGTACGGAACCTC | | MEEAGKRDWTVFQE |
| | | ACCAACGCTAAGGACGTTATCTTTATGGGGTTTCCCCTTGGAGCGCGGG | | PHIRDVTKELYKPDLI |
| | | GACAATGGTACAATAAGAACTTTGAACTTTGGACACTCTTGGCCTCCCC | | FVKEGLALVVDVTIRF |
| | | AGATCGAGGCAGGACATTATTGCAAAGACTTTATCCACGGACGCGCTCA | | ESTKTTLEEAAAEKVN |
| | | TTTCATCTGTGGACATTATACATATGTTTGCCAGTAGAGGCAGAAGACA | | KYKHLETEVRNLTNA |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GCATGCTTAGGGTAGATAATCTTTGTATAGTGGGGGGGGATCTCATGTA<br>CCGGGTTTCTTTTATTTGATTTTCAATAAAACAGACGGTAGCTAGGTTCG<br>CAAGGCAGCCACAAGCCAAAGATAGGGTAGGGTGCTCATAGTGAGTAGG<br>GACAGTGCCTTTTGATTCACAACGCGTCAATACCATCTGACACGGATACC<br>CTTACCGGACTTGTCATGATCTCCCAGACTTGTCCAAGGTGGACGGGCC<br>ACCTTTACTTAACCCGGAAAAGGAACATATATTAATTATATGTGTTCGGA<br>AAA (SEQ ID NO: 1540) | CGACTTGA<br>GAAGGTCT<br>GGTTACAA<br>CTGGGCAT<br>AGCTGCAG<br>AGATCGCG<br>CCTCCTCG<br>TGGCCCCG<br>CTGGTAAG<br>CCCTTAAC<br>AGGGTGA<br>CTAAGTCG<br>ATCTCTGC<br>CCCAGTCC<br>AGGAGCC<br>GCTGGGTT<br>TCACCAGC<br>CCAGCGAT<br>TCCTTCCA<br>AATTCGGT<br>GA (SEQ ID NO: 1142) | GTAGTC<br>ACATTG<br>CACTTTC<br>TGTAACT<br>TGCACT<br>GGGTGT<br>GGGATG<br>TGGGCC<br>TGGGGT<br>GTGGGT<br>TATGGG<br>GTATAT<br>ATGTGG<br>GATATTC<br>TGGTGG<br>GAATGT<br>CCATTCA<br>CTGTAT<br>GCCTATC<br>TTTTTAA<br>TAAAAA<br>GACGGT<br>AGCTAG<br>GTTCGC<br>GAAGCA<br>GCCACA<br>AGCCAA<br>TAGCCA<br>GTTAGG<br>TAGCTC<br>ATAGTG | KDVIFMGFPLGARG<br>QWYNKNFELLDTLGL<br>PRSRQDIIAKTLSTDA<br>LISSVDIIHMFASRGR<br>RQHA (SEQ ID NO: 1386) |
| R2 | R2-1_ZA | Zonotrichia albicollis | CGACTTGAGAAGGTCTGGTTACAACTGGGCATAGCTGCAGAGATCGCG<br>CCTCCTCGTGGCCCCGCTGGTAAGCCCTTAACAGGGTGACTAAGTCGAT<br>CTCTGCCCCAGTCCAGGAGCCGCTGGGTTTCACCAGCCCAGCGATTCCTT<br>CCAAATTCGGTGAAACAAATTCCTCGGTAAAAGCCGCGTGGCTTATTGC<br>CTGAAACCTGGCCCCCCGGTTTCAGACAGGGGCAAAGAGTTCGGAAGT<br>GGACTGACCACCCACCCCGAACCCGAGAGCGAATCTGGTCATGACCCAA<br>CTGTCCCAAATCCTGGTCCGTCTCTTGGAGCGGGGGAAGGTGCACAGCC<br>ACTACCCTTACTCAGGGTATCGGTGGGCACCCAAACCTGTGAAGAGGAC<br>TTTATAACATCTAGACCAACCAAATTACCCGGAATTGAATCAGAATTAGG<br>CCCGCTGGTGAAGTTTTTCTTTAGAGGTTTACAGGTCAGATCTTAAGGGG<br>GATGTGCAATTTGAGGGGATTCATTTTCCAGATAATTGGGGGGTACTGG<br>AGGGGTTTCCTGAGGTGTACGAACAACTGGCACCACAGCCAAACGGGG<br>GAGACGAGTTAAATCATAGTCTCCCAGGGGACAGGGAGGGGGATGTAC<br>TTGAGAAGGATAGCAGCGAAAAGGAGAAGGAGGCTGCACCAGAGGCA<br>TTGCCCTCAGTGCAAAGGGCCCGCAGTGAACAGTTGCCAGATAACATCG<br>TAAAGGTGACTGTTCCCGACAAAAATCCACCATGTCCCTGCTGTGGTGT<br>CCGCTTAAACTCAGTGTTAGCTCTGATTGAACATCTGAAGGGCTCACAC<br>GGGAGGAGGAGGGTGTGCTTTAGGTGTGCCAAATGTGGGAGGGAGAAA<br>TTTTAACCACCATAGTACTGTTTGTCATTACGCAAAGTGCAAAGGTCCAC<br>AGATTGAAAGGCCACCAGTGGGAGAGTGGATCTGTGAGGTATGCGGA<br>AGGGACTTCACGACCAAAATTGGCCTGGGACAACACAAAAGACATATG<br>CATGCAATGGTGAGAAACCAGGAAAGGATCGATGCTTCCCAACCGAAA<br>GAGACATCAAATCGAGGAGCCCACAAGAGGTGCTGGACGAAGGAGGA<br>GGAAGAACTGCTCATGAAGTTGGAGGTACAGTTTGAGAATCACAAAAA<br>CATCAATAAGCTTATCGCAGAGCAATTAACAACTAAAACAGCTAAACAA<br>ATTAGTGATAAAAGGAGAATGCTGCTCAAAAAAGGTAGGGGGACAACT<br>GGTAATTTGGAAACAGAGCCTGGGATGAGTCATCAATCGCAGGCAAAA<br>GTTAAGGACAATGGACTGGGTGGGGACCATCTGCCGGGAGGACCAGTT<br>GTCGATAAGGGAACAATAGGGAAGCCAGGACAACATCTTGACACAGAT<br>AACAGCCATCAAATAACTGCTGGCAAGAAGAAAGGGGGAGGGCTGCA<br>GGCTCGTTATAGAAGGAGAATAATGAAACGATTAGCGGCCGGGACAAT | | | |

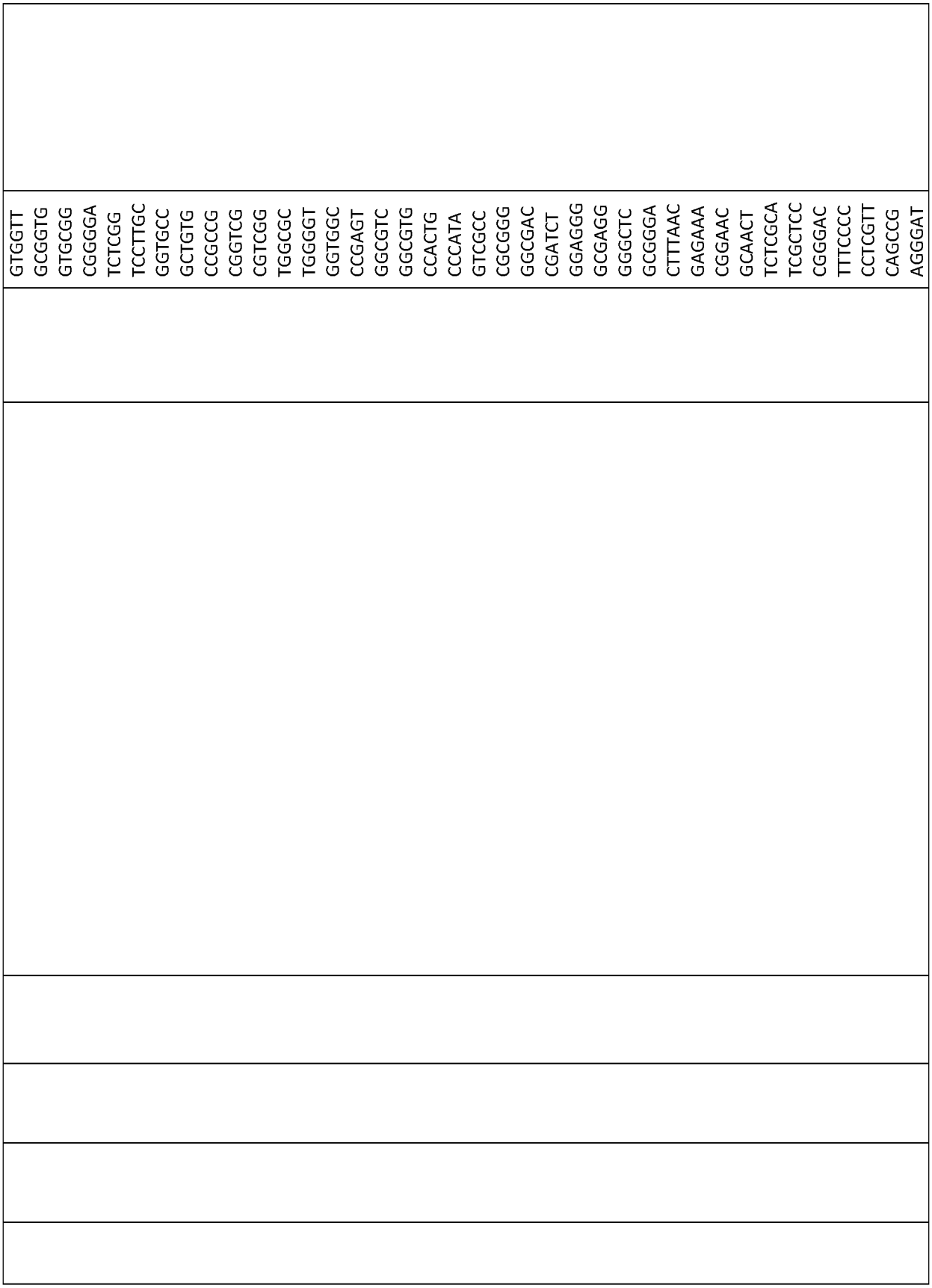
| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | TAACATCTTCCCCAAAGTGTTTAAAGAACTGATTAACGACCAAGAGGCG | | | GGTAGG | | ARPLINQTTEDCFGLL |
| | AGACCGCTAATCAATCAAACAACAGAAGACTGCTTTGGCCTCTTGGACT | | | TGACAG | | DSACQIRTALREKGK |
| | CTGCATGCCAAATTAGAACGGCACTCCGGGAGAAGGGCAAATCTCAGG | | | GAACCT | | SQEERPRKQYQKW |
| | AGGAACGACCAAGAAAACAGTATCAGAAGTGGATGAAGAAGAGAGCG | | | TTGACTC | | MKKRAIKRGDYLRFQ |
| | ATTAAAAGGGGGGGACTATCTCCGCTTCCAGCGATTATTCCATCTAGACA | | | AGAACG | | RLFHLDRGKLARIILD |
| | GGGGGAAACTGGCGAGAATTATCTTGGACAACACTGAGAGCTTGTCTT | | | CGTCCAT | | NTESLSCDISPSEIYSV |
| | GCGATATATCACCCAGTGAAATTTATTCGGTATTCAAGGCCAGATGGGA | | | TAACATC | | FKARWETPGHFNGL |
| | AACACCTGGACACTTCAACGGCCTTGGGGACTTTGAAATTAAAGGGAAG | | | TAGAAC | | GDFEIKGKANNKAFR |
| | GCCAACAACAAAGCCTTCAGGGACTTCATCACGGCTAAAGAAATTGAAA | | | GGACCA | | DFITAKEIEKNVREMS |
| | AGAACGTGCGGGAAATGAGTAAGGGTTCGGCGCCAGGTCCAGATGGG | | | AACTTC | | KGSAPGPDGIALGDI |
| | ATCGCCCTTGGGGACATCAAGAAGATGGATCCCGGGTATTCCCGGACC | | | GGACAT | | KKMDPGYSRTAELFN |
| | GCCGAGCTATTCAACTTGTGGCTGACAGCTGGTGACATCCCGGACATGG | | | GCACCG | | LWLTAGDIPDMVRG |
| | TGAGGGGGTGCAGGACTGTTTTGATCCCGAAATCGACGACACCGGAGC | | | ATTAACC | | CRTVLIPKSTTPERLK |
| | GCCTAAAGGACATCAACAACTGGAGACCCATCACGATTGGTTCCATCTT | | | GGATTT | | DINNWRPITIGSILLRL |
| | GCTAAGGCTGTTCTCCAGGATCATAACGGCGAGGATGACTAAGGCGTG | | | GTCCAA | | FSRIITARMTKACPLN |
| | CCCCCTCAACCCGAGACAGAGAGGCTTCATCAGTGCGCCCGGGATGCTCT | | | GGTGGA | | PRQRGFISAPGCSEN |
| | GAGAACCTGAAACTCCTGCAATCTATAATTCGGACTGCCAAAAATGAGC | | | CGGGCC | | LKLLQSIIRTAKNEHK |
| | ACAAGCCGCTGGGTGTTATTTTCGTGGACATTGCTAAGGCTTTTGACACC | | | ACCTTTA | | PLGVIFVDIAKAFDTV |
| | GTGAGCCACCAACACATCATACACGTTTTACAGCAACGGAGGGTTGACC | | | CTTAACC | | SHQHIIHVLQQRRVD |
| | CCCACATTGTTGGACTGGTGAACAATATGTACAAGGACATCAGTACGTA | | | CGGAAA | | PHIVGLVNNMYKDIS |
| | TGTCACCACAAAGAAGAACACACACGGACAAAATCCAGATCCGGGTT | | | GGGAAC | | TYVTTKKNTHTDKIQI |
| | GGAGTGAAGCAGGGTGACCCACTATCACCCCTTCTATTCAACTTGGCAA | | | ATATATA | | RVGVKQGDPLSPLLF |
| | TGGACCCCCTGTTGTGTAAGCTGGAAGAAAGTGGCAAAGGATTCCATC | | | GTTATAT | | NLAMDPLLCKLEESG |
| | GAGGACAGAGCTCAATAACCGCGATGGCGTTCGCCGACGATCTGGTCTT | | | GTGTTC | | KGFHRGQSSITAMAF |
| | GTTAAGCGACTCCTGGGAGAACATGAAAGAGAACATCAAAATACTGGA | | | GTAATA | | ADDLVLLSDSWENM |
| | GACCTTTTGCAATCTCACCGGTCTCAAAACACAGGGTCAGAAGTGCCAC | | | (SEQ ID | | KENIKILETFCNLTGLK |
| | GGCTTTTACATCAAGCCTACAAAGGACTCTTACACCATCAACAACTGCCC | | | NO: | | TQGQKCHGFYIKPTK |
| | TGCATGGACCATCAACGGCACACCCCTGAACATGATCAACCCCGGGGAG | | | 1265) | | DSYTINNCPAWTING |
| | TCAGAGAAATACCTCGGCCTGCAGATCGACCCATGGACTGGAGTAGCA | | | | | TPLNMINPGESEKYL |
| | AAATACGATCTCTCCACAAAATTGAAAATATGGCTCGAAAGCATTGACC | | | | | GLQIDPWTGVAKYD |
| | GAGCTCCACTTAAACCTCTGCAAAAATTAGACATCCTCAAAACATACACC | | | | | LSTKLKIWLESIDRAPL |
| | ATTCCTCGACTGACCTACCTGGCTGACCATTCAGAGATGAAAGCAGGGG | | | | | KPLQKLDILKTYTIPRL |
| | CTCTGGAAGCACTCGACCAGCAGATTCGAACAGCGGTCAAAGACTGGC | | | | | TYLADHSEMKAGALE |
| | TGCACCTGCCCTCGTGCACCTGTGATGCCATCTTGTACGTGAGCACGAG | | | | | ALDQQIRTAVKDWL |
| | GGACGGCGGTTTGGGTGTTACCAAGTTGGCGGGACTGATTCCAAGTGT | | | | | HLPSCTCDAILYVSTR |
| | GCAAGCCCGGAGGCTGCATCGCATTGCGCAGTCGCCGGACGAGACGAT | | | | | DGGLGVTKLAGLIPS |
| | GAAGGACTTCCTAGAGAAGGCGCAGATGGAGAAGATGTATGAGAAGTT | | | | | VQARRLHRIAQSPDE |
| | ATGGGTTCAAGCTGGAGGCAAAAAGAAGGGGATGCCGTCAATTTGGGA | | | | | TMKDFLEKAQMEK |

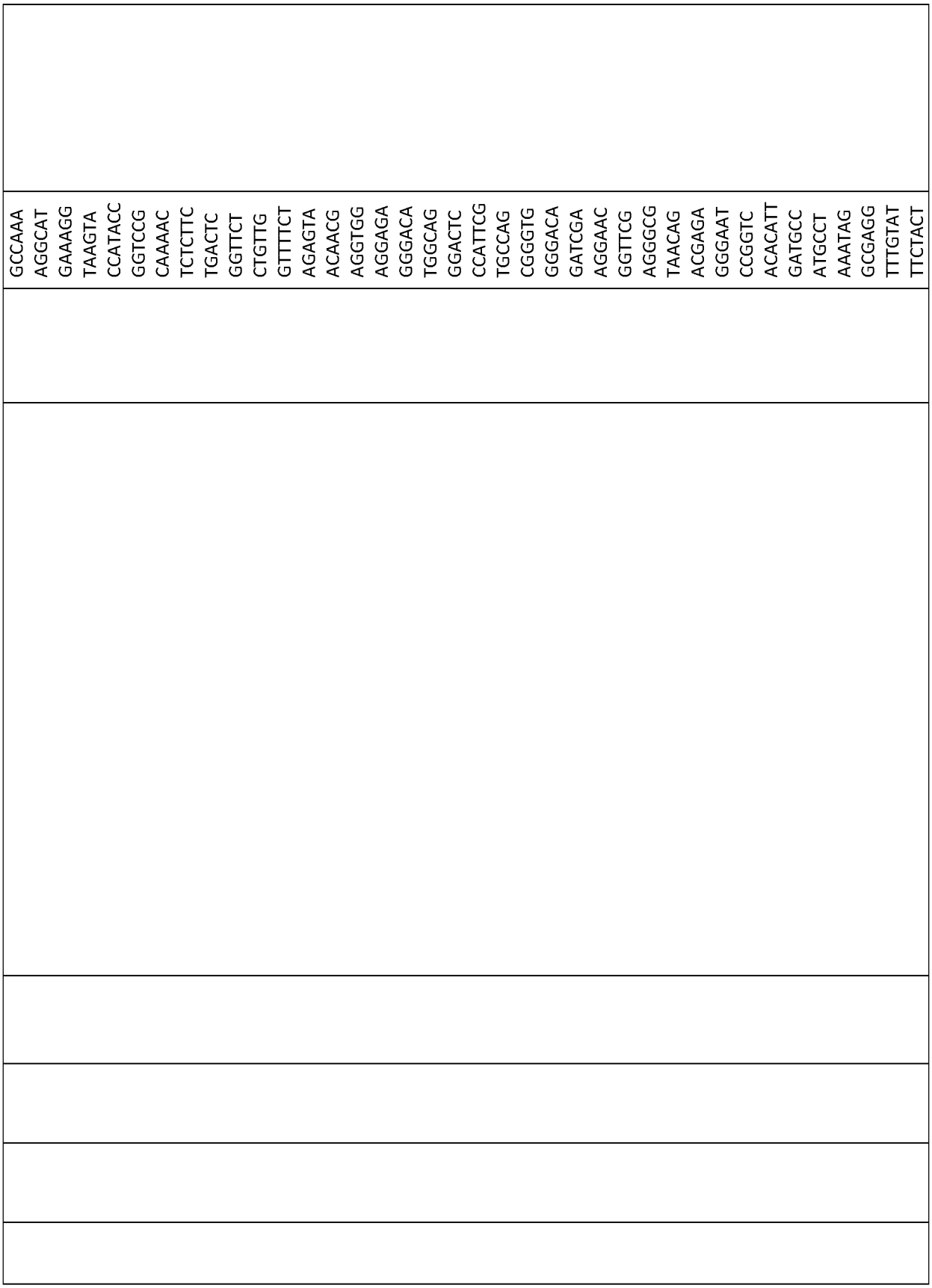
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GGCCCTACCGATGACTGTACCACCCACTAATACAGGTAATCTTTCGGAG | | | | MYEKLWVQAGGKK |
| | | | TGGGAAGCACCGAACCCCAAAAGTAAGTACCCAAAACCTTGTGATTGGA | | | | KGMPSIWEALPMTV |
| | | | GAAGGAAAGAGCTTAAAAAGTGGACAAAATTGGAGTCCCAAGGTCGTG | | | | PPTNTGNLSEWEAP |
| | | | GAGTCAAAAATTTTAGGAATGATACAATTAGTAACGATTGGATCCAATA | | | | NPKSKYPKPCDWRR |
| | | | TTATAGACGCATACCTCACAGGAAACTCCTCACTGCCATACAACTCAGG | | | | KELKKWTKLESQGRG |
| | | | GCCAATGTATACCCCACAAGGGAATTTCTCGCGCGGGGGAGGGGTGAT | | | | VKNFRNDTISNDWIQ |
| | | | AACTATGTTAAGTTTTGTAGGCACTGTGAAGCGGACCTTGAAACCTGTG | | | | YYRRIPHRKLLTAIQL |
| | | | GCCATATCATCGGCTTTTGCCCAGTAACGAAGGACGCCCGAATCAAGAG | | | | RANVYPTREFLARGR |
| | | | GCACAATCGCATATGCGACAGGCTTTGCGAGGAGGCAGCTAAGAGGGA | | | | GDNYVKFCRHCEADL |
| | | | ATGGGTGGTCTTCAAGGAGCCGCACTTGAGGGATGCCACCACGGAACT | | | | ETCGHIIGFCPVTKDA |
| | | | GTTTAAACCGGATGTGATATTCGTGAAAGAGGACCGTGCACTGGTTGTG | | | | RIKRHNRICDRLCEEA |
| | | | GATGTGACAGTACGATATGAATCAGCCAAGACAACGCTGGAGGCAGCT | | | | AKREWVVFKEPHLR |
| | | | GCTATGGAGAAAGTGGACAAGTACAAACATCTGGAGGCAGAAGTGAA | | | | DATTELFKPDVIFVKE |
| | | | GGAACTCACCAACGCAAAGGACGTTGTTTTTATGGGGTTCCCCCTTGGA | | | | DRALVVDVTVRYESA |
| | | | GCGCGAGGGAAATTCTACAAAGGGAACTTTAACTTGCTAGAGACTCTTG | | | | KTTLEAAAMEKVDKY |
| | | | GCCTCCCAAAAACGAGGCAATTGAGTGTGGCAAAGACTCTATCCACGTA | | | | KHLEAEVKELTNAKD |
| | | | CGCGCTCATGTCATCTGTGGACATTGTGCATATGTTTGCCAGTAGATCTA | | | | VVFMGFPLGARGKF |
| | | | GGAAACCAAATGTCTAGGTAGTCACATTGCACTTTCTGTAACTTGCACTG | | | | YKGNFNLLETLGLPKT |
| | | | GGTGTGGGATGTGGGCCTGGGGTGTGGGTTATGGGGTATATATGTGG | | | | RQLSVAKTLSTYALM |
| | | | GATATTCTGGTGGGAATGTCCATTCACTGTATGCCTATCTTTTTAATAAA | | | | SSVDIVHMFASRSRK |
| | | | AAGACGGTAGCTAGGTTCGCGAAGCAGCCACAAGCCAATAGCCAGTTA | | | | PNV (SEQ ID NO: 1387) |
| | | | GGTAGCTCATAGTGGGTAGGTGACAGGAACCTTTGACTCAGAACGCGT | | | | |
| | | | CCATTAACATCTAGAACGGACCAAACTTCGGACATGCACCGATTAACCG | | | | |
| | | | GATTTGTCCAAGGTGGACGGGCCACCTTTACTTAACCCGGAAAGGGAAC | | | | |
| | | | ATATATAGTTATATGTGTTCGTAATA (SEQ ID NO: 1541) | | | | |
| R2 | R2Dr | AB097126 | AATCCCCCCTACCCAATCCCCCCGTCGTGACCTCCAGGCCAGGAATCACG | Danio rerio | AATCCCCC | AAATCC | MESTAKGKSYWMA |
| | | | AGCGTACGACAGTGGCCATCCGGCAATGACACAATAGCGTGACTAACGAC | | 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 |
| | | | GGACGGGGAGCAGGGGAGCACCCGGGAGTGGTGGTGACCCTGCCGC | | AATAGCGT | TGGAGA | RHLAFRHGTVPVRFS |
| | | | TCAGGGACTTAAACTGCCCCCTATGTGGCGGGTCGGCGAGCACCGCGG | | GACTAACG | TCCTTTG | CESCGKTSPGCHSVL |
| | | | TGAAAGTGCAAAGACACTTGGCATTTCGCCACGGAACAGTGCCGGTTA | | ACAATGAG | GGGGGG | CHIPKCRGPTGEPPE |

| | | | | | |
|---|---|---|---|---|---|
| | | GATTCAGCTGTGAATCATGTGGAAAAACTTCTCCGGGTTGCCATTCCGTC | TCAGATCC | GTCGGG | KVVKCEGCSRTFGTR |
| | | CTCTGTCACATTCCGAAATGTCGCGGACCGACAGGCGAGCCGCCTGAGA | ATGACCCT | CTAAGTT | RACSIHEMHVHSEIR |
| | | AAGTGGTTAAGTGCGAGGGATGCAGTAGGACGTTTGGCACAAGGAGA | TGGAGTG | CCCCTCT | NRKRIAQDRQEKGTS |
| | | GCGTGTAGTATACATGAGATGCACGTTCACTCAGAAATCCGCAATAGGA | GGTTAACC | CGGGTC | TDGEGRAGVERADA |
| | | AAAGAATTGCTCAAGACAGGCAAGAAAAAGGGACCTCGACAGATGGA | TCCGCCTC | CTCCCAC | GEGPSGEGIPPKRPR |
| | | GAGGGGAGAGCTGGAGTCGAAAGGGCTGACGCTGGGGAAGGTCCCTC | TTTAAAAA | GGTGAC | RARTPREPSEPPANP |
| | | TGGGGAAGGGATCCCCCCCTAAACGTCCCAGACGTGCGAGAACGCCCAG | C (SEQ ID | GCTCTAC | PILSPQPDLPPGGLRD |
| | | AGAACCGTCTGAGCCCCCCGCGAATCCGCCGATTCTCTCGCCACAACCC | NO: 1143) | CCCTCCC | LLREVASGWVRAAR |
| | | GATCTGCCCCCAGGAGGCCTCCGGGACCTACTCCGGGAGGTGGCCAGT | | TCCTCGC | DGGTVIDSVLAAWL |
| | | GGGTGGGTAAGGGCAGCGAGAGACGGAGGTACGGTGATTGACAGCGT | | TCGTAG | DGNDRLPELVDAAT |
| | | GCTCGCAGCATGGTTGGATGGCAACGATCGGCTCCCTGAGCTGGTTGAC | | AACCCA | QRTLQGLPAGRLARR |
| | | GCGGCGACGCAAAGGACACTGCAGGGCTTACCTGCAGGGAGGTTGGCC | | ACGGTG | PATFVAPNRRRGRW |
| | | CGAAGACCCGCAACTTTTGTTGCGCCTAACCGGAGGAGAGGCAGGTGG | | AACACG | GRRLKLLAKRRAYHD |
| | | GGGCGCCGGCTCAAACTGCTCGCTAAGCGCCGCGCCTACCACGATTGCC | | GTTGGC | CQIRFRKDPARLAANI |
| | | AAATTCGGTTCCGAAAAGACCCAGCCCGCCTAGCCGCGAACATCCTAGA | | 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 |
| | | CCCAAAAAGTGCCACAGTTTCATGGTCAGGCCCTGCAGAGGTGCCTTCA | | 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 |
| | | AGAAATTCAAAAAGGCCTGGATGCGGGCCGGGGAGAGGAGAGTGCG | | 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 |
| | | ACAGTAAGGTACGAGTTCGCTCCGGATACCCTCCAGAATGCCGGAAAG | | CTCCGG | ICDLLKAEAETRGWE |
| | | GACAAGGTCAGCTACTACGGCCCGCACAAAGAAGCGATCGCTCGGGAG | | AGAGCC | VRREWAFRTPAGELR |
| | | CTGGGCGTAAGAAGGGTCGACATACATGGGTTTCCGTTGGGTGCACGC | | 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: |
| | | AGACATGCGTACGTGAGCGCGCATTTTTGCTGTTCTCTGGACTGGGTTTC | | CGAAAC | 1388) |
| | | GTCCCCCTCACAACCATCACTTACACTATAGGGGCACAGCGGCTCCTACC | | AGCACC | |
| | | TCCCTCCCTATGACCCCCCCTTCCCATACCGATCCATGGCTGTTCTAGTCT | | ATAACA | |
| | | GGACCGAGGGTCGGACGGGGCATTTGAAGGTAGCTGGAATCCTCCGCT | | GCGCCG | |
| | | GCTGCGAGCCTGAGGTCGATGGTTAGAGGTGAAATACTTGGGAGGAGA | | TAATAG | |
| | | CACAGCCTCCGGAGAGCCCCTCCCGGGTGGTCATCATGGCAACCGGGT | | CGCACC | |
| | | GAAACCTTACGGTTTCACTTACGAAACAGCACCATAACAGCGCCGTAAT | | GGTGTG | |
| | | AGCGCACCGGTGTGACTACTGTCCAGTGCTGATATTCTCATCTGGAGAA | | ACTACT | |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| | | | | | |
| | | | | | |
| | TACAACACGGGTAATGGCAGAGTATTCAAAACCCAAATGTTTACGATCG | | | GTCCAG | |
| | ACCAACGGAGTCGTTCCCTTGCATCTAGGCCGGACCCGAAACTGCCGTA | | | TGCTGA | |
| | ATTGCCCGTCCCCAAGGTAGCCTCTTAGAAAACCGAAGCCCGGTCGGGG | | | TATTCTC | |
| | CGGTGGTTGCGGCGGCGCTGCGGGGGCCTGCTGCTCGGGCGGCGTCG | | | ATCTGG | |
| | GTGTGCCGCGGTGGTTGCGGTGGTGCGGCGGGGATCTCGGTCCTTGCG | | | AGAATA | |
| | GTGCCGCTGTGCCGCCGCGGTCGCGTCGGTGGCGCTGGGGTGGTGGCC | | | CAACAC | |
| | CGAGTGGCGTCGGCGTGCCACTGCCCATAGTCGCCCGCGGGGGCGACC | | | GGGTAA | |
| | GATCTGGAGGGGCGAGGGGGCTCGCGGGACTTTAACGAGAAACGGAA | | | TGGCAG | |
| | CGCAACTTCTCGCATCGCTCCCGGGACTTTCCCCCCTCGTTCAGCCGAGG | | | AGTATTC | |
| | GATGCCAAAAGGCATGAAAGGTAAGTACCATACCGGTCCGCAAAACTCT | | | AAAACC | |
| | CTTCTGACTCGGTTCTCTGTTGGTTTTCTAGAGTAACAACGAGGTGGAG | | | CAAATG | |
| | GAGAGGGACATGGCAGGGACTCCCATTCGTGCCAGCGGGTGGGGACA | | | TTTACGA | |
| | GATCGAAGGAACGGTTCGAGGGCGTAACAGACGAGAGGGGAATCCGGT | | | TCGACC | |
| | CACACATTGATGCCATGCCTAAATAGGCGAGGTTTGTATTTCTACTTTGT | | | AACGGA | |
| | GGGTTCAGTATAGTCGGAGCATATGGTCGGTTGTCCCGTTGTTTTCACG | | | GTCGTT | |
| | GCGGGCAAGCGACTATCATGATAAAGTAGAATGGGAGACGGGCTCCCT | | | CCCTTGC | |
| | GACAAACCCGGAAAGGCGCCCCCCCGTGGTTCGTAGCAGCTGACGGAT | | | ATCTAG | |
| | CACGCTCGAAGAAAAATGAGTGAGAGGGGACGCCGCAACCAC (SEQ ID NO: 1542) | | | 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

| | | |

| | | | | | |
|---|---|---|---|---|---|
| | | GGTCGCGCCTCCGCGTTTGGGGTGTCGCAGGCGTGAGCCTTCGTCCCCT | CTGTTGCG | CCGAAG | VSLRRWSGPGLLGA |
| | | TAAGTTCAGACGGTCCCGGCTTCTTGCCGGGCCAACCCCCGGTGCAGCG | GCTCCGGT | GGCCCC | DGAGGGAAVASPRG |
| | | TTCTCCCATGTTGGATCGGCACCCAGCCCCGGGTGCCATGCGAGTTCAG | AAACGGTA | CCCGTTT | TQVLGSGAGRRWLG |
| | | ACATTTTGTTTATGTATCGTCTGCGTGGTTGACTTGCTAAGCTCATTTCCT | CCGGAGTC | CAGACC | HGSRGSSPSAARGLR |
| | | CCTCTCACTGCGTCCCCCCCAGGTGCTGATCGGTTGAAGAGGATTCGTCG | GGTTAAGC | TGATTCT | RLTVRLKRLSGGLLSP |
| | | TTGACCTCGGCGGTGAATTTGGGATTGTATTATACAGGTAGGTATAGAG | ATCGTTTG | AGGCTA | KACRDAEEGSSSSPG |
| | | GGCGTGCGGATGTTGCGTGGCGGTGTTGGTACTCCCCCCGGCTGGGGGA | GGCCCGCC | CCTGTG | FRNPKGLGGRGLTPL |
| | | GCGGGTGCGGTGGGGCCAGGCATGGCCTCGCCGGGTGGTTGCAGTGTC | TCCACGTG | CCTAATT | GSRRFCRLTVSLNRW |
| | | CGGTTCAGTCCCGGAGGGAGGCGACTGCTTGGCCACAGGACTGGAGG | GTGGTCCG | GGGGGG | RGSLVKLNASSRASG |
| | | GTTGAGTCCCTCCGTGTCCTGGAGGCTCAAGCGACTGTCTGTCTCTCTGA | CGGTAACA | GTCCCA | RRTPVKPACDSRAGR |
| | | GGCGCTGGAGCGGGGCCTGGGCTGCTAGGTGCGGATGGTGCGGGGGGGA | CCAATAGG | AAGAGA | GSEHAEGGGVSAAP |
| | | GGCGCTGCGGTGGCCTCCCCCCAGGGGTACGCAGGTCCTGGGAAGTGGG | GTGGCTAA | TGTTGTC | MVLRSRRKLTFSVDG |
| | | GCCGGGCGTCGGTGGCTTGGGCACGGGTCGCGAGGGTCTTCTCCTTCT | GAGGCCCA | TGTTGTA | DSNSGDRARSGSVSA |
| | | GCGGCCCGGGGGGCTAAGGCGGCTGACGGTACGGTTGAAGCGACTCAG | GTAATTTC | GAAGGG | ARPGHLLVDGESASS |
| | | CGGTGGCCTGTTGTCCCCTAAGGCGTGTCGGGATGCGGAAGAAGGAAG | CCCGAATT | TTTGCG | RSGPAGDARLAGPST |
| | | CTCCAGCAGCCCAGGGTTCCGGAATCCAAAAGGTCTCGGGGGAAGGGGG | GTCTTCCC | CCACTG | RSRRKGCLPPVDFEN |
| | | GTTGACGCCTCTCGGATCCCGTAGATTTTGTCGGCTGACCGTCTCCCTGA | CCCCGCGC | ACTGCA | PKKRTRLMAKMTNG |
| | | ATCGCTGGAGGGGCAGTCTGGTGAAGTTGAACGCTAGTAGCAGGGCCT | 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 |
| | | AGGGAATTCGGCACAAGAGCGGGGGCGCTCGTTGCACATGCGCAGGGCT | 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 |
| | GTCTCAGCAAGGCAGGAAATGGCTCCCTCGCGGTTGTTCGTCGACTT | 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 |
| | | AAGGTTAAGCGGTGCCGTATTCAGAGGCACCACAAGGTGTGCCAGGTC | TCGCAG | DLICWLDELALIVDVT |
| | | CTCGTCGCGGAGGCTGAGCGCCATGGATGGGAAAGTGGAAAGGGAAAA | 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 | |
| | | CGGGTGTATGGGCTCCGGCAGCCGTCGTCACTCCCATACAACACAGGG | GGTATG | |
| | | GCTGCATCCTGGTGGCCGGTGCTAGTTGGTTCTGGAAGCCCGCCCGGGC | AACAGT | |
| | | TGGTTCGCAGAAGCAGGGTGCGCCCAGGGTAGGTTTGGTATATCTGGG | TCTTGC | |
| | | TCCGGTGCGATACCTATCGATGGGCAGCGAGGGCCGCCTCGTGACGCG | GGATGT | |
| | | CTGTGTGGAGCTGGAGCCGGCCTGGGTATGAACAGTTCTTGCGGATGT | GGCGTA | |
| | | GGCGTAGCTAGATAGTACCCGTGGTTGTGGGCGTGGTGTCGACCAAAT | GCTAGA | |
| | | GTTGTCCTGTGTGCACATAGGCCAAGGGTTACGTGGGTGGCAGTCAGA | TAGTAC | |
| | | AGCACCCGCACCTGGAAGTGATTGCCCCGGGATCCCGGCTCTCTGTGAA | CCGTGG | |

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

| | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | GCTCCGAGGCTGGCGGAGTCACTCACTATAAGTGTGAGAGACGATGTC | TTTTAATTT | AGTCCA | | 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 |
| | | GGCAGTGTCATCCGGAAACAGGTGGGCCGGGGCGCCACCAGGGGGGA | 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 |
| | | ACCCGACTGTGCGGGGGTTCCGCCGGCGCAGATCCTGTGGGTCAGGATG | 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 |
| | | CTTGAGGGTGCGGGGGAGAAGACGAGCGTGCAGGGCTATGAGACCCAA | GGCTATCG | CTAATCC | | PVHLKAEMFNAWM |
| | | GACTGCAGGGCGCGGCGAAACGATTTGCACGATGATCGGACAGCTAGTGC | AATGGACG | ATGATG | | ARGEIPEILRQCRTVF |
| | | CCACAAAACCAGTAGACAAAAGCGCAGGGCAGAGTACGCGCGTGTGCA | GGACCAAT | ATTACG | | VPKVERPGGPGEYRP |
| | | GGAACTGTACAAGAAGTGTCGCAGCAGAGCAGCTGAGGTGATCGA | ACACCCCC | ACCTGA | | ILIASIPLRHFHSILARR |
| | | TGGCGCGTGTGGGGGTGTCGGACACTCGCTCGAGGAGATGGAGACCTA | 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 NO: 1268) | QGDPLSPILFNVVMD |
| | | ACCGATCTTGATCGCGTCGATTCCCCTGAGACACTTTCACTCCATCTTGG | CCTACCCA | | LILASLPERVGYRLEM |
| | | CCCGGAGGCTGTTGGCTTGCTGCCCCCCTGATGCACGACAGCGCGGATT | 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 |
| | | GACTTCGTGCAGTTCGTGCACACTCATATCAACGCCCTCCCATCCGCAT | GGCAGTGT | | VIVDVQVVSGQRSLD |
| | | TCGCGGATCGAGAGGGCGTAGAGGTGGGGGTGAGTCTTCGTTGACCTG | CATCCGGA | | ELHREKRNKYGNHG |
| | | CCGTGCTGGTTGCAAGGTTAGGGAGACGACGGCTCACATCCTACAACA | AACAGGTG | | ELVELVAGRLGLPKAE |
| | | GTGTCACAGAACACACACGGCGGCCGGATTCTACGACACAACAAGATTGTA | GGCCGGG | | CVRATSCTISWRGV |
| | | TCTTTCGTGGCGAAAGCCATGGAAGAGAACAAGTGGACGGTTGAGCTG | GCGCCACC | | WSLTSYKELRSIIGLRE |
| | | GAGCCGAGGCTACGAACATCGGTTGGTCTCCGTAAGCCGGATATTATCG | AGGGGGG | | PTLQIVPILALRGSHM |
| | | CCTCCAGGGATGGTGTCGGAGTGATCGTGGACGTGCAGGTGGTCTCGG | AGCAATCC | | NWTRFNQMTSVMG |
| | | GCCAGCGATCGCTTGACGAGCTTCACCGTGAGAAACGTAATAAATACGG | CTCCTG | (SEQ ID NO: 1390) | GGVG (SEQ ID NO: |
| | | GAATCACGGGGAGCTGGTTGAGTTGGTCGCAGGTAGACTAGGACTTCC | (SEQ ID | | |
| | | GAAAGCTGAGTGCGTGCGAGCCACTTCGTGCACGATATCTTGGAGGGG | NO: 1145) | | |

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

| | | | | | |
|---|---|---|---|---|---|
| | GAAAGTATAGATTCCTTCAGGAATGCTATAGGGAAAAGAGAGCTGAGA | GATCAGTG | AGGAAA | 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 |
| | TATAGACCTGTGACAATCGGCAAACATGCTGAACAGGCTTTACGCCAAAA | TAAGAGGC | ACATTTG | AFDTVLHDSIRKALW |
| | TCTGGGACAAACGGATCCGGAAGAACGTGCGTCTTCATGTGAGGCAAA | TATGGGGC | TAATAA | RKGVPSGVVKVVDSL |
| | AAGCATTTATCCCGGTGGATGGGTGCTTTGAGAACGTAAAAACGATCCA | TAACGATC | GTCAGT | YAGAVTSISVGKTKTR |
| | ATGCGTTCTCCAGTCTTACAGAAAGCGTAAGTTGGAACACAACGTCGTA | CTATAGTC | CTGGTA | SICNSGVKQGCPLSP |
| | 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 | |
| | | | ATACGTACGCTAGAAAAAAGCGACTTGCTGCACGGATGACGGTTCATCAG | | 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 |
| | | | GAGTGGAAGTTTTTCCATGAGTCGTCTCTCGTATAAACTGTGGTAAATCC | 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 |
| | | | GTACAGGTGTAATAAGTGTCCGTATGAGGGTGATGTCCAACTCAGTGTG | 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 |
| | TTGCTATAGAACGAAGAAGGAAGGAAACTGTTAGTAAGATTCTTGATGGT | 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 | KAFDTILHDSIEKALR |
| | 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 |
| | | | AATTACTGTCATCGATATGGCATGTCCGTACGAGAAAATCGGAAAGACAC | | 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 NO: 1270) | |
| | | | GAGCTTGCTCGGCGTTGGCACCTTTAGTCTGTAATATTTTCTTGATATTA | | | |
| R2 | R9Av | | TGGACGAAAAAGGTAGTATGGTTGCA (SEQ ID NO: 1546) | | | MNLPIREHAVSVHNI |
| | | GQ398057 | GAAATAGTTTGCAATGGTAGGTGTATGGCGCCTCTGTGTCTCTCTTTCGC | GAAATAGT | ACTAGT | NKFNYLCQLCSKSYD |
| | | Adineta vaga | TGGATATAGTTTGACGATTTTGTACCAGGTATCTGTTTCTTGTGAGTTCA | TTGCAATG | CTCCTTC | TINSVKAHYVACRRQ |
| | | | GCACCAGTTTGAACAGGCTTAGCGATAGACCTTCGAACTTGAAACACTG | GTAGGTGT | TTCTATT | KNASSTTAVPTNVIN |
| | | | TTGTGAAGCTGGCTGGGCCCCTGCAGATTTTCTCGATTAGAACGTGAGT | ATGGCGCC | AGTCAG | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | GTTACGTCCAGAATGACCCACCAGTGGTTAGTTCTACGTTGCCCTGGAA | TCTGTGTC | TCTAATT | NNQLAINTNQVISRN |
| | | AGGAGAAAAGTTGAGCTAAAATCGCACGGCCTAGTTGTTTATCAAATAG | TCTCTTTC | AATTTTT | PLQCVECLMKQVDF |
| | | GCACGGTGAGGAACTCTTCTATGTACCCTGACTAAAGTACTCACTTGTGC | GCTGGATA | CTTACAT | YAKDTKALVTHMRTK |
| | | GCTGGGTTTGCTCCCCCCTCGCATTGACTTATCTGATCGCACTACCCACCA | TAGTTTGA | TCTACAT | HAAAYEESKKVATRR |
| | | AACGAAACATAAACTTAGCTCGTGGTATCAGTCCACACAGCGTGTGCAGTC | CGATTTTG | CTAGTTC | VAWSPDEDQILAELE |
| | | GGATTCAGGGGAGCGTGTTAGTGACAAGCAGGATAAATATTAACATAGT | TACCAGGT | CATTATT | VKLKKIQKGQLLSRLV |
| | | TAATGTTAAGGCGTTCAACATTCCTTATCCAATTGGAAGAGTTGACTGTG | ATCTGTTT | AAATTG | VEYNKCADKSKAPSR |
| | | AAGTTTGTCATGAAGACATTGGACAAATGAATTTGCCGATTCGAGAGCA | CTTGTGAG | GTATGA | SKDAIRTRRQQHDYK |
| | | TGCCGTATCTGTACACAATATAAACAAATTTAATTATTTATGCCAGCTAT | TTCAGCAC | TCAGTG | LLLRSLQSQQPPVGS |
| | | GTTCTAAGTCTTATGATACTATTAATAGTGTTAAAGCTCACTATGTTGCA | CAGTTTGA | CTATCTC | EDSDSDISSSNNNPLT |
| | | TGCAGAAGACAGAAGAATGCCTCATCCACACAACAGCTGTTCCAACCAATG | 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 |
| | | ATAACAATCCTTTAACAACAACACATAATGTCACTCCAACGCCAGATTCA | GTGAGTGT | GCACTG | STIDDSPVNNIKTVNS |
| | | TCCAACGTTGTACTACTAATACAAAAGATCCGTGAATCTGTAGATTCCAT | TACGTCCA | TTTGCTT | DSIFAPISRDEIKLALS |
| | | TGTAAAAATAACGAACCTCAAATTGAACACGAAATATGCTGAACGCAGCA | GAATGACC | CAGAGA | NTKKDSAAGPDAVTI |
| | | AGTGCGTTCATTAATCAAAATAACAACATGGATCCACTTGAACTATCTAT | CACCAGTG | AACCAC | KEAKAIIDNLYVAYNI |
| | | GCGTGGTATCGAAGAGGAGGATGTGAAGGCAATTCGAGACAAAGAACTTCA | GTTAGTTC | TGTTCAT | WLGVQGIPEQLKLNK |
| | | GAAACCAACCAGGAACGTTCCTTCTTCAACAACTTCGAGAAAGCCAACT | TACGTTGC | ATAGTG | TILIPKGNSDLSLLKN |
| | | CGAAATGCCAAAAGGCTTGAGAAATCAAAAAAATATGGCTATTATCAAC | CCTGGAAA | AAGTTC | WRPITISSIILRVYNRL |
| | | ATCTGTACTATAATAACAAGAAAAAAATTAGTAGCGGAAATCCTCGATGG | 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 |
| | | AAACAAGTTGGATTCAAACCTGTTAATGGTTGTGGTATTAATATATATCTTG | 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 |
| | | CCAAAGTAGCAAATAAAATTGGTAATGAATGGTTATGGTGTATTTCGTGA | ACATTGGA | | SIELKYVAIGCMEDSL |
| | | GAAGAGTTATCCAGATCCAAACAACTCAGGTTCATACCTTCGACCGGAT | CAA (SEQ | | RIMSTFSKAIT (SEQ |
| | | ATAATTGCAGTAAAAAATGGTCATGTTATTGTTCTTGATGTAACGGTTGT | ID NO: | | ID NO: 1393) |
| | | GTACGAAGTAACTGGTGCTACGTTTATTAATGCCTACCAAACAAAAATA | 1148) | | |
| | | AATAAATATAATGCGATTATGGTACAAATCGAGCAAATGTTCAATTGTG | | | |
| | | TTAATGGTGAATTGCATGGTCTAGTAATTGGATCACGTGGTTCAATTCAT | | | |
| | | CACAGTCAACTCCACATCTGGCATCAAATGGGATTCTCTTCCATAGAACT | | | |
| | | TAAATATGTGGCTATAGGATGCATGGAGGATTCGCTCAGAATCATGTCC | | | |
| | | ACATTCTCAAAAGCTATCACATGAACTAGTCTCCTTCTTCTATTAGTCAGT | | | |
| | | CTAATTAATTTTTCTTACATTCTACATCTAGTTCCATTATTAAATTGGTATG | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | ATCAGTGCTATCTCTGCTACACTCAATGCTTAATCGTATGTTATTGACAG<br>TCTGACACTTGATTACTCTTACGACATATGCACTGTTTGCTTCAGAGAAA<br>CCACTGTTCATATAGTGAAGTTCCTCAGTTTTCTGTTGATATATTCTTCTT<br>TCATTCTCGCTTCTCCTTTTCTACTGTGTTCTTTTATCAGTTTTTTGTGGA<br>AAAATTGAGAATAAATAAAGT (SEQ ID NO: 1025) | | | |
| R2 | R2OI | LC349444 | Oryzias latipes | CGCACAGGGGACACAGAGCCTGCCCAAGTACCGCTCCCGAGGGAGCGG<br>GAAACGGGGGGGGTGACTATCCCCCTGGGGTCCGGCGAGAGCGCTGGTCT<br>ACGGACCAGGGGTGGCTGTGGGCAGGCTGCTCCTCAGGCCAGTTGATT<br>AGTTACGCATGGGCTGTACCTCCACGTGGTCCCGCTGGTAACGACTTGT<br>CGGCTAAATCAGCCCGCCCACCATCTGGGATATGGTTGACCGTCTAACC<br>CCAGTACTCAGGTCACAAACAAAATGGGAACAGATACAGTGTATGTCG<br>GCCAGGACTACCCTTCTGGCTTATCAAAACGGGTACCAGCACGGTTAGT<br>GGCGGGACCGATGCTGCGAGAGCGAAGCTGTCACGCCCATGTGTTTAG<br>GGCTGGACACATGTGGAACTGGCGAACCAGCCTTCCGAGCGGGCGCTG<br>GGACCAGCCCGCTTTGGAGAAGTCTCGGGTCCTAACCCGGTCGGTGGC<br>GACGGCCACCGACCCCGAAATTACCTCTTACCCAGGAAAGTCCGTATCG<br>ACAAGTACGCAGGTTCAGGAGGAGGACTGGTGTAGCCGGGAGAGCGG<br>GTGGATCTCGCCAGGACTTGCTCCTGAAGAACCCTCGGTGGTGTCCGAA<br>ATTACAGCCTCCATGGTAGCGACAATGAGGGTAGCAACCGAGGAGGTC<br>GTGCTGGAACCACAGCCTGAACAGGTCGTCACAATACTGCCGGAGCAT<br>GGTCGAAACGTTCCTCCGGGGCTGGCAGAACAGGACACCGCCAGCCCC<br>ATAGAAGTCTCGGTGCTCCTCCCAGACCTCGCTGAGAACTGCCCATTGT<br>GTGGCGTGCCGAGCGGGGGGCCTACGCTTGCTCGGGAAGCATTTTGCTG<br>TCCGACATGCGGGGGGTGCCTGTAACGTATGAGTGCCGTAAGTGTGCGT<br>GGCGGAGCCCCAACAGCCACTCAATCTCGTGTCACGTCCCCAAATGCCG<br>GGGGCGTGCGCGGATGCCCAGTGGCGATCCAGGGATCGCCTGCGATCT<br>CTGTGAAGCCCGGTTTGCCACGGAGGTTGGGGTCGCCCAACACAAGCG<br>GCACGTTCATCCGGTGGAGTGGAACAAGGTGAGGCTGGAAAGGAGAG<br>GTGCGCGCGGAGGGGGGAATTAAGGCGACGAAGCTCTGGAGTGTAGCG<br>GAGGTAGAGACGCTAATCCGGCTCATCCGTGAGCACGGAGATTCAGGT<br>GCCACTTACCAGCTCATTGCCGATGAGCTGGGAAGGGGCAAGACGGCC<br>GAACAGGTGAGGAGTAAAAAGAGGCTCCTGCGCATAGATACGGCAAGC<br>AATAGCCCAGATGATGCAGAGGTTGAGGAGGAGAGGTTGGAATCTCTG<br>GCGGTTCGGTCCTCGTCACGGTCACCCCCGAGCCTGGTGGCGACCAGG<br>GTCAGGGAGGCAGTTGCCAGGGGTGAATCAGAAGGTGGCGAGGAGAT<br>CAGGGCTATTGCTGCTCTCATTAGGGACGTAGATCAGAATCCTTGTCTG<br>ATTGAAACCTCGGCGTCGGACATCATCTCGAAGCTGGGAAGGAGGGTG<br>GATGGGCCCAAGAGACCCAGGCCCGTTGTCAGAGAACAGACCCAAGAG | CGCACAGG<br>GGACACA<br>GAGCCTGC<br>CCAAGTAC<br>CGCTCCCG<br>AGGGAGC<br>GGGAAAC<br>GGGGGGG<br>TGACTATC<br>CCCTGGGG<br>TCCGGCGA<br>GAGCGCTG<br>GTCTACGG<br>ACCAGGG<br>GTGGCTGT<br>GGGCAGG<br>CTGCTCCT<br>CAGGCCAG<br>TTGATTAG<br>TTACGCAT<br>GGGCTGTA<br>CCTCCACG<br>TGGTCCCG<br>CTGGTAAC<br>GACTTGTC<br>GGCTAAAT<br>CAGCCCGC<br>CCACCATC<br>TGGGATAT<br>GGTTGACC<br>GTCTAACC<br>CCAGTACT<br>CAGGTCAC | GGGGGA<br>CAGCTG<br>GGAGTC<br>TCGGCA<br>TGATTAC<br>AAATCTT<br>GCGCTG<br>CACTCG<br>GATGTC<br>GTCCCC<br>GTGACG<br>GACACA<br>TTAATCC<br>GGAAAG<br>CGAGTG<br>GTGACT<br>CGCCTC<br>AAG<br>(SEQ ID<br>NO:<br>1272) | MGTDTVYVGQDYPS<br>GLSKRVPARLVAGP<br>MLRERSCHAHVFRA<br>GHMWNWRTSLPSG<br>RWDQPALEKSRVLTR<br>SVATATDPEITSYPGK<br>SVSTSTQVQEEDWC<br>SRESGWISPGLAPEE<br>PSVVSEITASMVATM<br>RVATEEVVLEPQPEQ<br>VVTILPEHGRNVPPG<br>LAEQDTASPIEVSVLL<br>PDLAENCPLCGVPSG<br>GLRLLGKHFAVRHAG<br>VPVTYECRKCAWRSP<br>NSHSISCHVPKCRGR<br>ARMPSGDPGIACDLC<br>EARFATEVGVAQHK<br>RHVHPVEWNKVRLE<br>RRGARGGIKATKL<br>WSVAEVETLIRLIREH<br>GDSGATYQLIADELG<br>RGKTAEQVRSKKRLL<br>RIDTASNSPDDAEVE<br>EERLESLAVRSSSRSP<br>PSLVATRVREAVARG<br>ESEGGEEIRAIAALIRD<br>VDQNPCLIETSASDIIS<br>KLGRRVDGPKRPRPV<br>VREQTQEKGWVRRL<br>ARRKREYREAQYLYS<br>RDQARLAAQILDGAA<br>SQECALPVDQVYGAF |

| | | | | | |
|---|---|---|---|---|---|
| | | AAGGGATGGGTAAGGCGGCTTGCCCGGCGGAAAAGGGAGTACAGAGA | AAACAAA (SEQ ID NO: 1149) | | REKWETVGQFHGLG |
| | | AGCGCAGTACCTGTACTCAAGGGATCAAGCAAGGCTGGCGGCCCAGAT | | | 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 |
| | | ATCCTGTGTGTTCTCGAAGAAGGCGCGGCTTGACAGGCACGTTATCGGGT | | | PIQMKVGVKQGDP |
| | | TGATCCGAAACTCGTACGTGGATTGCGTGACCAGGGTGGGTTGTGTCG | | | MSPLLFNLAMDPLIH |
| | | AGGGCATGACACCACCAATACAAATGAAGGTTGGAGTGAAGCAGGGA | | | KLETAGTGLKWGDLS |
| | | GACCCCATGTCCCCCTTGCTCTTCAACCTGGCTATGGATCCCCTCATCCAT | | | IATLAFADDLVLVSDS |
| | | AAACTCGAGACGGCCGGAACTGGACTGAAATGGGGCGATCTTTCAATC | | | EEGMGRSLGILEKFC |
| | | GCCACGCTGGCCTTTGCCGACGATCTGGTGCTGGTGAGTGACTCTGAGG | | | QLTGLRVQPRKCHGF |
| | | AAGGCATGGGGAGGAGTCTCGGGATTTTGGAGAAGTTTTGCCAACTGA | | | FMDKGVVNGCGTW |
| | | CTGGGCTGAGGGTTCAGCCCAGGAAGTGTCACGGTTTCTTTATGGACAA | | | EICGSPIHMIPPGESV |
| | | GGGCGTGGTGAACGGCTGTGGAACCTGGGAAATCTGTGGGTCACCGAT | | | RYLGVQVGPGRGVM |
| | | CCACATGATTCCCCCGGGGGAATCAGTTCGTTATTTGGGAGTCCAGGTA | | | EPDLIPTVHTWIERIS |
| | | GGCCCGGGGCGCGGCGTGATGGAACCGGATCTTATCCCTACGGTCCAC | | | EAPLKPSQRMRVLNS |
| | | ACGTGGATCGAAAGGATCTCGGAGGCTCCTCTAAAGCCCTCACAACGCA | | | FALPRIIYQADLGKVT |
| | | TGAGGGTTTTGAACTCATTCGCTCTCCCCCGGATAATTTACCAGGCCGAT | | | VTKLAQIDGIVRKAVK |
| | | CTAGGGAAGGTTACGGTAACCAAATTGGCCCAGATAGATGGGATTGTC | | | KWLHLSPSTCNGLLY |
| | | CGGAAGGCTGTGAAGAAGTGGCTCCATTTGTCACCATCCACGTGCAATG | | | SRNRDGGLGLLKLER |
| | | GACTGCTGTATTCACGGAACCGCGACGGTGGTTTGGGCCTCCTAAAGCT | | | LIPSVRTKRIYRMSRS |
| | | GGAAAGACTAATCCCATCCGTGCGCACGAAGCGTATCTATCGGATGTCC | | | PDIWTRRMTSHSVSK |
| | | AGGTCTCCGGATATCTGGACACGGCGAATGACCAGCCATTCTGTGTCAA | | | SDWEMLWVQAGGE |
| | | AATCTGACTGGGAGATGTTGTGGGTCCAAGCGGGAGGTGAGAGGGGC | | | RGSAPVMGAVEAAP |
| | | AGTGCACCTGTAATGGGTGCCGTGGAGGCTGCCCCGACCGATGTGGAG | | | TDVERSPDYPDWRR |
| | | AGATCGCCAGACTACCCAGACTGGCGGCGTGAGGAAAACCTGGCATGG | | | EENLAWSALRVQGV |
| | | TCGGCCCTGCCGGGTGCAGGGTGTGGGTGCAGACCAGTTTCGAGGCGAC | | | 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 | Limulus polyphemus | TGGGAGG | ATTTTGT | GIDGYMFGYARASG |
| | | | 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 |
| | ATGTCCCTGGCGGCGAACTGTTACGTTTGCTGCGAACTTAAGCAATTTCT | AGAAGCTA | 1273) | | GREAVTCNLKKEEIK |
| | TGTGTGATCTTTGCAATGATAGTATTTTAAGACCAAATCAGGGCTTTCGCAA | 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 |
| | GGGAAGCCGTCACATGTAACCTCAAAAGGAGGAGATCAAAGACCACT | GGGAACC | | | YSGCSTSFRVGSQST |
| | ATGATCAGGTCTACGGTGTGTCAAATGATAGAGTTTCTCTAGATGACTG | GGTCGGCC | | | RPIVLMRGVKQGDP |
| | CCCCAGGCCACCAGGGGCCAATAACACCGACCTCCTGAAACCGTTTACG | CTATATCA | | | MSPILFNIALDPLLRQ |
| | CCAACCGAAGTGATGGACTCACTTCAGGGTATGAAGAACGGGGCGCCT | GAGTAGAC | | | LEEESRGFMFREGQA |
| | GGCCCTGATAAGATTACCCTACCGTTCCTCCAAAAACGTCTTAAAAATGG | CGTTTATT | | | PVSSLAYADDMALLA |
| | CATCCATGTTTCCTTGGCAAATGTGTTTAACCTTTGGCAATTCTCGGGTC | AAATATGG | | | KDHASLQSMLGTVD |
| | GCATCCCCGAATGCATGAAGTCAAATAGGTCAGTCCTCATCCCGAAAGG | GTGAAAAT | | | KFCSGNGLGLNIAKS |
| | GAAGAGCAATCTGCGGGGATGTCAGAAACTGGCGGCCAATCACAATCTC | 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 |
| | | TGAAGTGTCTGGCAATGAGAGCTAATGTATACCCTACAAAAAAGAGGCCT | CTATCTGT | | RSKYFEEFLCRRTILYT |
| | | CAGTAGAGGGAATCTATCTAAGAACAAAGATTCCGCCAAGTGTCGGGG | GGCGGTA | | LDLLWKSNNEQYLER |
| | | ATGCACATCAATGAGGGAGACCCTATGTCATCTAAGTGGTCAATGCCCG | GGTGCCGA | | LAP (SEQ ID NO: |
| | | AAATTGAAGTCGATGAGAATAAGGCGCCACACAATAAGATCTGTGAGCAC | 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 | | |

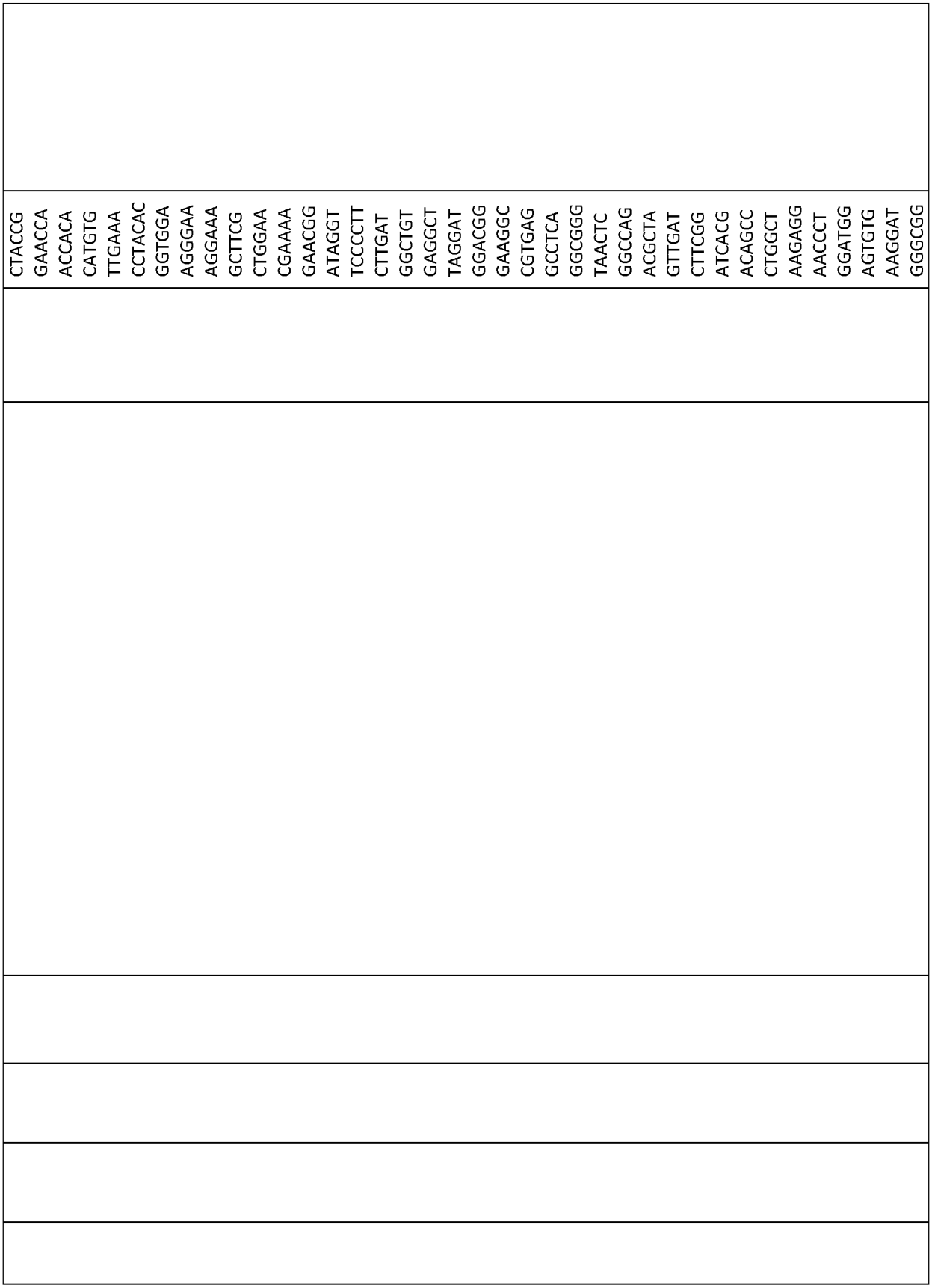
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NeSL | NeSL-1 | Z820588 | Caenorhabditis elegans | GCTCACTTTCTATCGTGTTAACCGTACGTTTACACTCCCAGTGAGTGTAA | TTCAGGCA GCATGTCT GTAAGTAT CCAGTCAT CGAGCGTG ACTGAGG GCGAAATT GATAATAA CTCTGAAA CTGA (SEQ ID NO: 1150) | CCTCCA | MLRRKGRHRMVMV |
| | | | | TAAAGGTTATTCGATAGAGGGGTGTCTCCCTCTTTCTTGGGTAATTCTTCG | GCTCACTT | | |
| | | | | GCGGTCCGGGGTCTCTCCCTCGTCTTTTTTTTTAAACTTTTCTTTCTCATCC | TCTATCGT | GGGCAC | NSVKWQPSAHAEAI |
| | | | | ACTCTTTTGCTCCTTTTTACTAACTCTTGTACTCTATAGTCTTTTCTCATCC | GTTAACCG | GCCGCA | GTGKSWAPQRSQAS |
| | | | | CCCATCCGCCGTTGGGCAAAGTTTATTTACTTTGTTAAATCCATATTTTAT | TACGTTTA | CGCCAA | EHGWQSNAMFDPP |
| | | | | CTCTCTCACCCGTACAGAAAGCGTCTCCTTCTCAAACGCTTTTCTGTACTT | CACTCCCA | AAGTCC | NRILFARDSWSLNQS |
| | | | | TTTCTTATATTTTCATTAACATATTTTTCCTGTTTATACTAACCTAACCTCC | GTGAGTGT | TGGCAT | THLQNQRSGSGLGIR |
| | | | | ATTGTCAATTACTAACTAACTTGTACAACGGATTTCGATGTTGCGCCGAA | AATAAAGG | AACTCT | PGQVRNNMVGGGP |
| | | | | AAGGACGTCACCGAATGGTTATGGTCAATTCTGTCAAATGGCAACCCAG | TTATTCGA | GCAAAT | HRAGDPKRRVELVSI |
| | | | | TGCACATGCTGAAGCAATTGGAACAGGAAAGTCCTGGGCACCACAGCG | TAGAGGGT | AACATC | QGSEVTVRTIYPSDEI |
| | | | | GTCCCAGGCATCCGAACACGGCTGGCAATCAAATGCAATGTTTGATCCC | GTCTCCCT | AAACGT | FSCYSKSCDIKTKAGY |
| | | | | CCCAACAGGATTCTCTTCGCCAGAGACTCATGGTCGCTCAACCAATCAAC | CTTTCTTG | CAATCA | GPEDLKHLTRHIKNE |
| | | | | GCATCTTCAAAATCAAAGGAGCGGATCAGGATTGGGTATAAGACCTGG | GGTAATTC | ACTCCAC | HGLKARWAYQCGLC |
| | | | | TCAGGTAAGGAACAATATGGTGGGGGGTGGGCCTCACAGAGCAGGGG | TTCGGCGG | AAACTCT | NEKSDPSVSEGHKW |
| | | | | ACCCAAAGCGTCGTGTCGAGCTGGTCAGCATACAAGGAAGCGAAGTGA | TCCGGGGT | CCACTCT | MEAHMVAVHQSSA |
| | | | | CCGTCAGAACAATCTACCCGTCGGATGAAATATTCAGTTGTTACTCCAAA | CTCTCCCT | CTTCAA | EKRIKSYQKCTGARV |
| | | | | TCATGTGATATCAAAACAAAAGCTGGCTATGGCCCTGAGGACCTAAAGC | CGTCTTTTT | GTCTTCT | AEQLQAAAPSLTVPG |
| | | | | ACCTGACTCGTCATATCAAGAACGAGCATGGTCTCAAAGCTCGCTGGGC | TTTAAACT | CGGTGC | KHKSGSRDAAKDSM |
| | | | | ATATCAATGTGGATTGTGCAATGAGAAGTCGGACCCAAGTGTATCGGA | TTTCTTTCT | TTCCAAC | TPTKDDDPKTRIYQT |
| | | | | AGGCCACAAATGGATGGAGGCACACATGGTCGCCGTTCACCAAAGCTC | CATCCACT | ACCACA | RSVVKKSTQKTAEPT |
| | | | | TGCGGAAAAAAGGATAAAGTCCTATCAGAAATGCACGGGTGCAAGAGT | CTTTTGCT | ATGGTG | DEGSRGPKYASIFQKS |
| | | | | TGCAGAACAGCTACAAGCTGCTGCTCCATCGCTTACTGTGCCGGGGAAG | CCTTTTTAC | AAAGCT | VKARKSLALLCELSSP |
| | | | | CACAAATCAGGCTCTAGAGACGCTGCCAAAGATTCGATGACACCAACAA | TAACTCTT | CCTTCAC | KPMNPLPTNELTLKE |
| | | | | AGGATGATGACCCGAAAACCAGGATCTATCAGACACGAAGCGTAGTTA | GTACTCTA | CTTTTCC | GNSRELAKEEAPSEGI |
| | | | | AAAAGTCGACTCAGAAAACAGCAGAGCCAACAGATGAAGGGTCTAGAG | TAGTCTTT | CTCCAA | DDIVIIDLDESEESPPR |
| | | | | GCCCAAAGTACGCATCCATTTTTCAGAAATCCGTCAAAGCAAGGAAGAG | TCTCATCC | AATTCTT | RKRFNTWCLDHESSR |
| | | | | | 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 | ACTTTTCT | 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 |
| | GAAGGAGAAAAAGCTCCAAAATTCCTGCTCTCAAGATAGGTAACTGGATG | | TTTTAAG | DPSGELLAKVYNECQ |
| | GAGAAGAAGTTTATCGGAAAGGAGGTGGCGTTCGCTCTGCGGAAAACA | | AGGAAG | RHRKIPKHWKEAETIL |
| | AAAGACACCGCGCAGGGTGCAGACGGACTGCGATACCACCACCTTCAA | | CACCAA | LFKNGDQSKPENWR |
| | TGGTTTGATCCCAGTGGTGAGTTATTGGCGAAGGTATATAACGAGTGCC | | TCCGGA | PISLMPVIYKLYSSLW |

| | | | | | |
|---|---|---|---|---|---|
| | | AACGACACAGGAAGATCCCAAACACTGGAAGGAGGCCGAGACCATCT | | 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 |
| | | GGTCAAAAGTGACCTGACGCCGCCGCAAAAGCTGGACTGTCTTAAAACA | | GGAAGC | NSIPLITEIKMFANIVI |
| | | TTCACGCTGTCCAAACTGACATACATGTATGGAAATTCCATACCACTGAT | | TCCTTCT | RGVKVMHRIPVRGS |
| | | CACGGAGATAAAAATGTTTGCAAATATCGTCATTCGAGGAGTCAAAGTG | | AGCAAG | PLEYIHLPVKDGGLG |
| | | ATGCATAGAATCCCAGTCCGAGGGTCACCACTGGAGTACATCCATCTTC | | AGCTAG | VACPKTTCMITFLVST |
| | | CAGTGAAGGATGGAGGGCTTGGTGTAGCATGTCCCAAGACAACCTGCA | | AGGGAG | LKKLWSDDEYIKTLFT |
| | | TGATTACGTTCCTTGTCTCTACTCTTAAAAAACTCTGGTCAGATGATGAA | | TTCCCAG | SLAEEVVKKESKKSTV |
| | | TACATCAAAACATTATTCACATCACTGGCGGAAGAAGTAGTAAAGAAAG | | TCCTGA | TMDDIADYLNVEERI |
| | | AGTCAAAGAAGAGCACACAGTCACTATGGATGATATAGCCGACTATCTCAA | | AACCCTT | NRSEFGYNSITRLRDV |
| | | CGTTGAGGAGAGGATCAATAGGAGCGAATTTGGGTACAATTCCATTAC | | GCGGTT | MRNLAITGDSPLYRL |
| | | GAGACTGCGGGATGTGATGAGGAACTTGGCCATCACTGGCGACTCCCC | | GATGAT | KMVVKNGKIALLVQ |
| | | ACTTTACAGGCTGAAAATGGTAGTAAAGAACGGGAAAATCGCTTTGCTC | | GGAATG | ATSESMERIYTEEDAK |
| | | GTCCAAGCCACAAGCGAAAGCATGGAAAGGATCTACACGGAAGAAGAT | | GAAGAG | KLQRSLKDQVNKALK |
| | | GCGAAAAAGCTGCAGCGCTCACTGAAGGATCAAGTGAACAAAGCACTC | | TACTTCG | HRFNTTKVVKSKVVR |
| | | AAACATCGATTCAACACCACCAAAGTAGTGAAAAGCAAAGTCGTCCGAG | | GTACTG | VVQQHPASNRFVTK |
| | | TCGTGCAACAGCACCCAGCAAGCAACAGGTTTGTCACAAAAGGTGGCA | | CTCGTT | GGNLSLACHRFVHKA |
| | | ACCTGAGCCTTGCATGTCACCGCTTTGTGCATAAAGCACGTCTGAATCTA | | GCTCTCT | RLNLLACNYNNYDKS |
| | | CTGGCCTGCAACTACAACAACTACGACAAATCCAAATCAAAAGTCTGTA | | CTGCGT | KSKVCRRCGKDLETQ |
| | | GGCGTTGTGGGAAGGATCTGGAGACGCAGTGGCACATACTGCAAAACT | | TTTACTG | WHILQNCPFGFSKKI |
| | | GTCCGTTTGGTTTCTCAAAGAAGATCACTGAGAGGCATGATGCCGTCTT | | 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 |
| | | GGGCACGCCGCCACGCCCAAAAGTCCTGGCATAACTCTGCAAATAACATCA | | 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 | |
| | | CATTGACGGGTGCTTCGGCACTAGGCAAAGGCGCCACCACACTGTCCAA | | AAGGTG | |
| | | TCTCTAAAAAGTTCACATTCATCGAAGAACTACCGGAACCAACCACACAT | | CCCTCGC | |
| | | GTGTTGAAACCTACACGGTGGAAGGGAAAGGAAAGCTTCGCTGGAACG | | CCTAGTT | |
| | | AAAAGAACGGATAGGTTCCCCTTCTTGATGGCTGTGAGGCTTAGGATGG | | AGCTCTT | |
| | | ACGGGAAGGCCGTGAGGCCTCAGGCGGGTAACTCGGCCAGACGCTAGT | | AGACCT | |
| | | TGATCTTCGGATCACGACACGCCCTGGCTAAGAGGAACCCTGGATGGAG | | TATGGG | |
| | | TGTGAAGGATGGGCGGGTAGGGGGTTAAGCCTGTTGACAGACCACCGA | | TCGCCAT | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | CTGCAGTCACACAAAATCAGTGATTATGCGGGTGGACCAATCTGTTGGCGG GTGTTTCCCTCTACCTGACCCCGCAATATGGTATGTACGATCCTCGGATC TAAAATTCATAATGGCCCACCACAACCATAAACCTCCCTAGCAGCTGGTG GTCCCGATAATTCGGGTTCTTGCCACTACTGCGACCCAGGCTCGCC (SEQ ID NO: 1028) | | | GGTTGT<br>GGACGG<br>GTATGC<br>TTGCCG<br>GAGCCG<br>AGTCGT<br>GTTTCTT<br>AGAACC<br>AACCTC<br>GACGAG<br>GCGAAA<br>GCTTGC<br>ACAAGT<br>TAGCAC<br>AATTGT<br>GGTAGG<br>GCCGAC<br>TAGAAA<br>ATGAGT<br>CCCTTAG<br>GGGGTT<br>ACGCCTT<br>GGCGAA<br>AGTGAG<br>GACAAT<br>TGGCAT<br>TGACGG<br>GTGCTT<br>CGGCAC<br>TAGGCA<br>AAGGCG<br>CCACCA<br>CACTGTC<br>CAATCTC<br>TAAAAA<br>GTTCAC<br>ATTCATC<br>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 |
| | | | | CCGTGGCTCTTCCCTAGTCGCTTGCCCTCGCTGCGGCACCCCCCTGCTCAG | CCTAATCA | TTTCTTT | SAGTGLSRHQSRYCG |
| | | | | CTGGCACTGGTTTATCTCGTCACCAGAGCCGGTATTGCGGTCTCACCGCT | ACGTTCTC | TCTTTAA | LTAPRIRRNRVGNST |
| | | | | CCTCGAATCCGCCGAAATCGCGTGGGAAACTCAACAAACACATCTCGCT | TGCACCTT | TAAGTT | NTSRCPPSNTAASPIV |
| | | | | GCCCTCCCTCCAATACTGCAGCTTCACCCATCGTTCCTTCGCCTTCCCCAG | AAACACCA | GTTTATT | PSPSPERPSPPQPAE |
| | | | | AACGCCCAAGCCCCCCCTCAGCCTGCTGAAGTTGTTGCCAGTCTCGAACC | CCAAC | TAAGTA | VVASLEPLSEAEEVLE |
| | | | | ATTGTCTGAAGCCGAGGAGGTGCTGGAGGTCGCCCAGGTTGATGCCGA | (SEQ ID | GTTTCTT | VAQVDAETVDTLEGT |
| | | | | GACTGTTGACACGCTGGAAGGGACCCGGAGAGCTCCGGAATCCGTTCC | NO: 1152) | 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 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CRE | CRE-12_CCri | . | Chondrus crispus | TCCTCTCAAGGTGTTCGACAGGGTGACCCCTTTGGCCCTCTCTTCTTCTC<br>GATCACCCTCCGACCAACCTTGAATGCCCTCAGTCAATCGCTAGGTCCGT<br>CTACGCAAGCACTCGCTTACCTCGATGACATCTACCTCTTCTCAAACGAC<br>TCGCAAGTCCTCAGCAAAACTACCCAATTCCTCGCCGACAAGCAGCACA<br>TCATCAAGCTCAATGAAAAGAAATGCAAGTTAATCAGCTTCGATGAGAT<br>CAGGCAGGAGGGCTTCAAGATGCTAGGGACGATGGTAGGAGGTAAGG<br>AGAAGCGAGCGGAGTTTCTGGAAGGCAGGATTCGGAAGGAAATGGCA<br>AAGGTGGGCAAGCTCAAGGATCTTCCACATCAACACGCGCTCCTTCTAT<br>TACGTTTCTGCATTCAGCAAAATCTACGACACCTGCAGAGAAGTCTGCG<br>CTCGGACGACCTTGTAGACCTATGGGAGAGGCTGGACACGATGCTATG<br>GGAGGAGGTGAAAAGGATGAGGATGAGGCAGCGGGAGGATACGGCG<br>GAAGAGGAGGCTCTAGGGAGATCGTTGACGAAGCTACCAGCGCGACTG<br>GGCGGACTAGGTCTACTTTCCTTCAAAGATGTAGCCCCCCTTGCTTACCG<br>CTCGGCAGCCGAGGCCTCCGACACTCTCCTCGATAACCTAGGTCTCCTTT<br>CTTCGCCAGAGGAACCTCCAACTCCGATCCCCCAACGAACTCGATGCGC<br>AGAACTCTGGGAATCGCAACAGGAAGCCATCCTACATAACCTCGGCGAC<br>ACTGAACGCAAGCGACTCACCGAGAATGCCTCCAGACTCGGCCGAAGTT<br>GGTTATCAGTTATCCCTTACCTTCAACCCCTGCGCCTTTCCAATGTCGAG<br>ATTGCCTCTGGTCTCCATGACCGCACCCTGGTCGGCTCCTCGATCCCTGT<br>CTGTCGCTTCTGTGGGTCGGACTCACCTTTGGGTCACGACGAGCTTTGCC<br>GCGCCCGCAACCCCTGGACCCAGCGCCGGCACAATGCCATCAACCGCGT<br>CATTTATCAACACCTCAAACAAATTCAAGGTGCCACGGTTGAGATTGAG<br>CCCCACACGCTGTCGGGACAAAGGAGAAACGACCTTCGGGTCAGAGGT<br>TCCAGCGCTCTGGCCTTCACTGACTACGACCTGAAGGTTTACTCCCTCGG<br>GGACCGAGACGCGAGAAGCACCGTCACACCCTGCGCCCCCCAACGGCAA<br>GCTGGCCGACTTCTGCTTGGACCGGTGCGTGAACTGGCTCGACAAGGT<br>GGGTCAGGTCGTCTCTAAGAACGCTCCGAAGGTCACTGGTGGGGGTCTTT<br>AAACCAATCATCCTTTCCACTGGTGGCTTGATGAGCAGGAGCACAGCAG<br>ACGAATGGAAGGACTGGAGGGACGCGATGCCGGTGGGGGGGTTCGAG<br>AAAATGGAGAAACGGATTGGTGTCGAGTTAGTAAAGGCAAGGGCGAG<br>GACGCTGGTCTTATGAGGAAGAGGAGGTTGGATTATTTTTTCTTTTCTTT<br>AATAAGTTGTTTATTTAAGTAGTTTCTTTCATTCGGGCAACCCACACGAC<br>AACCCAATAAATTAAACAACGAAAAATGCAACCTCTATAACCC (SEQ ID NO: 1029) | | | | TCKWAYGDSSDLVC<br>GDKILQSSQGVRQG<br>DPFGPLFFSITLRPTL<br>NALSQSLGPSTQALA<br>YLDDIYLFSNDSQVLS<br>KTTQFLADKQHIIKLN<br>EKKCKLISFDEIRQEG<br>FKMLGTMVGGKEKR<br>AEFLEGRIRKEMAKV<br>GKLKDLPHQHALLLL<br>RFCIQQNLRHLQRSL<br>RSDDLVDLWERLDT<br>MLWEEVKRMRMRQ<br>REDTAEEEALGRSLTK<br>LPARLGGLGLLSFKDV<br>APLAYRSAAEASDTLL<br>DNLGLLSSPEEPPTPI<br>PQRTRCAELWESQQ<br>EAILHNLGDTERKRLT<br>ENASRLGRSWLSVIP<br>YLQPLRLSNVEIASGL<br>HDRTLVGSSIPVCRFC<br>GSDSPLGHDELCRAR<br>NPWTQRRHNAINRV<br>IYQHLKQIQGATVEIE<br>PHTLSGQRRNDLRVR<br>GSSALAFTDYDLKVYS<br>LGDRDARSTVTPCAP<br>NGKLADFCLDRCVN<br>WLDKVGQVVSKNAP<br>KVTGGVFKPIILSTGG<br>LMSRSTADEWKDW<br>RDAMPVGGFEKMEK<br>RIGVELVKARARTLVL<br>(SEQ ID NO: 1397) |
| | | | | ACGCCCCCTCCATTTCTGCCAGCCTCCCATCGGCTCGCCGTCTCCGCA<br>ACCCCTCTTCCTCGGCTGTACCAGTTCCGCTCCCACAACCTCCCTCGCCAC<br>AATGAGTCAACCTAATATTTCGTCCGCTGAGACCCCCCTTGTCTCAGCTGC | ACGCCCCC<br>TATCCATT<br>TCTGCCAG | TAAGTC<br>CTTGAC<br>GCCTGC | | MSQPNISSAETPLSQ<br>LPTPVPTPPSPSNPSL<br>SLPTVRDLLLCPIRSSH |

| | | | | | |
|---|---|---|---|---|---|
| | | CCACGCCTGTTCCCACCCCGCCTTCTCCCTCCAATCCCTCTCTCTCTCCC | CCTCCCAT | CCCCGTG | VYSSIPSSCLHSFTML |
| | | TACTGTGCGTGACCTCCTCCTCTGCCCCATACGCTCCTCCCCATGTTTACTC | CGGCTCGC | ATACAG | LIKTVRAASATMTPT |
| | | ATCCATCCCTTCCTCGTGCTTACACAGTTTCACGATGCTCCTCATCAAGAC | CGTCTCCG | CATCGG | ESHRAFIHLHILPIAVL |
| | | TGTCCGCGCTGCGTCGGCCACAATGACCCCAACTGAATCMCATCGCGCA | CAACCCCT | TACCCCT | RRSFRGETGWRSRT |
| | | TTCATTCATCTACACATTCTTCCCATCGCTGTCTTGCGACGCTCGTTCCGT | CTTCCTCG | AGCATTT | GQHHALRQRIRRASS |
| | | GGAGAAACCGGATGGCGTTCCCGCACAGGGCAACATCATGCCCTCCGC | GCTGTACC | GAATAA | GRHWAALWHEALA |
| | | CAACGGATTCGCCCGCGCCTCCTCGGGTCGGCATTGGGCTGCCTTGTGGC | 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 |
| | | AGGCTCGGTATGCAAAGTCGAGCATGTTTTTCGCCTTACACCTCCCCACC | | | LSLLDEIPDARTRFHL |
| | | TCTCCCTCCCAGCAGCTACAAAATTTGATGAACAACAAATCGCTGCTTAT | | | HRVTGSVCKVEHVFR |
| | | TCTCGGTTGAATGATGTGGCCGTTTCCACATCCATGGCTACACAAATAG | | | LTPPHLSLPAATKFDE |
| | | GCCTTCCGTTTCGCCTCGGTGGACACGGCTTCACCCCACTGTCACCATTC | | | QQIAAYSRLNDVAVS |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | ATCCATGCGTCCTACGCTGCCAGTTTAATTGAGGCGGCACCTGTTCGTGT | | | | 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 | . | CNCCAGCCAMCGATCCCGCCGCCACTCGCMGCCCGGCCGTCTCGACCG | | CNCCAGCC | TAATTCA | MAXXPXISPPGAPPA |
| | | | CCACCTCCCCGASGCCCCAGCCCATCATGGCCTSTWMGCCCCWGATWT | | AMCGATC | CCTTCAT | PLRYRMLQCPPPLPK |
| | | | CTCCCCCGGGGGCCCCCCCCTGCTCCGCTGCGGTACCGGATGTTACAATG | | CCGCCGCC | ATCTGCT | XXXXPVPHPMSSPIR |
| | | | TCCACCSCCGCTACCSAAGCSMCASTMGTTKCCGGTCCCCCACCCGATGT | | 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 | | | TPPPXPPRPRPPPYRFPP |
| | | CCCCCGCGCCCGCCTCCCTACAGGTTCCCTCCCCCWACTCTCGACCAGCA | | | PTLDQHXFALSXAYP |
| | | CTTMTTCGCCCTCTCWTMAGCCTACCCCCACCCGMCTCCCMGGCGCCC | | | HPXPRRPPSPXRXLR |
| | | ACCCTCMCCTTKCCGTCSGTTGAGGCACTCCTTTCCTCCCCGATTCGGCC | | | HSFPPRFGXQTFSSIP |
| | | STCAGACATTTTCTTCCATTCCGGGACCGCGCCTTCATAGCACTGTATTAC | | | GPRLHSTVLLLIRLVR |
| | | TTCTCATCCGCCTCGTCCGCGCCGCTACAGCCGCCAACACTCCCGAAACC | | | AATAANTPETTTLXS |
| | | ACCACACTGMATTCTTGCACCTTCACCTGCTCCCGACTGCCSTTCTTCGA | | | CTFTCSRLPFFERPSX |
| | | GAGGCCTTCCGKGGCGAGSCTGGCTGGAGGTCCTCGCGCGGTCAACTTC | | | AXLAGGPRAVNFML |
| | | 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 | VAAVPRADSRGRTP |
| | | | | TTGCCTTCGCCTTCGTCTCGACGCCTTGCTCTCGWAATCGCTGCCACTAC | GCCCTAAT | ACCGCA | APTPSVAPPXAGPPP |
| | | | | GTGCCAGCTTTTTCGTGCCTTGTCTTCGTGTCGACCCGGACCGTTTGCAA | CGACTTTG | CGCGCC | RAAPRAAPSPLPCPP |
| | | | | GCCCTCGCCTTCGTACCCCGCTCTCGTAGCCGTTCTCATCGCTGAAGCGT | GCGCCAAA | ACGACA | ALPHPPPSASPPTSSV |
| | | | | TCTACGCGCTGGCAGCAAGCCTCGGCCCTAGCTTGTAGCGCCGCCGGTG | GCGACTTT | TATTGTC | TSPCSPPTTPPSQPSP |
| | | | | GCCGCTCGCCAACATGGCTACGACGACCATCTCACGATCCCCTTCGTCTT | CCCCGGCC | GCGCGC | DLFSGFANAPTTPSP |
| | | | | CTTCTTCTTCTTCCGCTCGCTCGGGCGTCAGCTTCCACGTCCGCCT | GATTTTCT | TGTACA | PSTPXSSPAGSPIPAA |
| | | | | CAGTTGCGTCGATACCCCGCCTCTTCCGCGATGGCCGCTTCCACTGCCCT | TCCTGCCT | GGCGGC | RRFVLPVATPYPAPA |
| | | | | CTCGCCCACTGTCAGACCCGCACGTCCACATGGCAGGACCTCTCCGCGC | TTTTCTTTT | TAGGTC | PRANRPKLSPVARPF |
| | | | | ACCTCACACGCATGCACGACGGTGACGTGCCCCGTGACGTCGCCGCCGC | CTCTCCAA | GAGCCC | VPKARAGAIPEASSP |
| | | | | CTGCGGCATCGTGCAGTGCCTACACGAGGGCTGCCGCAAGTGGTTTCGC | GCGACGC | AGCCGA | VTPQDRAVSRREDA |
| | | | | GGAGCTGCAGGACTTGCCTCTCATAGGGGCAAGGCCCGTCACGCCCCG | GCCTTTTA | CCGTTCT | AAAPSSAPGLGLADE |
| | | | | CCACCAGCCCCCCGCGCCGCCCTCGCCGTCGCCGTGCCCCGCGGG | CTTTGCCG | GAGCCT | HEDDDTYGGDTIALT |
| | | | | ATTCTCGCGGTAGGACCCCAGCCCCGACCCCCTCGGTAGCCCCGCCCWA | CCGTTCTG | CAGTCG | APHAPRETRAPFEFE |
| | | | | CGCCGGTCCTCCGCCGCGAGCTGCGCCACGCGCCGCCCCCAGCCCACTG | TTTTTTCTT | GCTTGA | ACFLEEEAPATAGDL |
| | | | | CCGTGCCCGCCTGCGCTCCCGCACCCGCCCCCCTCTGCCTCTCCTCCCAC | TTCTCTTTG | GCCCCC | PPYARAFLACPSARL |
| | | | | CTCCAGCGTGACGTCCCCGTGCTCCCCGCCCACGACCCCGCCGTCGCAG | CACTTCGC | GGCTTC | QEIPRRLKSAWQAAA |
| | | | | CCTTCGCCCGACCTGTTCTCGGGTTTCGCGAACGCGCCTACCACGCCATC | TTCTACTTC | CCAAGG | KTIAEAALDCHTAGD |
| | | | | GCCGCCCTCCACGCCGWCWTCGTCGCCAGCAGGCTCGCCCATCCCGGC | ACACCTCC | CCTACC | TQGYNAHLRLFIELPA |
| | | | | TGCCAGACGCTTCGTCCTGTGGCCACGCCCTACCCGGCCCCCGCG | TCCTCCTC | GGGGCG | RGLAVPTNCRGAART |
| | | | | CCGCGTGCTAACAGGCCCAAGCTGTCGCCGGTCGCGCGCCCCTTCGTCC | CTTCTCGA | GCTCTTT | KLQRERLLDIAAGRIP |
| | | | | CTAAGGCGCGAGCCGGAGCGATACCTGAGGCGTCCTCACCTGTGACGC | CCCGCGCG | TTCGCCC | AIPDPPCDAPGADDA |
| | | | | CTCAGGACCGCGCCGTCTCACGCCGCGAAGACGCCGCCGCGCCCCTTC | 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 |
| | GCCGGCCCCGCGGCTGGCAGTGCCCACCAACTGTCGAGGCGCCGCCCG | AGACGCCC | CTTAGTT | AAFLIDMAQRALRGT |
| | CACCAAGCTTCAACGAGAGCGCCTGCTCGACATCGCCGCCGGAAGGAT | CCGGTCTT | TTTCCTT | LHWRGLLTASRLVAL |
| | CCCGGCCATCCCGGACCCGCCGTGCGACGCCCCGGGCGCCGACGACGC | GCTCTCAG | CGGCCG | KKPDGGVRPIAVGEA |
| | CCTACGCGGTTTCCCCGTATCGGGGACGACAGCCGGCGACGTCAGCAA | TCTCCCGC | CGGCAG | LYRVIGRLVLKADRV |
| | CGACGACGACAGCGGAGGTGTGCACGATCGGCCCGCGGCCACCGCCAG | ATCGAAGC | CTTGTTG | MSSADATQYVGRHQ |
| | CGCCCGGCAAGCCAAACGGCTAGTGGAGCAAGGGCTCTCCTCCCGAGC | GGTAGTCG | CCCGGC | YGVAYPGGVEAPVH |
| | CCTCCGTGCCCTTGAACGGGGCGAGCCCGCGGTCGCCTCGGCAGACAC | GGGTACGT | ATAGTG | AVRELHDSGQLRAVV |
| | CCTCGGGCGTCTCGAGGCGCTCCACCCGCCTAACCCCACCGACACAGAGGA | GCTCAAGT | TTAATAT | SLDWRNAFNSLDRV |
| | CTATGGCCCGGCGCCCCGAAGGCCGCGATCCCGCGGGTCACCGCCAAG | GACTCAAG | GTTTAA | HTALLIADRAPALARL |
| | CACCTGGCCCAAGTGGCCAAAGAGCTCCCGCGCGTAGTGCGCCGGGT | 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 |
| | CCCACGGTGTCACGGACGCCGCGACCAAGGCCGCCGTCGCTGCCACCTT | TTCGCCTT | | VRDGSLQADIARLPV |
| | CGCTACGGCGGAACGCGAAGGCGCGGCGCCGCTGGCTTGCGGCTCAACCG | 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 |
| | | ACTGGCCGAAGCCGTGCGCATTAAGTTGCTCACGACGACGGCCAATGC | CGCTGGCA | | INDNSLLSDDDDDDD |
| | | GGCCGGCGTCTGTCCGGCATGCCACAAGACCGGCATCGACCCGTCCCAC | GCAAGCCT | | HDDNCHGEERGEGE |
| | | GCGTACACGTGCGTTTCGCTATCCCATTTGCGCACAGCACGCCACGACG | CGGCCCTA | | RNVTCPGHYTATPFA |
| | | TGGTCGTTCGCCGAGTCGAGCTCGCCTGCAAGACCGAGAAGCCGGTCC | GCTTGTAG | | ADDTLDNSDEDNED |
| | | GCGAACACGTGCTCGCCATCCCCCCCGTCGCGCCCACCGACAACAACAA | CGCCGCCG | | NAHEDDDEDGKDD |
| | | CAACGGCGACGAGGACGGCAGCCCAGTCACCACCGCCGACGACAACGC | GTGGCCGC | | NDDDVYNNCNSSSS |
| | | AGACGGCCACGCGGTCGCGACCAAACGTCGCCCCGAGACCCGCGCTTC | TCGCCAAC | | DGDEGGDDLDYEYS |
| | | AGCCAGAGCCGCCGCCGCCACCGCCGCCGCCGCCGCCAAT | (SEQ ID | | DQSVTRSVDAATGES |
| | | CATCAACGACAACAGCCTCCTGAGCGACGACGACGACGACGACCA | NO: 1155) | | PNPERPTTPTRALLRA |
| | | TGACGACAACTGCCACGGAGAGGAAAGAGGAGAGGGAGAAAGGAAC | | | DLWLPATSTAVDVM |
| | | GTCACGTGCCCCGGCCACTACACCGCCACACCTTCGCCGCGGACGACA | | | VAAACRRSRAKAFDR |
| | | CGCTCGACAACAGCGACGAGGACAACGAGGACAACGCTCACGAAGACG | | | AVSRKAAKYGPAVA |
| | | ATGACGAAGACGGTAAAGATGACAACGACGACGTCTACAACAACT | | | DGSIAKVVPFVVSPF |
| | | GCAACAGCAGCAGCGACGGGGATGAAGGCGGTGACGACCTGGAC | | | GVLSRPAKAFLKRAM |
| | | TACGAGTACAGTGACCAGAGCGTCACTCGAAGCGTCGACGCCGCGACG | | | GDTTAAKQAKARLRL |
| | | GGAGAGAGCCCCAACCCCGAGCGCCCTACCACGCCCACCCGCGCACTAC | | | AVAAVRGTARLSYA |
| | | TACGCGCGGACCTGTGGCTACCCGCCACCTCCACCGCGGTGGACGTGAT | | | WGACAALIVGGN |
| | | GGTCGCGGCCGCCTGCCGTCGGTCACGCGCCAAGGCCTTCGACCGAGC | | | (SEQ ID NO: 1400) |
| | | CGTCAGCCGCAAGGCCGCGAAATACGGCCCTGCGGTAGCCGACGGCTC | | | |
| | | GATCGCCAAGGTGGTGCCGTTCGTCGTGTCGCCCTTTGGCGTACTCTCG | | | |
| | | AGGCCGGCCAAGGCCTTCCTCAAGCGCGCCATGGGCGACACGACGGCG | | | |
| | | GCCAAACAGGCCAAGGCGCGTCTGCGCCTCGCCGTGGCCGCCGTCCGA | | | |
| | | GGCACGGCCCGCCTCTCCTACGCCTGGGGCGCCTGCGCCGCCCTCATCG | | | |
| | | TCGGCGGCAACTAAGCCGCGACGAGGACGGCCAGGACGACCAGGA | | | |
| | | CGACGGCGACGGCGACCACCTAGCACCGCACGCCACGACATATTGT | | | |
| | | CGCGCGCTGTACAGGCGGCTAGGTCGAGCCCAGCCGACCGTTCTGAGC | | | |
| | | CTCAGTCGGCTTGAGCCCCCGGCTTCCCAAGGCCTACCGGGGCGGCTCT | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | TTTTCGCCCTGGTTTTTGCCGGCCTGTTTTTTCTCTCCCCCTTTCCCCCCT | | | | |
| | | | TTTCCATTTGTACTTAGTTTTTCCTTCGGCCGCGGCAGCTTGTTGCCCGGC | | | | |
| | | | ATAGTGTTAATATGTTTAAAAAACGTGTAAATAAATAACTGTTTAACCCT | | | | |
| | | | AACCCTAACCCTAA (SEQ ID NO: 1032) | | | | |
| CRE | Cre-1_FCy | Fragilariopsis cylindrus | ATCAATCTAATACTGAAGGCAATACCAACTCAAACCCGAAATCAAAATC | ATCAATCT | TAGCAC | | MAPLPWNAATSSPP |
| | | | GTTAGAATCAATATACGACCCCCGCTGCTGTACATGTCCAGCCGGATCTC | AATACTGA | CACCATC | | SPVPLTNDKKKDSTLP |
| | | | GTTGTAAAGAAGATTGCAGCTGTAAAAAAAGTTGGACTTCTTTGTTCTTCC | AGGCAATA | TATTCAT | | TATSKNLSKNNNNK |
| | | | TGTGAAGAAGTTGATTGTGGCTGTTCAAATTCTTTTCATAATAAAGAATT | CCAAACTC | ATCCAC | | NNNTNRINNIKNND |
| | | | AATGGCTCCACTTCCTTGGAATGCTGCGACATCTTCACCACCGTCACCTG | AACCCGAA | ACACTG | | NTNDGSNKINLKLPP |
| | | | TCCCATTAACTAACGATAAGAAAAAAGGATTCTACTCTACCTACTGCAACA | ATCAAAAT | ACCACCT | | AAVKITNPYKNKKKN |
| | | | TCAAAAAATTTATCCAAAAATAATAATAATAAAAATAATAATACTAATAG | 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 | | | |
| | | ATGCTAGCAAATAGGCTTATTGCGCTAGACAAAATGCCCCGGAATTCGAC | | | |
| | | CAGTCGGTATAGGTGATACATGGCGTCGTTTCTTCGCCAAACTTGTTCTA | | | |
| | | GCAGTCTCTATGTCTTATGCTACTGACTGTTGTGGGTCAGACCAGCTCTG | | | |
| | | TGCCGGACTAAGAGCCGGAGTTGATGGTGCCATACATGGACTATCGGC | | | |
| | | TATGTGGAGGGAGATGGAATCTGAGGAAAACACAGGTTTCGTACTTATT | | | |
| | | GACGCAGACAATGCATTCAATGAGGTCTCACGCATCAACATGTTATGGA | | | |
| | | CGATCCGCCACGAATGGCCTGCTGGAGCTCGATTCGCCTTCAACTGCTA | | | |
| | | TCGGCACCACAGCCTACTAGTGGTACGGAATCCAGGCGGGAAACCCTTC | | | |
| | | ACTTTCTTTTCTAAAGAAGGTGTCACACAGGGCGACCCATTTGCGATGAT | | | |
| | | AGCATATGGTGTCGCTCTCCTACCACTCATCCGCAAACTGAAAGAATTAA | | | |
| | | ATGTATTATTAGTTCAATCTTGGTATGCAGATGATGCTAGCGCAGCTGG | | | |
| | | CAAATTTGATGAAATACTACGCCTTTTTCAAGATTTATTACGAATGGGAC | | | |
| | | CTGATTTTGGGTACTTTCCTAATGCATCTAAGAGTATCCTCATCACCCATC | | | |
| | | CCGACAATGTGGTTGCAGCTCACCACTTCTTCAACGAGACCCATGGCCT | | | |
| | | AGGTTTCAAGATCAGCACAGGAAGTCGTTTCCTGGGTGGTTTCATTGGA | | | |
| | | 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 |
| | | | TAGAATTGTCCGCTAGCACCACCACCATCTATTCATATCCACACACTGACCAC | | | | NKPYSVTCGIVRSRISI |
| | | | CTCCACCTTCACAACTCCCACTCTCAATTCCCCTGACTACTAAGAAATATTT | | | | ACARASNQCLRGSRI |
| | | | CATGGTGGTTACATTGGAGGTATCCCAACACCAAGCACACAAAATGAACCC | | | | PFKTMSRQIQWEDG |
| | | | ACTAACCCTCTCATCCTATCCACGGGGACCACCTTTGAGCAGAACACCAT | | | | AGAGLYRIVR (SEQ |
| | | | TCATATTACAACCTTTAGCTAGATTAAGATAATTATTTAGTACATATTTTA | | | | ID NO: 1401) |
| | | | TACTATTAAAAAAAAAAAAAAAAA (SEQ ID NO: 1033) | | | | |
| CRE | Cre-1_HM | . | Hydra vulgaris | TTTCTAATGTTACGTGATATGATATGGTTAGTTCATGGTTAGTTTATGTTT | TTTCTAAT | TAACTTG | MNMVSICKRCDRSF |
| | | | | ATGCTTAGTTTATGGAAAATCGTTTATTTATGGCACAATATTGTTTGCTG | GTTACGTG | TATTTTT | TTLKGLNIHKGQCKIF |
| | | | | TTTTTAAATTTATGTAACGTGTGCATTTGATGTATATTCTTGAACTTTTTA | ATATGATA | AAATTG | VSNTNKQINNVVVNN |
| | | | | 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 |
| | | | | AATTTATCAATCTTTTTTTCGTTACGGTAAACAATAAACAATAAAATAACT | TGCTGTTT | | RRITLWRNGNFSELM |
| | | | | ATAGTTATTTTATTGTTTACCGCATATTGTTTAACTATAGTTAAACAAAGT | TTAAATTT | | REIRYIQSKINTSKKKR |
| | | | | ATTTGTTTATGGAACATTACCAGTATCTCTTGTTAAGGTAAACAACAAAA | ATGTAACG | | TFEDISRIFAKLMME |
| | | | | CATAGACGGCATCTCTTTTTAAGGTAATTAAGTATACGGCTAATAATAAA | TGTGCATT | | GKVAAALKVLDRESS |
| | | | | AATATACAGCTAATAATAAAATCTTCAATGAACATGGTTTCTATATGCAA | TGATGTAT | | GILQCSESVLKELKSK |
| | | | | AAGATGTGATCGTAGCTTTACTACCCTTAAGGGACTAAATATTCATAAA | ATTCTTGA | | HPDETPVQDNCLLYG |
| | | | | GGTCAATGTAAGATCTTTGTTTCCAATACAAATAAACAAATAAACAATGT | ACTTTTTA | | PLQNTPECLFDSIDEI |
| | | | | AGTTAACAATGAATTAACAACACCACGAATAAAAACAAGGTGGAAATTAAT | 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 |
| | | GGCACTTAGGAGCTGCACTTGGTTCGGAAGCATACAAAAAAGTGTATTG | 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 |
| | | AAAAAAGAAGAGGTTAAAAAAAATACATGCAGGACTTCCAGAAAACCTT | GGGGGCT | | | GFGDECKRFTALLAQ |
| | | AAACTTCTGGTTGAACAGGCCTGTGACAAAGGTGCCAGCAGCTGGTTAA | GCTATATT | | | KLSLKMGERYGAVIN |

| | | | | | |
|---|---|---|---|---|---|
| | | ACACCTTACCAATTAAAGAACAACATCTAGATCTGAATAAGGAAGAGTT | TTTGTTTG | | WLRTRLSMEITRASLL |
| | | TAAGGACGCACTTAGATTGAGATATAATGTGCCACTTGCCAATTTACCAT | AAGTTGTT | | CLRGSRTPFRHYNTD |
| | | CCTACTGTGCTTGTGGAGAAAAATTTGACGAGCTACACGCAATGTCATG | TATCTTAA | | DVGLENVQCGLI |
| | | CAAAAAAGGTGGCTTTGTTTGTAACAGACATGATAACATCAGAGATTTA | AACTTTAA | | (SEQ ID NO: 1402) |
| | | TTAACTGTTTTGCCTAAATAAAGTTTGTACTGATGTTCAAGCGGAGCCGCA | ACTTGTGT | | |
| | | TTTAATTCCATTGACAAATGAAAAATTTAATTTCAAAACTGCCAATACCA | TCAACCAA | | |
| | | ACGACGAAGCTAGATTGGATATAAAAGCAAAAGGGTTTTGGAGAAAAG | CCGTAAAC | | |
| | | GAGAAACTGCATTTTTTGATGTTAGAGTAACGCACGTAAACTCCAAATC | ATTAGTTC | | |
| | | CTCCAAAAAACAACCAACAAAACACATATTCCGTAGGCATGAAGATGCA | GCTGTTCG | | |
| | | AAAAAACGTGAGTATTTAGAACGAGTTCTAGAGGTTGAACACGGGACA | CTCAAATT | | |
| | | TTTACCCCATTAATTTTTGGTACGAATGGTGGGTTTGGAGACGAATGCA | ATCTACAA | | |
| | | AACGCTTCACGGCACTACTCGCACAAAAACTGTCCTTAAAAATGGGTGA | TATAAAAT | | |
| | | GCGGTACGGAGCTGTTATAAATTGGCTAAGGACACGTCTTTCCATGGAG | TTATCAAT | | |
| | | ATTACTAGAGCCTCCCTACTCTGCTTAAGAGGGTCACGAACCCCATTTAG | CTTTTTTCG | | |
| | | 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 | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| CRE | CRE-1_LSa | . | Lactuca sativa | ACATTAAATTAGAGAGGTTGATGTTTCAATGGAAGAAGATGAAATTCCA | ACATTAAA | MASSSTSSSDICLCPF |
| | | | | AGAAGCTATTTTTGTTGCCCACCAAGTGTTTGATAAAATGTCCAAACTAA | TTAGAGAG | RSFHCCPNGEVGSKG |
| | | | | TTTTTCTCTTGTTGCAGCTTTATTGTTCAAGATAATGTAGTTTGCTTAGTT | GTTGATGT | IXRMISHIKRHHLLTE |
| | | | | TGAGCGTTCCTTGTGCACACCAACAGTGTGTTGGTGTGCCATTTCCTTTC | TTCAATGG | DRKCVLREALSSDVG |
| | | | | CTTCCTTTTAACTATTGCTTCATAGCTTAAGCTTCATCTCGAGGCTTGTT | AAGAAGAT | LFMAVEETLKAFGQ |
| | | | | CTCTTGTATGGCTTCTTCTTCTACAAGTTCGAGTGATATTTGTCTGTGCCC | GAAATTCC | WMCGKCMTLHALS |
| | | | | GTTCAGAAGCTTCCATTGTTGCCCAAATGGTGAAGTGGGAAGTAAGGG | AAGAAGCT | RYCHHPDGRVXFVT |
| | | | | GATTGKCCGTATGATTTCACACATCAAAAGGCATCATCTACTTACTGAAG | ATTTTTGTT | GADGSSRYIVGILKPS |
| | | | | ATCGTAAAATGTGTTTTACGTGAAGCTCTTTCTAGTGATGTTGGTTTATTT | GCCCACCA | TKESVTNALGGLVFD |
| | | | | ATGGCGGTGGAAGAAACTTTGAAGGCCTTTGGTCAATGGATGTGTGGG | AGTGTTTG | VGLLDRVFKEPITTVK |
| | | | | AAGTGTATGACTTTGCATGCTCTTAGCCGTTATTGTCATCACCCGGATGG | ATAAAATG | SIPHSCRLAFSQALKT |
| | | | | TCGTGTGAGKTTTGTTACAGGGGCTGACGGCTCGAGTCGTTACATTGTC | TCCAAACT | ALYKVIAQPGSVDA |
| | | | | GGTATTCTAAAGCCGTCTACTAAAGAGTCGGTGACAAATGCTCTTGGAG | AATTTTTCT | WICLLLLPRCTLQVFR |
| | | | | GTTTGGTTTTTGATGTTGGGCTCCTTGATCGTGTTTTTAAAGAGCCTATC | CTTGTTGC | PKNRQECRSGNRKSL |
| | | | | ACTACTGTCAAGAGTATCCCCCATAGTTGTCGCCTTGCTTTCTCTCAGGC | AGCTTTAT | QQSSILKSLDTWGKE |
| | | | | TTTGAAAACTGCTCTTTACAAGGTGATTGCCCAACCTGGCTCGGTTGATG | TGTTCAAG | DGIRKLVQNMLDNP |
| | | | | CATGGATTTGTTTGTTACTTCTTCCTCGCTGCACACTGCAGGTGTTTAGG | ATAATGTA | EVGAMGQGGGILQK |
| | | | | CCCAAAAATAGACAAGAATGTAGGTCTGGGAATAGAAAATCCTTACAAC | GTTTGCTT | ESTSSNTNIRQCLRKV |
| | | | | AAAGCTCCATCCTGAAGTCCTTGGATACATGGGGGAAAGAGGATGGTA | AGTTTGAG | ADGHFTAAVKVLCSS |
| | | | | TCAGGAAGTTAGTTCAAAATATGTTAGACAATCCCGAGGTTGGGGCCAT | CGTTCCTT | GVAPYNGDTIKALED |
| | | | | GGGACAGGGTGGAGGCATCCTTCAGAAGGAGTCTACATCAAGTAACAC | GTGCACAC | KHPFRPPPPSMPSPIIS |
| | | | | CAACATCAGGCAGTGTCTCCGTAAGGTTGCAGATGGKCATTTTACCGCA | CAACAGTG | EPPLVADFDCVFGCIK |
| | | | | GCAGTGAAAGTGTTATGCTCATCGGGTGTTGCGCCATATAATGGTGATA | TGTTGGTG | SFPKGTSCGRDGLRA |
| | | | | CTATTAAAGCTTTGGAGGACAAACACCCTTTCAGGCCACCCCCATCCATG | TGCCATTT | QHXLDALCGEGSAIA |
| | | | | CCGAGCCCCATAATTTCTGAACCTCCCCTTGTAGCAGACTTTGACTGTGT | CCTTTCCTT | TDLIRAITSVVNLWLA |
| | | | | ATTTGGTTGCATCAAATCCTTCCCTAAAGGAACTTCWTGCGGGAGAGAT | CCTTTTTA | GRCPTILAEFVASAPL |
| | | | | GGCTTGAGGGCTCAACACWTACTAGATGCCCTTTGTGGAGAAGGGTCT | ACTATTGC | TPLIKPDNGIRPIAVG |
| | | | | GCTATAGCCACACAGATCTCATACGTGCTATCACTTCAGTGGTTAATTTATG | 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 |
| | AGCTCCTTGGGGGGGGCTGTTAGCAGAGACGCAGGGTTTATTAGCGGGC | | | 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 |
| | ACCTTTGGGGAACATGCCGGTTCATTGTAGAGAGCTCCCTGGTTTCAAGT | | | RELPGFKYRHDVVRD |
| | ACAGACATGATGTGGTTAGGGATGTTCTCTTTGATGCTTGTCGGCGTGC | | | VLFDACRRAGISAKKE |
| | TGGTATTTCTGCGAAGAAAGAAGCGCCAGTGAACTTTTTGACGGACCCG | | | APVNFLTDPQDGRST |
| | CAGGATGGAAGATCCACACTTAGACCGGCTGACATTTTGGTCTTTGGAT | | | LRPADILVFGWVGGK |
| | GGGTAGGAGGGAAGCACGCGTGTGTGGATCTTACTGGGGTCTCTCCTC | | | HACVDLTGVSPLVGL |
| | TCGTCGGTTTGAGGAGCGGGGGTTTCACAGCAGGGCATGCCGCTTTGA | | | RSGGFTAGHAALKA |
| | AAGCCGCTGCGTGCAAAGTGGCAAAGCACGAGAATGCATGTATAGAAA | | | AACKVAKHENACIEN |
| | ATCAACATGTGTTTGTACCTTTTGCATTTGATACATTGGTTTTCTCGCAC | | | QHVFVPFAFDTFGFL |
| | CAGAGGCGGTGGAGCTCCTCAACAGAGTCCAACGGGTCATGCATTCTA | | | APEAVELLNRVQRV |
| | ATGTCATATCTCCTAGATCCACAGATGTTGTTTTCAAAAGAATTAGTTTT | | | MHSNVISPRSTDVVF |
| | GCCATCCAGAAAGGGCTAGCGGCGCAGCTTGTTGCCCGTTTGCCTTCCA | | | KRISFAIQKGLAAQLV |
| | TCGATATGTATTGAACTATATTTTATATATTAAAAAAA (SEQ ID NO: 1035) | | | ARLPSIDMY (SEQ ID NO: 1403) |

| CRE | Cre-1_MB | . | Monosiga brevicollis | CATCTTGGCGTGAACCACGTTGTCAGACAAAATCTGCAACCCCGCTCTTT<br>GCGGCCCGCGTTTTGGCGGCGCCCTCGCTCCCACCGTGTCCGCTCGCTT<br>GCTCGCTTGCTTGCCCCGCGGACATGGCCACTGAGTCCGGCGGCGGAGG<br>ATTCTTGGACCCAGGTCCGCGGTGCTAAACGCCCGAGTGCCGAATCACC<br>TCCAAGCAACACCACCACCTCGCCTTCCCAAACTCATCGTTCTGCAAAAC<br>ACACAAAACATGGCAGCGCTCGCCACGACCGTAACCATGTTTTCCCTGA<br>CCCCATGACCACCCCGCTTCGCCCTCACGCCCGCCACTCTGTCCCTACCG<br>CCCGTGCCTCGTCTCATGTGCCCTCGACGTCCCCCGCTGCCGGTGCGACC<br>GAGTCTTCGGCACGTGCCGTCGTGCCCCGCGGCCGAGCCAGTGACCAGG<br>ACGTCAAACGGCGCGGCGGGGAGCAACATCCCATCATCGGAAACACTTCC<br>AATGCTTCTCCCCGCACCCCTCGCACGCCATCCAGCCCCTCGCTCCTTTGCT<br>CAAGTTGCTGCGGCAATGCCTGCCGCCGCCACTGCCACACATCTTCGGCCC<br>CTATGACCGAGGATTTGTCAGCATCGGTGCCCTCTGAGCCAAATGGCAG<br>CGGGGAGCAACAACCCTCGCCCGAGTCCACAGGGCAGACACATCATTC<br>GATTCCTAACACACCATCGGATTTTTTGACCATGTCTTCGGATGAAAGCG<br>ACTCCCCTCCTCGCTCCACCGCACTCCGCGCGCCCACCCCTATCGCCCCCTC<br>CCGCGCATGATGGTGACGGTGACACACAAACGGCAGTGCCACGCCTGAGC<br>CATTGGTGCAATCACCTACACCCGCTCAAATGGTGCTGCCATATCCATCG<br>GGTACACAACAAACCCATTCCGATCCCTCTCCGCCCTCTGCTTCACCCCCT<br>GCCACTACCATTTTGCCCGCTGCCATTTCACATCCTGTCGAACACAGTGA<br>GCATGCAAACTCAGCCCCACTTGGCGAAGTCAGTGAGAGTGAAACACA<br>CAATACAGCGGGCGAACACAGTGAGAGTGAGCAAGATGTTCTTCTCAG<br>CGATCCCGCTCCGCCCATCGCTGCCAACGTGCTGGATGCCCAGCGCAAG<br>GTCCTGCTGAAGACATCTGGCCACAGGCAACTCCTCGCCTGCCCATTTG<br>GGCTTTGCAAATGCAAGGGGCCCCGCCTTGACCGCAAAGCCTGGGTCA<br>ATCATGTACTACGCGAGCACCCCTACGACGAGCAAGCCACTGATCTGGT<br>CAAGCAGGTGATGGAGGCCAAATTGGTCGCCCAGTGCAACAAGTGTCA<br>CCTCTTCTTCGAAGCTGCTGGTATCAGTCAACACCGCTCCCGATGTGGTG<br>CCAATCTGAAGCGAGCGACCGAGGCGTTGTTTCATGCGGCTGGACACG<br>ACCTGCTTGAGATTATGCGTGGCGCTTGGCCCCAACAGTGTGTAGGGTC<br>TCGCATCAGTGTCTGCGAGCTGCTCAAGCTCGCCCGGCATCCACTGATG<br>CAGCGCAGCCGCTACCCATCCAACGCCACCGAGACCAAGCTGATGGCTG<br>CCACCCTGAGCCAGCTGTATTGGTCTGCCGTCCACTCGGACTATACCGCT<br>GAAGAGCGAGAGATGTGCTGGGCTTTGATTTTGGCCTTGCCTAGCATGT<br>TGTTGTCTGCTCCCTCGACCGCACTGTCTACGATTGACCTGCGCAATATG<br>TTTCACGATCGTCTCCGTTGGCTTGTGACGGGGCAACTAGGTCGGGTCG<br>TGGACGCCATGCGCAAGGCAGTCGCACGCAAGCAGAGCCGTCGAGGAC<br>AGCTGAACGCCGGCGCGGGCGCCCACCCGAACGACGCAGTCGACCAGA | CATCTTGG<br>CGTGAACC<br>ACGTTGTC<br>AGACAAAA<br>TCTGCAAC<br>CCCGCTCT<br>TTGCGGCC<br>CGCGTTTT<br>GGCGGCG<br>CCCTCGCT<br>CCCACCGT<br>GTCCGCTC<br>GCTTGCTC<br>GCTTGCTT<br>GCCCCGCG<br>GAC (SEQ<br>ID NO:<br>1159) | TAGGTA<br>GGCACC<br>GTCTCG<br>GGGGTC<br>CCTCTGT<br>GGGGAT<br>CCCTGT<br>GTGCAC<br>CTGTCG<br>CTCCCTA<br>GGTGGT<br>TCCTCGT<br>TGTGTCT<br>TTTGATG<br>GCTTGA<br>CTTGTAT<br>TTTTGTT<br>TTAATTT<br>TGCTTTA<br>ATTTTTG<br>CTGTATT<br>TGTGTG<br>GTATTTT<br>TGCTGA<br>ATTTTTG<br>TAAGGT<br>CCTTTGT<br>ATGATG<br>TCTTTGT<br>CTTCTGT<br>GGTCGG<br>TTGTTTT<br>CCTCAAT<br>CCGACG<br>TTGTGTC<br>TCGTTG<br>GATGTG<br>AGCGTG | MATESGGEDSWTQ<br>VRGAKRPSAESPPSN<br>TTTSPSQTHRSAKHT<br>KHGSARHDRNHVFP<br>DPMTTPLRPHARHS<br>VPTARASSHVPSTSP<br>AAGATESSARAVVPA<br>AEPVTRTSNGGGEQ<br>HPIIGNTSNASPRTPR<br>TPSSPRSFAQVAAA<br>MPAAATATSSAPMT<br>EDLSASVPSEPNGSG<br>EQQPSPESTGQTHHS<br>IPNTPSDFLTMSSDES<br>DSPPRSTALRAPTPIA<br>PPAHDGDGDTNGSA<br>TPEPLVQSPTPAQM<br>VLPYPSGTQQTHSDP<br>SPPSASPPATTILPAAI<br>SHPVEHSEHANSAPL<br>GEVSESETHNTAGEH<br>SESEQDVLLSDPAPPI<br>AANVLDAQRKVLLKT<br>SGHRQLLACPFGLCK<br>CKGPRLDRKAWVNH<br>VLREHPYDEQATDLV<br>KQVMEAKLVAQCNK<br>CHLFFEAAGISQHRS<br>RCGANLKRATEALFH<br>AAGHDLLEIMRGAW<br>PQQCVGSRISVCELLK<br>LARHPLMQRSRYPSN<br>ATETKLMAATLSQLY<br>WSAVHSDYTAEERE<br>MCWALILALPSMLLS<br>APSTALSTIDLRNMF<br>HDRLRWLVTGQLGR<br>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 |
| | | CACATTGCCAACACAACACCGTGCCCGAGTCGATTCGACATGCCCTTTGCTC | 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 |
| | | AGCTCCTTGAGGCTCGCCTCTTCACACGGGCCAACACCGGGGGGCTTTGG | TTGCGT | | AMQYVPSLINHLYTL |
| | | CCTGCACGACTTGGTTGAGCGCGGTCCGGTGGCTTATGTCTGCAACATG | GCCGTC | | PPSLTSGLSELLNRAC |
| | | GCCAAGCTGGCCACTCGCTACCCTCGGGTCTACGATCGACTTTTGGAGG | GTGATTT | | KDTFVKAFFAKVNLS |
| | | ATGCATCGAGGGCTGCCGACTTTGAGGCCCACGTGCAGCGAGCTGGCT | TGATGT | | APAGAEGHDVTLEQ |
| | | TCCAGATGGCCACGGTCAAGGACGCGGCGACCCAGCGACCAGCTGAGA | CGGGGT | | LLEARLFTRANTGGF |
| | | TCATTGCCCTCCGCTCCAAGGCGGCACTGGACGACCTGATGGCCAAGTG | TGCACA | | GLHDLVERGPVAYVC |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CGCGCTGGACCTGCAGCAGGCATATCTGGCCTCACGCGAGTGGGGCGT | | GCTTTG | NMAKLATRYPRVYD |
| | | | CAGCACTGTCTTGACCATGCGGGGTCGGGACAAGTTGCGTCGCTTGAGC | | CTTTCAG | RLLEDASRAADFEAH |
| | | | GACACGACCTTTGCCATTGCGGTCGTGTCCATGATGGGTTTTGGCCTCCA | | CTCTGA | VQRAGFQMATVKD |
| | | | TGAACTCATCAACGTCAAGCCGACGGACAAGTGCCCGCTCTGCAGCAGC | | GGTTCA | AATQRPAEIIALRSKA |
| | | | AAGACACCTCAGCCGCGACTGACCCGCGAGCACCTGCTGACCTGCCGTC | | AACACC | ALDDLMAKCALDLQ |
| | | | CCATCAAGCGTCACAACGCCCTTCGCGACGAGATGGGCCGCCTGCTCAG | | TAATTTA | QAYLASREWGVSTVL |
| | | | GTACGCCACCCTCTCCCATGTCTGGGTGGAAAAGTCTGGCTACAACGCC | | (SEQ ID | TMRGRDKLRRLSDTT |
| | | | AACGGTCAGAGCTGCCGCATCGACCTGCACTGCCGCAACCCCTTTCCCG | | NO: | FAIAVVSMMGFGLH |
| | | | GCGGTGCTCTGGGCCCAGCTCTGCCCGACCTGGGCATTGACGTGACTGT | | 1282) | ELINVKPTDKCPLCSS |
| | | | GCGCACAGCCCAACCCCCGACCACCTCGCAAGCCTGCATCAAGGTGGGC | | | KTPQPRLTREHLLTCR |
| | | | GCTGCCCTTCGCCGAGCCGAAAAGGAGAAGCGCGACTACTACACCGGT | | | PIKRHNALRDEMGRL |
| | | | TTCAACCATGGAAAAACTCTGATCGTCCCTGCGGCGATGACGACAACCG | | | LRYATLSHVWVEKSG |
| | | | GTGGGTTCGCCTCCTCCTCTTTGTGGATCTGCTTGGTCAGCTCGCCCGCTGC | | | YNANGQSCRIDLHCR |
| | | | GCCGAGGCCCGTGGTGTGTACCAGCCGGGGCTGGATGAGGCCTTTGTT | | | NPFPGGALGPALPDL |
| | | | CCTCGGTGGAAGGGTCGCTTTGCGGCGCTGGTCCATCAGATGAACGCT | | | GIDVTVRTAQPPTTS |
| | | | GACCACATCCAGCGCCACTTTGGCGGTGTCTGCCTCGCGCTCGTCGTAGG | | | QACIKVGAALRRAEK |
| | | | TAGGCACCGTCTCGGGGGTCCCTCTGTGGGGATCCCTGTGTGCACCTGT | | | EKRDYYTGFNHGKTL |
| | | | CGCTCCCTAGGTGGTTCCTCGTTGTGTCTTTGATGGCTTGACTTGTATTT | | | IVPAAMTTTGGFASS |
| | | | TTGTTTTAATTTTGCTTTAATTTTTGCTGTATTTGTGTGGTATTTTTGCTGA | | | FVDLLGQLARCAEAR |
| | | | ATTTTTGTAAGGTCCTTTGTATGATGTCTTTGTCTTCTGTGGTCGGTTGTT | | | GVYQPGLDEAFVPR |
| | | | TTCCTCAATCCGACGTTGTGTCTCGTTGGATGTGAGCGTGCCGTGGTGTT | | | WKGRFAALVHQMN |
| | | | CTTTGTGTTTGTGCTGTGATGGCTTGTAGTTGTGATGTGTGACTGCCTTT | | | ADHIQRHFGGVCLRS |
| | | | TTGGGTGTCTTGTGTTTGAAATGGCCGTATCTCTGGTTATACTTGGTCGT | | | S (SEQ ID NO: 1404) |
| | | | TTTGTACGATTTTTGTTTCTATGTGCGTGATTCTTCGCGCTTGTACTTCTT | | | |
| | | | GGCATGATAGAAGCCAATGAATGTGTCTTGTTCTCTTGTGTTGTTTTGCG | | | |
| | | | TGCCGTCGTGATTTTGATGTCGGGGTTGCACAGCTTTGCTTTCAGCTCTG | | | |
| | | | AGGTTCAAACACCTAATTTA (SEQ ID NO: 1036) | | | |
| CRE | CRE-2_HMa | Hydra vulgaris | AATTTAAAAAAAAAAAAATCGTTTATTTATGGCATAATACTGTTTGTAATT | AATTTAAA | TGAGCT | MSSCKVTIPHVCPYC |
| | | | TTTGAAAATTCGTGCAACAACTGCAGTTAAATTGAAGAGCTGAAATTTA | AAAAAAAA | CTTATAA | KVELKTICGINRHILKC |
| | | | AGATCTGAGCTTTTCAATCAGAGTTTTTTACCCTAAAACATTAAATTTTAT | ATCGTTTA | ATTTATA | KKNPLQIPSLQKTNTS |
| | | | CATAACAAAAATCGTTCTAATATTATTAAACTTAAAGAAATTCGTTCTTAT | TTTATGGC | TTATAGC | LTLEPNTKVIPSITKQ |
| | | | ATCAAATCTTATTTCAGTGTTTCACAGACGAAGGGTTTTACTAGATTTTT | ATAATACT | ATTTTGT | NDIIIASTSSNNLAFN |
| | | | ATTTTTTCAACTTTTGAATTTGTTTATTATAAAACTGTAAACTAGTGTGCA | GTTTGTAA | TTTA | QKKDYTLTPTYSRKTT |
| | | | ACCAACCGTAAAAAWTAGTTTAGCTGTTCACCCAAAATATTATTCAGTAT | TTTTTGAA | (SEQ ID | PVSILSSMKMTPISITS |
| | | | GAAAATATTTAATCTTCTTTATTCGCAGTAAACAATAAAAATATCTAGTTA | AATTCGTG | NO: | HIVRRKLPELPSQTTN |
| | | | AACAAAATATTTCTTAATAATAAAAAACAAAAACTTTTTCTTAACAAGTA | CAACAACT | 1283) | HLFNENFINVPFLPEI |
| | | | CAATGAGCAGCTGCAAAGTTACTATACCTCATGTGTGCCCTTATTGTAAA | GCAGTTAA | | MNHLPVPNNNVM |
| | | | GTAGAACTTAAAACAATATGCGGAATAAACCGTCACATTTTAAAATGCA | ATTGAAGA | | WGVYSYQQFKLFVD |

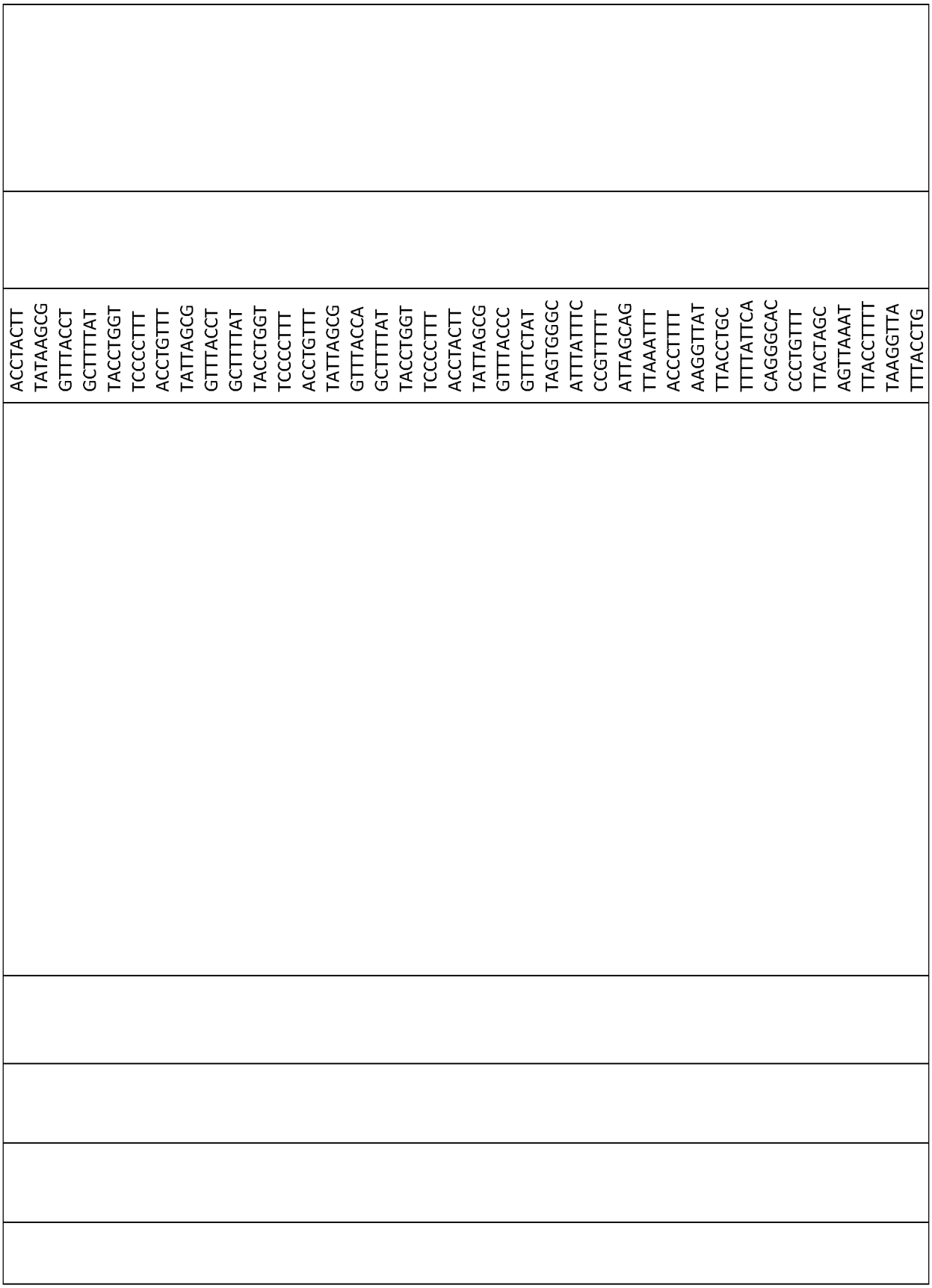
| | | | | | |
|---|---|---|---|---|---|
| | | AAAAGAATCCTTTACAAATACCCAGTCTACAAAAAACTAATACCTCTTTA | GCTGAAAT | | STYDEIVNYRRNIFNI |
| | | ACACTCGAACCAAATACTAAAGTAATACCCTCAATTACAAAACAAAACG | TTAAGATC | | PSGKAGKEFIEELTFW |
| | | ATATTATAATAGCATCCACTTCGTCTAACAACTTAGCGTTTAATCAAAAA | TGAGCTTT | | LRKFNSTSSLNSIALK |
| | | 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 |
| | | CCCCCTCGTTAGCAGAGAATGCAGGCATAGTATTTGGGTTCAGTAATAGGTTCGCAA | 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 |
| | | | ACTGACAGACTTAAGCATCGCAAAGAGAAAATAATTGCAGTTGATTCAA | 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 | MoTe R1 | JQ747 487 | CCCGAACCCGAACCCAAACCCAAACCCAAACCCAAACCCAAACCCAAAC | CCCGAACC | TAATAG | MVCPTCNGVYADYN |
| | | Magnaporthe oryzae | CCAAACCCAAACCCGGAGGGTTCCCAAGTCGCCTAAACCCGAAGGGTTT | CGAACCCA | GTAACG | DHIRKKHPDERYTAL |
| | | | AGGATATTATTTCGTTTATTAGAATTGGATAAATTATTTACCCCTGTTGGA | AACCCAAA | TCCCTAT | QLQPLGLTPCPICKTA |
| | | | CAGGGGGGTTGCAGGGGTTAAATTAAGGTTTTTTATTATTTATGCGCCG | CCCAAACC | TTTTGTC | CKNDLGVKTHLSKIH |
| | | | TTTATTTGTTTACCCCCCCCAAATATTATAAAAGCGCGTTCCATCCTCTTAG | CAAACCCA | TTTGGTT | KISGASKISTQPRIRTE |
| | | | GAAAAGCGAAGCTTTTCCTTGTAAAAGTCGCTAGACTTTTACTATAAAA | AACCCAAA | TTGTTTT | NTDNTNSVPTSSFNP |
| | | | GTCGCTAGACTTTTATACCAATCTTTTAACAAAAAGCGTAGCTTTTTGTT | CCCAAACC | TATCTTT | VLPEIQTLTPGLNNSR |
| | | | GCCAATCTATTAAAAAAAAGCGGAGCTTTTTTTAACTTTTTCTTTTTTTTTT | CAAACCCG | GTTTTTG | WADNPRKRRADTPS |
| | | | TTTTTCTTTTTTTTTTTTTTTTTTTCTTTTTTTTTTTTTTTTTTTATATAT | GAGGGTTC | TTTTTGT | PTRGRNTRPRRFSYT |
| | | | TATTATTATTATTAGCGGTGGGGCTATTTATGCGCTTTAATTTGTGC | 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 |
| | | TACCCTTTTAAGGTTATTTACCTGCTTTTATTCACAGGGCACCCCTGTTTT | TGCGCCGT | GTTTTAT | | GREKAPGLSGWTRSL |
| | | TACTAGCAGTTAAATTTACCTTTTTAAGGTTATTTACCTGCTTTTATTCAC | TTATTTGTT | TTGTTTT | | LDAAIKIPTQNDVIPA |
| | | AGGGCACCCCTGTTTTTTACCAGCAGTTAAATTTACCTTTTTTAAGGTTATTT | 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 |
| | CAAAACGCCCGTTAAAGATAAAATAGCCAAAATAATCGAAAAAAAGCCC | TTTGTGCG | | PIIISAGGLMEKDTAQ |
| | GTTTACCCTAAATTCCGCCAAAACGACAGAAACGGATATCGATACGTTA | GGGCTATT | | AYKQIQKLIGPVAAH |
| | AAAACCAATGGTTTAAAAAACGCTCGGCTCGTTTATTGGACCAACGGAAT | AATGCGCT | | WLDTSISLILLRSRTTA |
| | TACGGAAGGAATTTTTGCAAAATAAAATTCAAAATTTCGAATCGTCCATT | TTAACTTT | | AISIAKNRPRA (SEQ |
| | AACGCCCTGAAAAAAACTCCCTAAACAATACGGATTGCTAATCTTGCGTA | ACAAATTT | | ID NO: 1406) |
| | AAAGTACACAATTACTTTTACGCCATTTGCTCCGTACTTTAAATTCCCAG | TATTTATG | | |
| | GACCTGTGGGAATTATGGGAAAAAAACAGATAAATTAATAGCGGATTTC | CGCTTTAA | | |
| | GTTATAAATTTAACTGTTACAAAACGGAAAAAAACGGCCAATTACGGATT | TTGCTGCG | | |
| | TCGTTACGCCGTTAATTACGTTACCTATAAAGGACGGAGGTTTTGGATTA | GGCCTGTT | | |
| | TTACGGCATAACGGAATAGCCCAAGATATTTATTTTGCGGCCAAGGATT | AATGCGCT | | |
| | TAACAACCGAAATTCGGCACAAAATCCAACGTATATCCAACGATTTTCCA | TTAATTTA | | |
| | CAAAATCAAAGCCCTACCGCCACCGAGATTTTGCATTTGTTGCATAACG | CAAATTTC | | |
| | GGGTTTTAGCAGATTGCAAAAACGGGTTAACAAACGCCCAATTAAACGC | 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 | | |
| | | | ATTTTTGTCTTTGGTTTTGTTTTTATCTTTGTTTTTTGTTTTTGTTTTCGTTTTT | TTTATCCC | | |
| | | | TGTTTTTGTTTTTCGTTTTTTGTTTTTTTTTTTTGTTTTTTGTTTTTTGTTTTTGCCTT | GTTTCTTTT | | |
| | | | TGTTTTTGTTTTTTATCTTTATTTTTTGTTTTTTGTTTTTTACTTTGTTTTTATTTGTT | CCATCCCA | | |
| | | | TTATATTTACCTTTTGATTTTTTTCTATTTTTTCCCACCCTTATTATTATAACCC | TCGCGCGT | | |
| | | | CAACCTACTAATATTTTTTTCTTTTTTTCTTTTTTTACGGTTTTTATTTTTC | TTTCGTAA | | |
| | | | CCGTTTGTTTTTTTCTATTTTTATTTGTACGACAAAACCCTTAGCAAATAAGC | 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<br>CACGAATAGGGTGGTGACAGCTGGATAGACAGACGACAGCTCGGAAA<br>GACGGCATTGGGGCAGTATGGGTTGGCACCCCTAACTGCATCTCCCCTA<br>GGAGAGCATCCCGCAACACGCTACAAAGAACCACAAAGAGCAATACCC<br>CCAGGGATGCCCGAGAGAGGGGGGAGGATGAGCATCCCATTCGGACGG<br>TCCAATCGGTATTGACCCCAGCAAACGGAGAATCGACAATGAATGCAGT<br>CTGTGTGTGTGGCAAGGTATGTAAGAACCAGAGAGGTTTGAGAATCCA<br>CCAAACAAAGATGGCCTGCTTAAGGAGGGGTGCAGGCGGAGCACCGCTC<br>AGGGGCTGTGGCAACCACTGTAGAACCAGTGTTGTCAGCATCAGCCCCT<br>GGTCAGACGGAGGAGGATCAGGGCCCGGAAGCTCCCCACAGTGCCCG<br>GAACCTCCGCGCAACGCCTGCCCCTCCACAAGGCAGGAAGTCAGATCAT<br>CACCGAGTGAAGTGGCCAGCCGCAAACTCCAAGGAGTGGTCGCAGTTT<br>GACGAGGACGTTGACATGATCTTGGAGTCGGTGTCAAGAGGTAGTACA<br>GACCAAAAGCTTCAGTCCATGTGCACAGTGATTATGTCCATGGGGGCAG<br>AACGATTTGGCACGATTGGGCAGAGGAAACCGACAGACACAATGAAGC<br>CAAATCGCCGGGAAGTAAAGATCCGTCAACTGAGGCAGGAGCTAAAGT<br>CGTTGAGGCGGAGCTTTAAGGCGAGTACGTCGGGAGAAGCTTAGGACCCTCA | TTTTCAGT<br>CTGGCTCA<br>GCCAGTGA<br>CCGCCGGG<br>AAAGTCCG<br>GCTGACTA<br>CCACGAAT<br>AGGGTGG<br>TGACAGCT<br>GGATAGAC<br>AGACGACA<br>GCTCGGAA<br>AGACGGC<br>ATTGGGGC<br>AGTATGGG<br>TTGGCACC<br>CCTAACTG<br>CATCTCCC | TGATTA<br>AAGACC<br>CGAAAC<br>ACCCAA<br>TGACCC<br>CGGGTT<br>CATCACT<br>GATGAT<br>GTGTCC<br>CTGTTC<br>GCACTA<br>CCAGAG<br>TGTATTC<br>TAGAG<br>(SEQ ID<br>NO:<br>1285) | MNAVCVCGKVCKN<br>QRGLRIHQTKMACLR<br>RVQAEHRSGAVATT<br>VEPVLSASAPGQTEE<br>DQGPEAPHSARNLR<br>ATPAPPQGRKSDHH<br>RVKWPAANSKEWS<br>QFDEDVDMILESVSR<br>GSTDQKLQSMCTVI<br>MSMGAERFGTIGQR<br>KPTDTMKPNRREVKI<br>RQLRQELKSLRRSFKA<br>STSGEERAALAELTH<br>HLREKLRTLRRAEWH<br>KKKGKERARKRSAFIT<br>NPFGFTKRLLGQKRS<br>GNLTCPVEEINLHLSN<br>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 |
| | | | GTTCCCTGGGAGGAGCGGATAAGCGAAGCCAACGAGCGGAAGAGGGC | | | 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 |
| | | | CCAAGCCCACCACACCCCTACACTATGAACCGTAGGGCTACCCAGATACA | 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 |
| | GCTGCTTGGAACACACTGGAGTAGTTACACAACTCATCAGAGAGGCCCA | TCGGACGG | | YNNFKMRVTSGSETS |
| | TGAGAACAGAGGGGACTTGGTTGTCTTGTGGTTGGACTTGGCAAATGC | ATCAAAGG | | SWHRIGKGIITGCTIS |
| | CTATGGGTCCATACCCCACAAGCTGGTTGAGCTCGCTCTACACCGCCACC | TTA (SEQ | | VILFALAMNMVVKS |
| | ACGTTCCTAGTAAGATTAAGGACCTAATTCTGGATTACTACAATAATTTC | ID NO: | | AEVECRGPLTKSGVR |
| | AAGATGCGGGTCACATCTGGGTCAGAAACATCAAGCTGGCATCGCATC | 1163) | | QPPIRAYMDDLTITTT |
| | GGGAAAGGAATAATAACAGGCTGCACCATCTCAGTTATTCTTTTCGCTCT | | | TVPGSRWILQGLERLI |
| | CGCCATGAACATGGTGGTCAAGTCAGCCGAAGTGGAATGCAGAGGGCC | | | AWARMSFKPSKSRS |
| | CTTAACTAAGTCAGGTGTGCGACAGCCCCCTATTAGAGCATATATGGAT | | | MVLKKGKVVDKFHF |
| | GACCTTACCATCACACAACAACGGTCCCAGGGAGCAGGTGGATCTTAC | | | 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 |
| | ACCTTAGTAAGCCAGGTCAAGGCAAGGAAAGACACCACCTACTCCAG | | | QGAWTRWNTLQRRI |
| | GGAGAGGTTCGAGCAAGTGTGGAGGAAGAGAGTCAGTAGGGTGGT | | | TWANILQADFQRVR |
| | AGGACTCCGGCAGCAGGGAGCATGGACTAGGTGGAATACACTGCAACG | | | FLVQAVYDVLPSPSN |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | TAGGATCACCTGGGCGAACATCTTGCAGGCGGATTTCCAACGTGTCCGT | | | | LHVWGKNETPSCLLC |
| | | | TTCCTAGTACAAGCTGTCTACGATGTACTGCCAAGCCCATCAACCTCCA | | | | SGRGSLEHLLSSCPKA |
| | | | CGTTTGGGGAAAGAATGAGACACCTTCCTGCCTTCTTTGCTCTGGAAGA | | | | LADGRYRWRHDQVL |
| | | | GGCTCTCTAGAACATCTCCTCAGCAGTTGCCCCAAGGCTCTGGCTGATG | | | | KAIAASLASAINTSKN |
| | | | GTCGCTATCGTTGGCGCCATGACCAGGTGCTTAAGGCAATTGCTGCGAG | | | | HRAPRKAVHFIKAGE |
| | | | CTTAGCTTCAGCCATTAACACGAGCAAGAACCATCGTGCTCCAAGGAAG | | | | KPRALPQLTTGLLHK |
| | | | GCAGTCCACTTCATCAAGCTGGAGAAAAACCCCGGGCCCTCCCACAAT | | | | ASDWQLEVDLGKQL |
| | | | TAACAACAGGCCTCCTTCACAAAGCCTCGGACTGGCAGCTGGAGGTCGA | | | | RFPHHIAATRLRPDII |
| | | | CCTGGGAAAACAGCTGAGGTTTCCTCATCACATCGCTGCAACACGTCTC | | | | AISEASRQLIILELTVP |
| | | | CGTCCAGACATTATAGCTATCTCAGAAGCTTCAAGACAGCTAATTATTCT | | | | WEERIEEANERKRAK |
| | | | GGAGCTTACAGTGCCGTGGGAAGAGCGTATTGAAGAAGCAAATGAGA | | | | YQELVEECRERGWRT |
| | | | GGAAGCGCGCTAAGTACCAGGAATTAGTGGAGGAGTGCAGGGAGAGA | | | | YYEPIEIGCRGFAGRS |
| | | | GGCTGGAGAACTTACTATGAGCCCATAGAAATTGGATGCAGAGGCTTT | | | | LCKVLSRLGITGVAKK |
| | | | GCAGGGCGTTCACTTTGCAAAGTCCTCAGTCGTTTGGGCATTACAGGCG | | | | RAIRSASEAAEKATR |
| | | | TGGCGAAGAAAAGGGCCATTCGATCCGCAAGCGAAGCCGCAGAGAAG | | | | WLWIKRADPWTAV |
| | | | GCCACAAGGTGGCTGTGGATTAAGAGGGCAGATCCGTGGACTGCTGTT | | | | GTQVGT (SEQ ID |
| | | | GGGACACAAGTCGGGACTTGATCAACCCCGGCTGGGTCACCTGGGTGA | | | | NO: 1408) |
| | | | GAGTGTATGATGTTGAGAGACCCGAAACACTCAATGATCCCAGGATACA | | | | |
| | | | TCACTGATGATGTGTCCCAAATGCATCCATGAGATGTTTCTTGCATAA | | | | |
| | | | (SEQ ID NO: 1040) | | | | |
| HERO | HERO-3_BF | . | Branchiostoma floridae | CTGACCAGCAGACGGGAAGCCCGCGACCAACTAGTCTCCGCAAATATTG | CTGACCAG | TAGAAA | MALPAVRSGPASTW |
| | | | | CACACAGGGCGACCCTATGGAGCTGATTCAGTCAAATTTCCTCTGAGAT | CAGACGG | CCCACA | TLLITLVIVAAKGTDG |
| | | | | ATACCGATAACTATCTACAGAAACTGCACAGTTAGTTTGGAAAGAGCTT | GAAGCCCG | AGGCTG | FMSFKLPLLSTDTWS |
| | | | | TTCTACTGAAAGACAGCAAAATCCGCCACTTTAGACGAGCGTCAAGACT | CGACCAAC | AGAAAT | GYNNDVKTLLGPLHH |
| | | | | GCCCTCCCCATAACCAATATGGCGCTACCTGCTGTACGTTCTGGACCAGC | TAGTCTCC | GTAGAG | ELATNEMSPKLAGEG |
| | | | | CAGCACCTGGACACTGTTAATCACGCTGGTCATCGTCGCTGCTAAAGGT | GCAAATAT | CATCTGT | FSDIMCDFMASKPEF |
| | | | | ACAGATGGTTTTATGTCTTTTAAACTGCCACTGCTGTCTACTGATACCTG | TGCACACA | ATGGAC | SHTTEESHSEGYISHE |
| | | | | GTCTGGGTATAACAATGATGTGAAAACCCTGCTAGGCCCGCTCCACCAC | GGGCGAC | AATATT | PQSLAQVKRLKNKLR |
| | | | | GAACTGGCCACAAATGAAATGTCCCCCAAACTAGCTGGGGAGGGGATTC | CCTATGGA | GATGAT | KKAFRADATPEDRKA |
| | | | | AGTGACATCATGTGCGACTTTATGGCCAGTAAACCAGAGTTCAGCCACA | GCTGATTC | TGAAAT | FRDAIKTYSFMKRQQ |
| | | | | CTACCGAAGAAAGTCACTCAGAAGGCTATATAAGCCACGAACCACAGTC | AGTCAAAT | GTTGTG | KRKETTKSAAHQEKE |
| | | | | TCTCGCACAAGTAAAAACGCCTGAAAAACAAGCTACGTAAGAAGGCATTC | TTCCTCTG | ATTTTAG | YHKNFWKFAGKCAK |
| | | | | AGAGCTGACGCAACACCTGAGGATCGAAAGGCTTTCAGAGATGCAATT | AGATATAC | ATCAAA | GQLDIPPVKPAFSVYY |
| | | | | AAAACATACTCCTTCATGAAGCGACAACAGAAACGAAAGGAAACTACA | CGATAACT | TTTAGA | ANEYYKNKYSHPTRV |
| | | | | AAATCGGCAGCACACCAAGAGAAAGAATATCATAAGAACTTTTGGAAG | ATCTACAG | AATATG | DFNKLLWFPHLPVEE |
| | | | | TTTGCCGGAAAATGTGCAAAGGACAGCTCGATATCCCCCCAGTAAAAC | AAACTGCA | AAAACC | QLPANSFDMSPVRP |
| | | | | CGGCATTCTCTGTTTATTATGCAAATGAGTACTACAAAAACAAATACTCA | CAGTTAGT | GAACTA | KDIKAVLSKRCATSAP |
| | | | | CACCCAACCCGTGTTGACTTCAACAAACTGCTCTGGTTTCCTCATTTGCC | TTGGAAAG | AACTAA | GPDGIMYGHLKHLP |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 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 |
| | | GACTTGCAATGATCTGATAGACCCAGAAACACAAAAGGCATTCCTGACC | CCTCCCCA | GCAATA | | RRTVHITWFDLADAF |
| | | GGAATCAACGGCTGTGTGGAGCATGTCCAAGTTATGCGGGAGATCTTA | TAACCAAT | CTATCTA | | GSVEHELIYYQMERN |
| | | GCACATGCCAAGAAAAACCGCCGAACAGTCCACATTACATGGTTTGACC | (SEQ ID | TGTTATG | | GFPPIITTYIKNLYSRL |
| | | TCGCGGATGCCTTTGGTTCTGTAGAACACGAACTGATCTACTACCAGAT | NO: 1164) | TCCTTTG | | KGKVKGPGWESDPF |
| | | GGAGAGAAACGGCTTCCCGCCAATTATCACCACGTACATTAAAAACCTG | | TCCCCCC | | PFGRGVFQGDNLSPII |
| | | TATTCTCGCCTGAAAGGGAAAGTGAAGGGTCCAGGCTGGGAAAGTGAT | | TGCATG | | FLTVFQPILQHLKGVE |
| | | CCGTTCCCGTTCGGAAGAGGAGTGTTCCAAGGAGACAACTTGTCACCCA | | 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 |
| | | ACAAGGGGTCTCAACTTCAAAGCATCTCAGACTTGTACCTAGAAGCCC | | 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 | | | | DTIIESIDTSRINATIYA |
| | | | ATCTATGCGGACATACAAGGTTACACAACTAACGGAGGTACCATCCCGG | | | | DIQGYTTNGGTIPVH |
| | | | TCCATACAATACCCACTACCCAGAAACCAGACCTGATCATATATTTACCA | | | | TIPTTQKPDLIIYLPEQ |
| | | | GAACAGAAGACCCTCCACATCCATGAACTGACTGTACCCTTTGAAAAGA | | | | KTLHIHELTVPFEKNIK |
| | | | ACATCAAAACAAGTCATGACCGAAAGGTCAACAAATACAGCACCCTAGC | | | | TSHDRKVNKYSTLAA |
| | | | GGCAGATTTAGAAACTGCTGGCATTTCCGCTACACTAACCTGCTTTGAA | | | | DLETAGISATLTCFEV |
| | | | GTCGGATCAAGGGGACTCGTCACGCCAGAGAACAAGACCAGGCTTAGA | | | | GSRGLVTPENKTRLR |
| | | | ACACTGTTCAAAATAGTTAAAGCCAAACCACCGAAGACTCTGTTTACTGA | | | | TLFKIVKAKPPKTLFT |
| | | | TATAAGCCGCATTGCGATGTTATCGTCATATGCTATTTGGAACTCACGCC | | | | DISRIAMLSSYAIWNS |
| | | | ACGAACCGTATTGGGAGTCAGAAACGCTATTGTAGAAACCCACAAGGCT | | | | RHEPYWESETLL |
| | | | GAGAAATGTAGAGCATCTGTATGGACAATATTGATGATTGAAATGTTGT | | | | (SEQ ID NO: 1409) |
| | | | GATTTTAGATCAAATTTAGAAATATGAAAACCGAACTAAACTAAATATA | | | | |
| | | | ATGTTTTTTTTAAAGTAATGATAAGCAATACCCACATTGTGCAATACTAT | | | | |
| | | | CTATGTTATGTCCTTTGTCCCCCCTGCATGTTTGGTCAATAATGACCATCG | | | | |
| | | | TGTCCTGGGCTCCGTGTACCTTTCTTTACTATGAATAAAGAATGATTTTA | | | | |
| | | | CTAC (SEQ ID NO: 1041) | | | | |
| HERO | HERO Dr | . | Danio rerio | AAAGCAGTAGAGATGACGACACATCGCGCGCAGAAGTTACAACTTCTGGT | AAAGCAGT AGAG (SEQ ID NO: 1165) | TAGCAT | MTTHRAEVTTSGKT |
| | | | | AAGACGCAGGAGGAGCCAGGCCCGGAGGCAACCCACAGTGCCCAGAG | | GCCACTT | QEEPGPEATHSAQSL |
| | | | | CCTCCTAGTGTCGCCAACACCTGCTGCCGGCCGCTCGCCTGCTACTCAAA | | GGACAC | LVSPTPAAGRSPATQ |
| | | | | GCTGCCCTCAAGTGACAGCAGCTCATAACAGTCCACAAAGCCCCCAAAG | | AGGCCG | SCPQVTAAHNSPQSP |
| | | | | TCAGCAAGTGGCAGTTACAAGATCTGACTGTGTTCCCTTGGCACAGCCA | | GGGTCT | QSQQVAVTRSDCVP |
| | | | | AGAATCCAGTGGCCCCAATCCTCCAAAGAAAGCTGAGTGGCTCCAGTTCG | | GATCAG | LAQPRIQWPQSSKKA |
| | | | | ACAAGGACGTGAATCAGATCCTGGAAGTGACAGGCAAGGGGGGGTGTG | | CCTCGG | EWLQFDKDVNQILEV |
| | | | | GACCAGCGACTGTCAACAATGACCACGCTCATAGTGAACATTGCAGCTG | | TCGGGT | TGKGGVDQRLSTMT |
| | | | | AGCGATTCGGAACTGTGACACCCAAACCCACTCCATCGACATATACTCCA | | CGCCTG | TLIVNIAAERFGTVTP |
| | | | | AGCCACAGAGTAAAGGAAATCAAACGTCTCAGGAAAGAACTTAAGCTA | | GAGGAG | KPTPSTYTPSHRVKEI |
| | | | | CTAAAGAGGCAGTACAAGGCAGCAGGGGAAGTAGAAAGAGCGGGCCT | | GGTGTC | KRLRKELKLLKRQYKA |
| | | | | AGAAGATCTGAGAGGAATCCTGAGGAAACAGCTCGTGAACCTATGTAG | | TGTTGC | AGEVERAGLEDLRGI |
| | | | | GGCAGAGTATCACAGGAAGAGGCGGAGAGAGAGCAAGGAAAAGG | | AAGACC | LRKQLVNLCRAEYHR |
| | | | | GCAGCATTTTTGGCCAACCCTTTCAAGTTGACCAAGCAGCTCCTTGGCCA | | CGAAAC | KRRRERARKRAAFLA |
| | | | | AAAGAGGACTGGCAAACTCACCTGCTCCAAGGAGGCTATCAACAATCAC | | ACCCTGT | NPFKLTKQLLGQKRT |
| | | | | CTCAAGGCCACTTATTCTGACCCGAATAGAGAACAACCCCTGGGGCCTT | | GAGCCC | GKLTCSKEAINNHLK |
| | | | | GCGGTGCACTGCTGACACCACCTGAGCCCACATCAGAGTTCAACATGAA | | AGGAAA | ATYSDPNREQPLGPC |
| | | | | GGAACCCTGCCGGAGTGAAGTAGAGGAAGTGGTGAGGAGAGCAAGGT | | CAACAC | GALLTPPEPTSEFNM |
| | | | | CAAGCTCAGCACCAGGCCCAAGCGGAGTGCCTTACAAGGTATATAAGA | | TGATGA | KEPCRSEVEEVVRRA |
| | | | | ACTGCCCAAAGCTTCTACACAGGCTCTGGAAGGCCCTGAAAGTCATATG | | TGTGTC | RSSSAPGPSGVPYKV |
| | | | | GAGAAGAGGGAAGATTGCCCAGCCATGGAGGTATGCGGAGGGAGTGT | | CAAGGT | YKNCPKLLHRLWKAL |
| | | | | ACATCCCAAAAGAGGAGAAGTCGGAGAACATCGACCAGTTTCGAGTCA | | TGTGCA | KVIWRRGKIAQPWR |

| | | | | | |
|---|---|---|---|---|---|
| | | TCTCCTTGCTCAGTGTGGAGAGCAAAATATTCTTCAGCATTGTGGCCAAA | | TCAGGA GATGTTT CTGTAA C (SEQ ID NO: 1288) | YAEGVYIPKEEKSENI |
| | | AGACTCTCCAACTTCCTATTGAGCAATAAATACATCGACACGTCTATGCA | | | DQFRVISLLSVESKIFF |
| | | GAAGGGAGGCATACCAGGAGTCCCAGGCTGCCTGGAACACACAGGCGT | | | SIVAKRLSNFLLSNKYI |
| | | GGTAACTCAGCTCATTAGGGAGGCAAGAGAAGGCAGGGGGGGACCTGG | | | DTSMQKGGIPGVPG |
| | | CTGTGTTGTGGTTGGATCTCACCAATGCCTATGGCTCAATACCCCACAAG | | | CLEHTGVVTQLIREA |
| | | CTGGTGGAGGTCGCACTGGAGAAACATCATGTACCCCAGAAGGTGAAA | | | REGRGDLAVLWLDLT |
| | | GACCTCATCATCGACTATTACAGCAAGTTCAGCTTGAGAGTCTCCTCTGG | | | NAYGSIPHKLVEVALE |
| | | CCAGTTAACATCAGATTGGCACCAGCTTGAGGTAGGAATAATCACTGGT | | | KHHVPQKVKDLIIDYY |
| | | TGCACCATCTCAGTGACCCTCTTTGCACTGGCAATGAACATGATGGTCAA | | | SKFSLRVSSGQLTSD |
| | | AGCAGCTGAGACAGAGTGCAGAGGCCCCCTCAGCAAGTCCGGAGTAAG | | | WHQLEVGIITGCTISV |
| | | GCAACCTCCCATCAGAGCCTTCATGGACGACCTCACAGTGACACAACAACG | | | TLFALAMNMMVKA |
| | | TCGGTACCAGGAGCAAGATGGATCCTCCAAGGGTTGGAGAGGCTCGTG | | | AETECRGPLSKSGVR |
| | | GCATGGGCACGCATGAGCTTCAAACCTGCAAAATCCAGATCCTTGGTGC | | | QPPIRAFMDDLTVTT |
| | | TTAGGAAAGGCAAAGTCAGAGATGAGTTCCGCTTCAGGCTGGGACAAC | | | TSVPGARWILQGLER |
| | | ACCAAATCCCATCAGTCACTGAGAGACCAGTAAAGAGTCTCGGGAAGG | | | LVAWARMSFKPAKS |
| | | CCTTTAACTGTAGCCTCAATGACAGAGACTCCATCAGGGAAACCAGCAC | | | RSLVLRKGKVRDEFR |
| | | TGCCATGGAGGCTTGGTTGAAAGCAGTGGATAAATCAGGGCTCCCTGG | | | FRLGQHQIPSVTERP |
| | | AAGATTTAAGGCTTGGGTTTACCAACATGGAATCCTTCCAAGACTCCTCT | | | VKSLGKAFNCSLNDR |
| | | GGCCCTTGCTAATCTATGAGGTCCCCATGACTGTGGTTGAAGGTTTTGA | | | DSIRETSTAMEAWLK |
| | | ACAAAAGGTGAGCAGCTATCTACGCAGATGGCTGGGATTGCCACGCAG | | | AVDKSGLPGRFKAW |
| | | CCTAAGTAACATCGCTCTGTATGGGAACACCAACAAGCTCAAACTTCCTT | | | VYQHGILPRLLWPLLI |
| | | TTGGCTCAGTCAGGGAGGAGTTCATTGTGGCACGGACACGAGAACATC | | | YEVPMTVVEGFEQK |
| | | TGCAGTACTCTGGATCCAGAGATGCGAAAGTGTCCGGGGCAGGGATTG | | | VSSYLRRWLGLPRSLS |
| | | TCATCAGGACAGGGAGAAAGTGGAGGGCAGCAGAGGCAGTCGAACAA | | | NIALYGNTNKLKLPFG |
| | | GCGGAAACCCGGCTGAAGCACAAGGCCATCCTGGGGGCAGTAGCACAA | | | SVREEFIVARTREHLQ |
| | | GGCAGAGCTGGACTTGGGAGCCTAGCAGCAACCCGATACGACTCGGCC | | | YSGSRDAKVSGAGIVI |
| | | AGTGGGAGGGAGAGGCAGAGGCTGGTGCAGGAGGAGGTGCGTGCTTC | | | RTGRKWRAAEAVEQ |
| | | AGTTGAGGAGGAGAACCAGCAGAGCAGTGGCCATGCGGCAACAAG | | | AETRLKHKAILGAVA |
| | | GTGCCTGGATGAAGTGGGAGCAGGCGATGGAGCGGAATGTCACCTGG | | | QGRAGLGSLAATRYD |
| | | AAGGACATCTGGACATGGAACCCCCTGAGAATCAGGTTCTTGATCCAAG | | | SASGRERQRLVQEEV |
| | | GGGTCTACGACGTTCTTCCCAGCCCATCGAACCTGTACATATGGGGCAG | | | RASVEEERTSRAVAM |
| | | AGTAGAGACACCTGCATGCCCGCTGTGTTCCAAGCCAGGGACACTAGA | | | RQQGAWMKWEQA |
| | | ACATATTTTGAGCAGCTGTTCCAAGGCACTAGGTGAAGGTCGGTATCGA | | | MERNVTWKDIWTW |
| | | TGGAGACACGATCAGGTCCTTAAATCCATTGCTGAGGCAATCAGCAAGG | | | NPLRIRFLIQGVYDVL |
| | | GGATCAAGGACAGTCGATACCGCCAAGCCACGGCCAAGGTCATTCAGT | | | PSPSNLYIWGRVETP |
| | | TCATCAAGGAAGGACAAAGGCCAGAGAGAACAGCAAAGAACTGCTCTG | | | ACPLCSKPGTLEHILS |
| | | CTGGGTTGCTCTCCACGGCCCGAGACTGGGTGATGACAGTTGATCTTGA | | | SCSKALGEGRYRWR |
| | | GAGGCAGCTAAAGATTCCACCACACATCACCCAGTCTACGTTGAGACCT | | | HDQVLKSIAEAISKGI |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GACATAATCTTGGTCTCTGAGGCCACAAAGCAATTAATCCTGCTGGAGC<br>TGACGGTGCCCTGGGAGGAGAGGATGGAGGAGGCTCAGGAGAGAAA<br>GAGGGGAAAATATCAGGAGCTAGTGGAGCAATGTAGGGCGAATGGAT<br>GGAGGACCAGGTGCATGCCAGTGGAAGTGGGCAGTAGGGGATTTGCC<br>AGCTACACCCTGAGCAAGGCCTATGGTACACTGGGAATAACAGGCACA<br>AACCGAAGAAGAGCCCTAAGCAACAACGTGGAAGCAGCGGAAAAAGC<br>ATCCAGATGGCTCTGGTTGAAGAGGGGGGAACAGTGGGGGCAGTAGC<br>ATGCCACTTGGACACAGGCCGGGGTCTGATCAGCCTCGGTCGGGTCGC<br>CTGGAGGAGGGTGTCTGTTGCAAGACCCGAAACACCCTGTGAGCCCAG<br>GAAACAACACTGATGATGTGTCCAAGGTTGTGCATCAGGAGATGTTTCT<br>GTAAC (SEQ ID NO: 1042) | | | KDSRYRQATAKVIQFI<br>KEGQRPERTAKNCSA<br>GLLSTARDWVMTVD<br>LERQLKIPPHITQSTLR<br>PDIILVSEATKQLILLEL<br>TVPWEERMEEAQER<br>KRGKYQELVEQCRA<br>NGWRTRCMPVEVG<br>SRGFASYTLSKAYGTL<br>GITGTNRRRALSNNV<br>EAAEKASRWLWLKR<br>GEQWGQ (SEQ ID NO: 1410) |
| HERO | HEROFr | . | Takifugu rubripes | AGACTAGGTGACAACCAAGAACAGTTWGGTCGACTACTGGAAAGACA<br>GTTGGCAGCTCGGAAAGACGGCACCCGGGACAGTATGGGTTAGCACCC<br>CAGCCTGTATCTTTCGCGAGAAGGAACCCAAACAAGCTACGGAAAGCCC<br>TACAGAGAAACACCCCCAGGAGATCCCGAGAGGGGGGGAGGATGAGA<br>TCTCCAATCGGACGGACCTAACGTTAATGACCCCTGCAATGGAAATGAC<br>TACGACAGTAACATGTATCTGCAGCAAGCTGTGCAAGAACCAGCGTGGC<br>TTAAAGATCCATCAGGCCAGAATGAAATGTCTGGAGCGGGAGGTTGAG<br>GTGCAACGCACAGGTCCTGGACCTGGTGAGACGCAGGAGGAGCCCGG<br>ACAGGAGGCAACCCACAGATCCCAGTCCCTCCACGTACCGGAGCCTCCC<br>AACCCTAACAGAGTAGTTCAACAGCAGCGGATTAAGTGGCCCCCAGCAA<br>ATAGACGGAGTGAGTGGCTGCAGTTTGATGAGGATGTGTCCAACATCA<br>TCCAAGCCACAGCCAAAGGAGAGATGTCGACAGCAGACTCCAGGCGATAA<br>GTACCATCATCGTCAGCTATGGCTCAGAAAGATTTGGACGGATCGAGAA<br>GGGCAACACTGAGACCACCTCTTACACCATGAACCGCAGGTCCTTTAAG<br>ATACACCAACTGCGCAAGGAGCTGCGAACCCTCAAGAAACAGTTCAAG<br>AGAGCTKCTGATGGGGACACAAGCAAGCTTTAAAAGAGCTGTATAACATCC<br>TGCGGAAGAAGTTGAAAACTCTCCGCAGAGCAGAGTGGCACAGGAGGC<br>GCGGGAGAGAGAGCAAGGAAGCGAGCAGCCTTCATTGCCAATCCCT<br>TCCGGTTTTCTAAACAGCTGCTCGGGGACAAGCGGAGTGGCCGACTTGA<br>GTGCTCAAGGGAGGAAGTGAATCGCTTCCTCCAAAACACCATGAGCGA<br>CCCACTGAGGGGTCAAGACCTAGGACCCAACAGAGCGCTCATCAGCCCT<br>GCCCCACCATCGGCAGAGTTCAAGCTGGCAGAGCCTAGTTTGAAGGAG<br>GTTGAAGAAGTCATCAAGGCAGCCCGTTCTGCATCTTCCCCGGGCCCCA<br>GTGGTGTACCTTACCTCGTCTACAAGCGCTGTCCAGAAATTCTCCGGCAT<br>CTGTGGAAGGCCTTGAAAGTGATCTGGCGAAGGGGGGAGAGTAGCCGA | AGACTAGG<br>TGACAACC<br>AAGAACA<br>GTTWGGT<br>CGACTACT<br>GGAAAGA<br>CAGTTGGC<br>AGCTCGGA<br>AAGACGG<br>CACCCGGG<br>ACAGTATG<br>GGTTAGCA<br>CCCCAGCC<br>TGTATCTT<br>TCGCGAGA<br>AGGAACCC<br>AAACAAGC<br>TACGGAAA<br>GCCCTACA<br>GAGAAAC<br>ACCCCCAG<br>GAGATCCC<br>GAGAGGG<br>GGGGAGG<br>ATGAGATC | TGATCA<br>CCCCGG<br>CTGGGT<br>CGCCTG<br>GGCGAG<br>GGTGTA<br>TGATGT<br>CGTGAG<br>ACCCGA<br>AACACC<br>CTATGA<br>ACCCAG<br>GATACA<br>TCCTGAC<br>GATGTG<br>TCCCAGT<br>GCATCC<br>AGGAGA<br>TGTAKCT<br>TTAAGT<br>(SEQ ID NO: 1289) | MTPAMEMTTTVTCI<br>CSKLCKNQRGLKIHQ<br>ARMKCLEREVEVQR<br>TGPGPGETQEEPGQ<br>EATHRSQSLHVPEPP<br>NPNRVVQQQRIKWP<br>PANRRSEWLQFDED<br>VSNIIQATAKGDVDS<br>RLQAISTIIVSYGSERF<br>GRIEKGNTETTSYTM<br>NRRSFKIHQLRKELRT<br>LKKQFKRAXDGDKQ<br>ALKELYNILRKKLKTLR<br>RAEWHRRRGRERAR<br>KRAAFIANPFRFSKQL<br>LGDKRSGRLECSREE<br>VNRFLQNTMSDPLR<br>GQDLGPNRALISPAP<br>PSAEFKLAEPSLKEVE<br>EVIKAARSASSPGPSG<br>VPYLVYKRCPEILRHL<br>WKALKVIWRRGRVA<br>DQWRCAEGLWIPKE<br>EDSKNINQFRTISLLS<br>VEGKVFFSIVSRRLTE |

| | | | | |
|---|---|---|---|---|
| | CCAGTGGAGGTGTGCTGAGGGACTTTGGATACCCAAGGAGGAGGACTC | TCCAATCG | | FLLKNNYIDTSVQKG |
| | GAAAAACATCAACCAGTTTCGGACTATCTCACTACTGAGTGTGGAAGGG | GACGGACC | | GIPGVPGCLEHNGVV |
| | AAGGTGTTTTTTAGCATCGTCTCCCGAAGACTGACCGAGTTTCTCCTCAA | TAACGTTA | | TQLIREAHESKGELAV |
| | GAACAACTACATCGACACTTCAGTGCAGAAGGGTGGGATCCCTGGAGT | (SEQ ID | | LWLDLTNAYGSIPHK |
| | CCCCGGCTGTCTAGAGCACAATGGTGTAGTCACACAGCTCATCAGAGAG | NO: 1166) | | LVELALHLHHVPSKIK |
| | GCCCATGAGAGCAAAGGAGAACTAGCCGGTTTTGTGGTTGGACCTGACT | | | 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 |
| | AGTAACACAACAGTGGAATCCTTTGAAAGGAAGATCAGCAGCTTTCTGCGC | | | 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 |
| | GCTCCGAAGAAGGCAATCTCCTTCATAAAAGCTGGAGAGAGACCTCGTG | | | KAGERPRAGPQITTG |
| | CAGGCCCACAGATAACAACGGGACTCCTCCACACAGCTMCTGATTGGC | | | LLHTAXDWQLHVDL |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | AACTGCACGTTGACCTGGGAAAACAACTGATATTCCCCCAGCACATCGC<br>AACAACGTCTCTACGGCCAGACATGATCATCATCTCAGAGGCTTCGAAA<br>CACCTGATCATGCTGGAGCTTACAGTGCCCTGGGAAGAGCGGATTGAG<br>GAAGCCAACGAAAGGAAACGTGCCAAGTATCAGGAGCTGGTGGAGGA<br>GTGCAGGGGCAGGGGCTGGAGGACCTTCTACGAGCCCATAGAAGTTGG<br>CTGTAGAGGCTTTGCAGGACGCTCCCTCTGCAAAGCCTTTGGCCGACTG<br>GGAGTCACAGGGACAGCCAAAAAGAGGGCCATTAAAKCCGCGAGTGA<br>AGCTGCAGAGAGAGCCACGAGGTGGSTGTGGCTKAAAAGGGCAGATCC<br>GTGGGTTGCTACTGGGACACAAGCCGGGTCTTGATCACCCCGGCTGGG<br>TCGCCTGGGCGAGGGTGTATGATGTCGTGAGACCCGAAACACCCTATG<br>AACCCAGGATACATCCTGACGATGTGTCCCAGTGCATCCAGGAGATGTA<br>KCTTTAAGT (SEQ ID NO: 1043) | | | | GKQLIFPQHIATTSLR<br>PDMIIISEASKHLIML<br>ELTVPWEERIEEANE<br>RKRAKYQELVEECRG<br>RGWRTFYEPIEVGCR<br>GFAGRSLCKAFGRLG<br>VTGTAKKRAIKXASE<br>AAERATRWXWLKRA<br>DPWVATGTQAGS<br>(SEQ ID NO: 1411) |
| HER<br>O | HEROT<br>n | . | Tetraod<br>on<br>nigrovir<br>idis | AGATTGGTCTGGCTAAGCCAGTGACGTCCAGGAACAGACTGGCTGACG<br>ACCACGAATAGAGTGGTGACAGCTTGGATAGACAGCTGACAGCAGGGA<br>AAGACGGCAACCGGGGCAGGAAGGGCTAGCAACCCAGCCTGCATCTTC<br>CGTGAGGAAGAACCCAAAACTTGCTACGAAGAGCCCGAAGCAAAGATA<br>CCCCCAGGGGAGCCCGAGAGGGGGGGAGAATGAGCTCCCCAAACGGA<br>CGGATAACATGGCAACGACCCAGGCTAGCGTTAAACCGACAGCGGTTG<br>CCACATGTGTATGTGGCAAAATCTGCAAAAACCCACGAGGTCTGAAGAT<br>CCACCAGACCAAGATGGGGGTGCTTGGCAAGTGTGCAACCAGAGCAGCG<br>CGCAAGGTTCAGCCTCAGCGAGTCGCGGGAGGTGCCAGCCAGGGCCGA<br>GCCCTATGGCCCTCAGCAACCGCATTCTCCTGAGGCCCTTGGTGAGACG<br>CAGGAGGAGCGGGGCCAGGAGTCACCCCACAGTGCCCAGAACCTCCGT<br>GCTCAGGTAGCACAAGCGCCAGACAACCCACAACACCACCGGCGGGTT<br>AAGTGGCCCCCAGCCAGCAAAGTGAGCGAGTGGCAGCAGCTTGATGAG<br>GATTTGGAAGGTATTCTGGAGTCCACCGCAAAAGGTGGAGTAGACAGA<br>AAACTCCAAACAATGACCACGCTGGTCATCAGCTTTGCCACCGAGAGAT<br>ATGGTACAATGGAGAAAACGCGCTGCTCCAGAGAAGTACACCACAAAACC<br>GCAGGGCAGAAAAGATCTCCCAACTGCGGCAGGAACTTCGGGTCCTGA<br>AAAAGCAGTTCAAGGGCGCCAGCGAGGATCAGAAGCCAGGATTGGCA<br>GAGCTTCGTTGCACCCTTAGGAAAAAACTGCTTACCCTTCGCCGAGCAG<br>AGTGGCACCGGAGACGGGCCAAGGAAAGAGCCAAGAAACGCGCTGCA<br>TTTTTAGCCAACCCTTTTGGGTTCACTAAACAACTTTTAGGCCAGAAGCG<br>TAGCGCCCACTTGGAATGTGCAAAAGAGGAGGTTGATTCCTACCTCCAC<br>GACACATTCAGTGACGCAGAACGGGAGAACAGCCTAGGCGAATGTAGA<br>GTGCTGATCAGTCCACCTGAGCCAGCCTGCAGTTTCAACACCAAGGCTC<br>CAACTTGGAAAGAAATCCAAACTGTGGTCAGGGCTGCAAGAAACAACT<br>CAGCTCCTGGACCCAATGGAGTCCCATATCTGGTGTACAAAAGATGCCC | AGATTGGT<br>CTGGCTAA<br>GCCAGTGA<br>CGTCCAGG<br>AACAGACT<br>GGCTGACG<br>ACCACGAA<br>TAGAGTGG<br>TGACAGCT<br>TGGATAGA<br>CAGCTGAC<br>AGCAGGG<br>AAAGACG<br>GCAACCGG<br>GGCAGGA<br>AGGGCTA<br>GCAACCCA<br>GCCTGCAT<br>CTTCCGTG<br>AGGAAGA<br>ACCCAAAA<br>CTTGCTAC<br>GAAGAGC<br>CCGAAGCA<br>AAGATACC<br>CCCAGGG | TGATCA<br>CTCCCA<br>GTCGGG<br>TCGCCT<br>GGGTGA<br>GGGGGT<br>CTGATG<br>TTGAAA<br>GACCCG<br>AAACCC<br>CCGATG<br>ACCCCA<br>GGTACT<br>ATCACT<br>GACGAT<br>GTGTCC<br>AAGACA<br>TGCATC<br>AATAGG<br>TGTATTT<br>AGAAAT<br>C (SEQ ID<br>NO:<br>1290) | MATTQASVKPTAVA<br>TCVCGKICKNPRGLKI<br>HQTKMGCLASVQPE<br>QRARFSLSESREVPA<br>RAEPYGPQQPHSPEA<br>LGETQEERGQESPHS<br>AQNLRAQVAQAPDN<br>PQHHRRVKWPPASK<br>VSEWQQLDEDLEGIL<br>ESTAKGGVDRKLQT<br>MTTLVISFATERYGT<br>MEKRAAPEKYTKNR<br>RAEKISQLRQELRVLK<br>KQFKGASEDQKPGLA<br>ELRCTLRKKLLTLRRA<br>EWHRRRAKERAKKR<br>AAFLANPFGFTKQLL<br>GQKRSAHLECAKEEV<br>DSYLHDTFSDAEREN<br>SLGECRVLISPPEPAC<br>SFNTKAPTWKEIQTV<br>VRAARNNSAPGPNG<br>VPYLVYKRCPKLLARL<br>WKILRVIWRRGKVA<br>HQWRWAEGVWVP<br>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 |
| | | GTGCAGGCCTGGGGAACAACACCGACCACCCACCTCAGCAGGCTCAAGG | | | DQVQLEVRASIEEQR |
| | | GCAAGGAAAGGCGGGATCAGGTCCAACTAGAAGTGAGGGCCAGTATT | | | ASQWVGLRQQGAW |
| | | GAGGAACAGCGAGCTAGTCAGTGGGTGGGGCTGAGGCAGCAAGGCGC | | | TRWEEAMARKISWP |
| | | TTGGACTAGGTGGGAAGAGGCCATGGCCAGAAAGATCTCATGGCCTGA | | | ELWRAEPLRIRFLIQS |
| | | GCTGTGGAGGGCTGAGCCCTTGCGCATCCGCTTCCTTATTCAGTCAGTTT | | | VYDVLPSPSNLFLWG |
| | | ATGACGTCTTGCCCAGCCCATCAAACCTCTTCCTGTGGGGCAAGGTGGA | | | KVESPSCPLCQGRGT |
| | | ATCCCCATCATGTCCCTTGTGCCAGGGAAGGGGCACCTTGGAGCACATC | | | LEHILSSCPKALGEGR |
| | | CTCAGCAGCTGTCCCAAAGCACTTGGAGAGGGTCGCTATCGCTGGCGTC | | | YRWRHDQVLKAIAES |
| | | ACGACCAGGTGCTGAAGGCAATCGCTGAGTCTATCAGCTCCGCCATGGA | | | ISSAMEYSKRLPLPGR |
| | | GTACAGCAAGCGCCTACCCTTACCGGGACGCGGAGTTAGGTTTGTCAG | | | GVRFVRAGEQPPPQ |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GGCCGGTGAACAACCTCCTCCCCAACCAAGGGCCCAACCAGGCCTCCTT<br>GCAACAGCTAGGGACTGGCAACTAAGGGTTGACCTGGGGAAACAATTA<br>AAGTTCCCGGAAAACATCGTAGAAACCAACCTGAGGCCAGACATTGTTC<br>TGCACTCACAGTCGTCCAAGCAAGTTATTTTGCTGGAGCTGACTGTGCCC<br>TGGGAGGAGAATGGAGGAAGCGTATGAAAGGAAGGCAGGGAAGT<br>ACGCTGAGCTGGTGGAGGATTGCCGCAGAGCAGGGTGGCGCAGTAGA<br>TGCCTGCCTATAGAGGTTGGGGGTAGGGGCTTTGCAGGGAAGTCACTC<br>TGCAAGGCCTTTAGCCTCCTGGGCATCACAGGCATGCGCAGGAGGAAA<br>GCCATCTGCGCCGGCCTCAGAGGCTGCAGAGAGGGCGTCCAGATGGCTG<br>TGGATCCAGCGGGACAAGCCGTGGACGAGCGCTTCTTGGACACAGGCC<br>GGGAACTGATCACTCCCAGTCGGGTCGCCTGGGTGAGGGGGTCTGATG<br>TTGAAAGACCCGAAACCCCCGATGACCCCAGGTACTATCACTGACGATG<br>TGTCCAAGACATGCATCAATAGGTGTATTTAGAAATC (SEQ ID NO: 1044) | | | PRAQPGLLATARDW<br>QLRVDLGKQLKFPEN<br>IVETNLRPDIVLHSQS<br>SKQVILLELTVPWEER<br>MEEAYERKAGKYAEL<br>VEDCRRAGWRSRCL<br>PIEVGGRGFAGKSLC<br>KAFSLLGITGMRRRK<br>AICAASEAAERASRW<br>LWIQRDKPWTSASW<br>TQAGN (SEQ ID NO: 1412) |
| NeSL | LIN9_SM | . | Schmidtea mediterranea | AAACGACATCATGAACGCTTGGCCGCAACAATCCAGTTATCCCTGCGGT<br>AACATTGTGGAACTCATAAGACAAGTACTAAAAGAAGAATTAGAAAAAT<br>TAGAAGAAAAAAATTGAAAATAATTTATTTATAAAATTTAAAAATTTAAAT<br>AAATTTAAAAATTTAAATTTAAATTTAAATGAAGATAAAAATTTATTTAA<br>TCCAATAAATAATCAAGAAAATCAAGAAAATGATGGATTCAAGACAATT<br>AAATACTCCAAAAATAAGAAAATATCAGAACCCAAAAATGACAAACGAC<br>ATCATGAAAAGCTACAACTACGCGGTTTTGAGCGATGTCACGCCTCAAG<br>AAACCACTCAAACAACAACCCACTTAAATGTCGATATAGACAATGAAAC<br>CACCCAACCAAAACAGCCACTTACGAAGTCTGGCAAACCAAAATCTAAA<br>CCAATTGCGGTATCATACAAATTTAAAGATGCCACCTTCATCTGGGACAC<br>TACCCCACAAACAAATCCACCAAGAGATTGCACCAAACTTATTGATAAA<br>ACAAGACCAAGAAAGACCATCTTCAAAAAATCAGCATTTCAAAGCTACC<br>TCAAAAAAGAACTGTCCAATGAGACATTTGTGGAAGTAAAAACCTTCCT<br>CATGGCAACTCACAAATATCGTTTTAAAGACGAAAACTCAAGACTCTTG<br>GCATACCGAATAATTAATCGCTATGTCATGGAGACAGCAAATGAATTCA<br>AAGAAACCGAATTTGACATGGCTCGCTTTGCCAAATTCTTCACAATCCCA<br>GAGAATTGGTTAAAACATCTAAAACCATACTCTACAGCTACCGAAACAT<br>CACCGGCTGATAGAATAAAAGTACAAAAATTAGTGGATCTCACATGCAG<br>ATACCCATTCAAAACTCAAGAAGAGCAAACAAGTGTAGCAAACTTCCTA<br>CACTTCTTCACCCAAAGATCAATAATTGGAATCTCAAGAGATTATAAATT<br>CCAAAAATTTATACCATTTATGGCAAGAAAAAACACCAGGCCGGAGACA<br>ACCTCCACTATGGTTACGACTTCTCCAACAGAACAAAACAGACTACCAAT<br>GGTAATAATCACACCACTTGAAGAACCAAAAAGTGAACATCGTAGACCA<br>GAGAAAAGAGGCGCAAGCAATGACACAATTGTGCTTAGCGACGAAGAG | AAACGACA<br>TCATGAAC<br>GCTTGGCC<br>GCAACAAT<br>CCAGTTAT<br>CCCTGCGG<br>TAACATTG<br>TGGAACTC<br>ATAAGACA<br>AGTACTAA<br>AAGAAGA<br>ATTAGAAA<br>AATTAGAA<br>GAAAAAAT<br>TGAAAATA<br>ATTTATTTA<br>TAAAATTT<br>AAAAATTT<br>AAATAAAT<br>TTAAAAAT<br>TTAAATTT<br>AAATTTAA<br>ATGAAGAT<br>AAAAATTT | TAAAAT<br>GGCAAA<br>AAGATA<br>TTTCAAG<br>ATGAAT<br>TGTGGA<br>CTCATCT<br>AAAAAA<br>TGACCA<br>CCTTGA<br>GTCCAA<br>ATATGC<br>CTAGCT<br>ATCATG<br>GTTGCT<br>GATGGA<br>AACAGT<br>AAGGCA<br>CCTGAT<br>AGCTAA<br>CTTTTCA<br>CTGTGA<br>ATATCTT<br>CAGATA | MMDSRQLNTPKIRK<br>YQNPKMTNDIMKSY<br>NYAVLSDVTPQETTQ<br>TTTHLNVDIDNETTQ<br>PKQPLTKSGKPKSKPI<br>AVSYKFKDATFIWDT<br>TPQTNPPRDCTKLID<br>KTRPRKTIFKKSAFQS<br>YLKKELSNETFVEVKT<br>FLMATHKYRFKDENS<br>RLLAYRIINRYVMETA<br>NEFKETEFDMARFAK<br>FFTIPENWLKHLKPYS<br>TATETSPADRIKVQKL<br>VDLTCRYPFKTQEEQ<br>TSVANFLHFFTQRSII<br>GISRDYKFQKFIPFMA<br>RKNTRPETTSTMVTT<br>SPTEQNRLPMVIITPL<br>EEPKSEHRRPEKRGA<br>SNDTIVLSDEEFPLLK<br>RRTLPTRKSKNPTGA<br>GNVPTETECTDEVKFI<br>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 ACGGTC TGTATTC GTACTG AAAATC AAGATC 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 SKRIRKEIGELILEEMK 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 | CPLSMTLFSLCIQYLI |
| | | GCCTCTGTATACAATATCTTATACAAGGCATAGCAGAAAAGAAAAAAGG | | ACCTGA | QGIAEKKKGATIAGQ |
| | | AGCAACAATTGCAGGTCAAGAAGTTTGCATATTGGCTTATGCCGGACGAC | | CACTGTC | EVCILAYADDLVIVAN |
| | | CTAGTAATTGTTGCAAACACAGCAAAAGACATGCAAATGCTGTTAACAA | | CTCAAA | TAKDMQMLLTTIENL |
| | | CAATCGAAAATCTGGCAAAACAAGCCGATCTCATATTCAAACCGGCAAA | | ACAGTT | AKQADLIFKPAKCGY |
| | | ATGTGGATATTACAGAGACCCAAGAGATAAAAAGTCCATGATGAAGAT | | CAATTG | YRDPRDKKSMMKIY |
| | | ATATGGCAAAGAAATCAGCATAGTAGACGAAAAGAATGTTTACACCTAC | | CATCCG | GKEISVDEKNVYTYL |
| | | CTAGGTGTAAGAATCGGTGACACAAAGAAAAAAGACCTAAATGTCAGA | | AAGATC | GVRIGDTKKKDLNVR |
| | | TTCGAAGAGGTCAAAAAGAAAACGACAGCAATCTTCAAATCGAAATTGC | | GCAATTT | FEEVKKKTTAIFKSKLR |
| | | GAAGTGACCAAAAACTAGAGGCATACAACATCTTTTGCCAATCAAAATT | | TTTCACT | SDQKLEAYNIFCQSKF |
| | | TGTGTACATCCTACAAGGCGAAGATATCGCAAAAACCAAAATTGAAACT | | AAAATA | VYILQGEDIAKTKIETY |
| | | TACGACGAAGAAATCAAGAAAATGATAAAAAGAAGATATATTAAAATTAC | | AATTAA | DEEIKKMIKEDILKLQ |
| | | AAGACAAAAGTCCGTTCACAGACTTCGTTATCTACTCCCCAAGAGAAAA | | CAAAAG | DKSPFTDFVIYSPREK |
| | | AGGGGGGTTAGGAATAACAAAGATAATAGATGAACAAACAATTCAAAC | | TTAATTA | GGLGITKIIDEQTIQTI |
| | | TATTAATAGAACGGCAAAACTCCTAAATAGTAGCCATAGAGCAATCCGG | | TACTGCT | NRTAKLLNSSHRAIR |
| | | GCTATTATTTATGAAGAGCTAATACAAGTAGCTAACCTAAGAGGAGAAAA | | TCATTGA | AIIYEELIQVANLRGEK |
| | | AAGAAATCAACACCATTGAAGAAGCACTAAAATGGTTGGAAGGTACCA | | GTAAGT | EINTIEEALKWLEGTN |
| | | ACAAATACAAAAAGAACTCCAACGCCAAGACCACCTGGATAACAAGGG | | AGAAAA | KYKKNSNAKTTWITR |
| | | TTCGGGAGGCCTTTCAAACTCTAGAAAAGAAACACAAAATCAAGGTTAG | | ACAATC | VREAFQTLEKKHKIKV |
| | | ATTTGTGCCCAAAGAAAACTGCATTGGATATAAAATCAAATGCGACACC | | (SEQ ID | RFVPKENCIGYKIKCD |
| | | CAAGAAAAGATAGTGGAGCTTGATAACTCAAAAGAGTTATCAAAAAGC | | NO: | TQEKIVELDNSKELSK |
| | | TTACACTGGATGATAAAAGAGGCATATTATAAAGAATGGAAAGCCCTAA | | 1291) | SLHWMIKEAYYKEW |
| | | AATGCCAAGGATATATTATAAGCCTAAAAACCTCCGAATTTATGGAGTG | | | KALKCQGYIISLKTSEF |
| | | GAAAATGCCCAGAGGCCTTCCGGACCCTGATTGGAGATTCCTAACAAAA | | | MEWKMPRGLPDPD |
| | | GTAAAGGCAAATATGTTGGACGTAAACATGAAACAAGCCAACCAGGGA | | | WRFLTKVKANMLDV |
| | | GGAAGGTTGGGAAGCACAAAATGCCGAAAATGTGAAGATAAAGAATC | | | NMKQANQGGRLGS |
| | | GGCAAGCCATGTTATAAACCACTGTGCCTCAGGTAACTGGAGTAGAGTG | | | TKCRKCEDKESASHVI |
| | | GAAAAGCACAACCAGGTGCAAAATGAGCTAGCAAAAGAACTGACAAAG | | | NHCASGNWSRVEKH |
| | | CGGAATATCAGCTTCGAAAAGGACAGCATCCCAAAAGAAACAAAAGAG | | | NQVQNELAKELTKR |
| | | AGCCTAAGACCAGATTTGGTTATAAGACTCAAAGACAAGATAATGATAG | | | NISFEKDSIPKETKESL |
| | | TGGACATCAAATGCCCATTTGATGAGGAATCTGCTATCGAGAGTGCCAG | | | RPDLVIRLKDKIMIVD |
| | | AAACAAGAACATAGACAAATATCGAGAACTGGCCAAAGAGATCCAAGC | | | IKCPFDEESAIESARN |
| | | AAAAACTGGGTTACAAACAACAGTCTCAACTTTCGTTGTCTGTTCTTTGG | | | KNIDKYRELAKEIQAK |
| | | GAACCTGGGATAAGAGGAACAACGAGCTCCTACGGCAGATGGGAATAA | | | TGLQTTVSTFVVCSL |
| | | GATATGAAGAATCCAAAGAGATGAGGATCAATATGATCCAAAAAGCCA | | | GTWDKRNNELLRQ |
| | | TCCACGGGTCTAGAAAAACCTACGACCACCACAGAAATTTTAACAATGG | | | MGIRYEESKEMRIN |
| | | TTAAAATGGCAAAAAGATATTTCAAGATGAATTGTGGACTCATCTAAAA | | | MIQKAIHGSRKTYDH |
| | | AATGACCACCTTGAGTCCAAATATGCCTAGCTATCATGGTTGCTGATGG | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | AAACAGTAAGGCACCTGATAGCTAACTTTTCACTGTGAATATCTTCAGAT<br>ATTCACAGTGACACGAAAGGACACCACTAGTAAAAACCACTAGTTTTTTC<br>TGACACCTCTTGCTACAAACTCTGTAAAAATCAAAAGGATCGATAGGCC<br>GCGCTTTCACGGTCTGTATTCGTACTGAAAATCAAGATCAAGGAAGCTT<br>TTCCCCTTTTAGTCAACACCAGGTTTCTGTCCTAGTTGAGCTTCCCTTGGG<br>ACATCTGCGTTACCATTTGACAGATGTACCGCCCCAGTCAAACTCCCCAC<br>CTGACACTGTCCTCAAAACAGTTCAATTGCATCCGAAGATCGCAATTTTT<br>TCACTAAAATAAATTAACAAAAGTTAATTATACTGCTTCATTGAGTAAGT<br>AGAAAAACAATC (SEQ ID NO: 1045) | | | HRNFNNG (SEQ ID NO: 1413) |
| NeSL | NeSL-1_CBre | . | Caenorhabditis brenneri | AAGGACGCTGGTTTAAGGCCGAATTCGTTCGTTCTTTTTCTGGCGGTCTT<br>GCTTTGAGCTTGGTTTCCGATCCTATGCCCTTGWGCATCAGCGATTGCG<br>TTCATCTGGTCTCGGCCGAAGGAGATACGATGAACGGGAGGTCCACTT<br>GTGGGCCATTGTCTCGTTCATCCTCTGTCGTAAGTAGGTCGAGGTCTTCC<br>CCTTCCCCCAGTGTTCCCCCCCACCCTTCCCCCAGTATCGGTCCAGATACA<br>GGATTGTCGGCTGGAATCATCGGCACATCGAGAGGATGTAGCCTTTGGT<br>TGCCAGAGGTGGACAATGCCTTATCACAGTGGCTGAGAAAAGGGTTGG<br>AACGAGACCATGAAGTTCTGGTTTGTGGATTTGAGGCAGCAAAGCCACT<br>GTCACTTTCCAAAGCTAGACTTCTAAGAAAGACCCCAAGGAACACTGGT<br>GTGGTTAGGCACATATTAGAATTTGACGGAAGGTTAGTTCATACTAACT<br>GTAACGAGACCGAGTGTGTTCTTTCTACTTTGTKCAGTGAMGWGGCTG<br>TCGAAGTAGTCAGGATATCTCTCAAATGTGAACCCCGTGAACCCTGTGA<br>ACCCAAATGTGTTCTTTCTATTTTATGCAGTGATAAGATAGTCWGGATAT<br>CATTTGAATGTGAAACWCGTGAACCTTTTCCTTTCTTCMCGGATCGGAA<br>ATTCAGAGAACCTATTCCTTTCGTTTTCGAGAGGATGTATGACCCAAGA<br>GACCCTATTCCTTCATTTATTTGTTGGATGTATGACCTGAGACAAAGGAT<br>GACCCCTGGCACSTTGCCAAGWAATCCCCCTTTCTCMAGAGAACAAAGA<br>CAGCTGGGGACGCCCAGCTGTCATAAAGAATGAGATAAGATCTATGAG<br>ATCTTATCTCGAAGAGAATGTGAAGGAAAACCGCCTGAACCTTTTGAGA<br>AGGTTAAGAGGTGGTGGTGAAGGAAAGAAGATGATCAGAAAGTTGGT<br>TGCAGAAAAGAAAAGCGACACAGAGGCTGTCTGCAGGATACTGTACCC<br>ACTTGATGATCGTTATGAGTGTTTTGTTGATGGTTGTGAGACAACATCAA<br>CGATGGGATACGGGTCTAGTGACCTGAAATACATGACCACACACATAAA<br>GAAAGAGCATGGTGTGAAAGTCCAATGGACATATGAGTGCTCCCTGTG<br>TAATAAGCAAGCTCCTTTCATGGGGGGAGCTGCGTCCAAGTGGGTTACA<br>GCGCACATGGCAACAAAGCATACCGAAACGGTGAAGTTGAAGCTCAAA<br>CCAAGCATCTCGACTACTGCCAAGGTTGCTGCGAAGCTAGATGAGATCG<br>CCGTGTCGCTACCCAAACCGAGACAAGTACGTGTATTGAGAGACCCAGA<br>TGAAGTGAAAGAGAAGGTTGCAAAACCAACACTTGCTTCCACGAGAGA | AAGGACG<br>CTGGTTTA<br>AGGCCGA<br>ATTCGTTC<br>GTTCTTTTT<br>CTGGCGGT<br>CTTGCTTT<br>GAGCTTGG<br>TTTCCGAT<br>CCT (SEQ ID NO: 1169) | TAAACC<br>CACACG<br>AGAMCT<br>ACGACG<br>CCATAA<br>GATCAG<br>GCATGT<br>ACGGAT<br>GTGAAT<br>GAGACT<br>GATGAA<br>CGGAAT<br>GAGCAC<br>GTGCCC<br>ATAAGA<br>TCGGGT<br>ATKAAA<br>GAWCA<br>GAGACG<br>ATCCCTA<br>MCATCG<br>GGAAAA<br>CACGAG<br>TTATACT<br>GCTTCAC<br>TGAMCT<br>CGCTAA<br>GCTCTCA<br>TAATGA | MPLXISDCVHLVSAE<br>GDTMNGRSTCGPLS<br>RSSSVVSRSRSSPSPS<br>VPPHPSPSIGPDTGLS<br>AGIIGTSRGCSLWLPE<br>VDNALSQWLRKGLE<br>RDHEVLVCGFEAAKP<br>LSLSKARLLRKTPRNT<br>GVVRHILEFDGRLVH<br>TNCNETECVLSTLXSX<br>XAVEVVRISLKCEPRE<br>PCEPKCVLSILCSDKIV<br>XISFECETREPEPFFXD<br>RKFREPIPFVFERMY<br>DPRDPIPSFICWMYD<br>LRQRMTPGTLPXNPL<br>SXENKDSWGRPAVIK<br>NEIRSMRSYLEENVK<br>ENRLNLLRRLRGGGE<br>GKKMIRKLVAEKKSD<br>TEAVCRILYPLDDRYE<br>CFVDGCETTSTMGY<br>GSSDLKYMTTHIKKE<br>HGVKVQWTYECSLC<br>NKQAPFMGGAASK<br>WVTAHMATKHTETV<br>KLKLKPSISTTAKVAA<br>KLDEIAVSLPKPRQVR<br>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 |
| | | GAAAGAGATTCGGAGAAAACCGAAACAACAAATCGAAAAGAAAAAGAA | CACACG | TFTIEQAEDPAEEDA |
| | | AGGTGCCAACAGGGAAACCAGATGAACTGTTGGTCAGAGTGCGATTAT | GCGGGT | QHLESPVKPVKLMEL |
| | | GGTTGGAAAGAGAAGTCCAATCATACTTCGACTCTGGAAAGAGATTCCA | 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 |

| | | | | | |
|---|---|---|---|---|---|
| | | CAAAGACACCGCTCCGGGTGTAGACGGACTACGGTATCACCATTTGAAA | | GGGAAA | FEGTTAETNVPAEVL |
| | | TGGTTTGATCCAGAGTATAAGATGTTGACACTTCTCTACAATGAATGTAA | | GGAGAA | KEMGSRLPKLEALD |
| | | GAACCATCGAAAGATTCCAAGTCATTGGAAAGAGGCAGAGACGATTCT | | AGTTGTT | WMEANFIESEVSDA |
| | | CCTTTATAAAGGAGGTGATGAAACGAGGCCCGACAACTGGAGACCTAT | | TGTTGG | MKKTKDTAPGVDGL |
| | | AAGTTTGATGCCCACGATCTACAAACTGTATTCTAGTCTTTGGAACCGAA | | GCTGGC | RYHHLKWFDPEYKM |
| | | GAATTAGATCAGTTGGTGGTGTGATGAGCAAATGTCAACGAGGTTTCCA | | AAGAGT | LTLLYNECKNHRKIPS |
| | | AGAGAGAGAAGGGTGTAATGAAAGCATAGGAATCCTTAGAACGGCTAT | | GAAGTT | HWKEAETILLYKGGD |
| | | CGATGTCGCTAAGGGAAAGAGAAGGAACCTGTCAGTTGCATGGTTAGA | | TGAATG | ETRPDNWRPISLMPT |
| | | CCTTACGAATGCGTTTGGTTCAGTACCCCATGAACTGATAAAAAGTACTC | | TGAACC | IYKLYSSLWNRRIRSV |
| | | TGGAATCGTATGGATTCCCAGAAATGGTGACAGAGATTGTCATGGATAT | | ACCGTC | GGVMSKCQRGFQER |
| | | GTACAGAGGTGCATCAATCCGAATCAAGAGCAAGAATGAGAAAAGTGA | | ATGCAA | EGCNESIGILRTAIDV |
| | | ACAGATTGTTATCAAATCTGGAGTGAAGCAGGGAGACCCCATCTCCCCC | | CCACTA | AKGKRRNLSVAWLD |
| | | ACGCTGTTCAACATGTGTTTAGAGAATGTGATAAGAAGACATCTGGATA | | AACCAG | LTNAFGSVPHELIKST |
| | | GTGCTTCGGGACACAGATGCATAAAGACAAAAGTCAAGGTCCTGGCTTT | | TGGCGA | LESYGFPEMVTEIVM |
| | | TGCAGATGACATGGCCATACTGGCCGAGAACAGATCAGTTACAAACT | | TGCGGG | DMYRGASIRIKSKNE |
| | | GAACTAAACAAGTTGGACAAAGAATGTGAATCACTAAACCTCATTTTCA | | TGGAGT | KSEQIVIKSGVKQGD |
| | | AACCGGTAAAGTGTGCCAGTTTGATAATTGAGAGAGGAATGGTGAATA | | CATCAC | PISPTLFNMCLENVIR |
| | | AGAATGCGGAAGTGGTTCTGAGAGGGAAGCCAATCAGAAACCTAGATG | | AGGAAA | RHLDSASGHRCIKTK |
| | | AGAATGGTTCCTATAAGTACCTGGGAGTACATACAGGAATCGCAACAAG | | ATGTTTC | VKVLAFADDMAILAE |
| | | AGTTTCAACAATGCAATTGTTGGAAAGTGTCACGAAAGAAATGGATCTA | | TGTTGCT | NRDQLQTELNKLDKE |
| | | GTGAATCAGAGCGGCATGGCTCCGTTTCAAAACTAGACTGCCTAAAGA | | TGACTTA | CESLNLIFKPVKCASLI |
| | | CGTTTGTTTTGCCGAAACTAACGTACATGTATGCGAACGCAATACCTAA | | TCAGTG | IERGMVNKNAEVVL |
| | | GTTAACGGAGCTTAAGGTCTTTGCAAACTTGACGATGAGAATGGTGAAA | | TTTGATA | RGKPIRNLDENGSYK |
| | | GAGATTCATGAAATCCCCATCAAGGGATCTCCGTTAGAGTATGTACAGC | | TCGCCCT | YLGVHTGIATRVSTM |
| | | TACCTCCAAGTCAAGGCGGATTAGGAGTGGCTTGTCCAAAGATAACAGC | | CAGGCA | QLLESVTKEMDLVN |
| | | GTTGATTACCTTCTTGGTCAACGTCATGAAAAAGCTATGGTCTTCTGACA | | CAAGTA | QSGMAPFQKLDCLK |
| | | GTTACATCAGAAAACTATACAGGGACTACCTGGATGAAGTCGCAGAGA | | TGAAGG | TFVLPKLTYMYANAIP |
| | | CGGAGACAGGTATGGAAGAGATGACGAAAGAAGATATTGCAAAATATC | | CCCCCAC | KLTELKVFANLTMRM |
| | | TGAGTGGTGATGTGCCGATCGACAAGAAGGCGTTCGGTTACAACACMT | | CCACAT | VKEIHEIPIKGSPLEYV |
| | | TCACAAGAGTAAGAGATGTGTGCAACAGCCTCACTAASATAGKGGGAG | | AAACTC | QLPPSQGGLGVACPK |
| | | CTCCACTGCACAAGTTAAAGATTGTGGAGAGAGACGGTGACTTTGCCAT | | CCTAGC | ITALITFLVNVMKKL |
| | | TCTAGTGCAAGCCACCAAAGAAGGAATGGAGAAAATCTTCACCTGTGCT | | AACTGG | WSSDSYIRKLYRDYLD |
| | | CAGGAGAAGAAACTCCAACAGCTTCTGAAAGCAGAAGTMAACACGGCT | | TAGTCC | EVAETETGMEEMTK |
| | | CTAGCGCACCGTTTCTTCACCGAGAAACCMGTGAAAAGTGCAGTGATG | | AGCAAG | EDIAKYLSGDVPIDKK |
| | | AGTGTAATGAGACAGTATCCACACAGAGCAATGCCTTTGTGAAGAATGGA | | CGCTGG | AFGYNTFTRVRDVCN |
| | | AAGAATGTGAGCATTGCTGTCCACTCGTGGATACACAAAGCAAGGTTGA | | TWCTTG | SLTXIXGAPLHKLKIVE |
| | | ATGCGCTGCATTGCAACTTCAACACGTACGGTGAAAACCAAGTCAAAAGT | | CTACTAT | RDGDFAILVQATKEG |
| | | GTGCCGACGTTGCGGCAAAGACGTGGAAACCCAACTGCACATCCTGCA | | TGCGCC | MEKIFTCAQEKKLQQ |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | GWCATGCGAGTACGGGTTACCAAAGCTAATCAACGAAAGACATGATGC<br>GGTGTTACATGTGGTGAGAAACCTCATCCGCAAAGGCTCAAAGAAAGA<br>CTGGAAGCTAAAGATAGATGAAACTGTGTCAAGTTGTAATCAACTTCGT<br>CCAGACATCTATATGTGTAGCCCAGATGGGAAAGAGGTCATAATGGCA<br>GATGTAACCTGTCCTTATGAATCAGGAATGCAAGCTATGCAAGAGAGTT<br>GGAACCGAAAGGTCACSAAATACGAAGGAGGCTTTAGCCACTTCCAWA<br>AGATGGGAAAGAAATTCACAGTGTTGCCAATAGTGGTTGGATCACTGG<br>GAACGTGGTGGAAACCCACAACGAACAGCTTAGTTCAACTAGGCATAG<br>AGAAAGASACGATAAGAAGAGTGATCCCCGAGCTGTGCTCAATGACCA<br>TGGAATACAGTAAGGATGTCTACTGGAACCATATATTCGGGGACACCTT<br>CAGGAAGCCACCAATGAGATTTGGTGTAGAGAAGCCAAAGGGTAATAG<br>TTGGAAGAAGGAAGGCAGCGAGCCGAAAGGTGCTGCTTCCTCCGACTA<br>AACCCACACGAGAMCTACGACGCCATAAGATCAGGCATGTACGGATGT<br>GAATGAGACTGATGAACGGAATGAGCACGTGCCCATAAGATCGGGTAT<br>KAAAGAWCAGAGACGATCCCTAMCATCGGGAAAACACGAGTTATACT<br>GCTTCACTGAMCTCGCTAAGCTCTCATAATGACCGAACTTGTTCGCAAC<br>TGCCTCCTAACCGGGCGGGTGTGAGAAGGGAGGTCGCCTTGAGGCGGA<br>CGCAATGAGGGATGTGTGCAGGTTCCCCCTCTTGAGATCCGAAAGTCTA<br>AAAGTACTAGACCGAAAGATCGAGGACGGACGGGATGGCCGCGAGGC<br>ACACGGCGGGTAACACAGCCAGATAACCTAGTAGATCTTCGGATCTCGT<br>CGGCCTGGAGATATGTGGAACCCTGGGAAAGGAGAAAGTTGTTTGTTG<br>GGCTGGCAAGAGTGAAGTTTGAATGTGAACCACCGTCATGCAACCACTA<br>AACCAGTGGCGATGCGGGTGGAGTCATCACAGGAAAATGTTTCTGTTG<br>CTTGACTTATCAGTGTTTGATATCGCCCTCAGGCACAAGTATGAAGGCCC<br>CCACCCACATAAACTCCCTAGCAACTGGTAGTCCAGCAAGCGCTGGTWC<br>TTGCTACTATTGCGCCCCAGGCTCGCCC (SEQ ID NO: 1046) | | GCGCCCCG<br>GGTTACAT<br>TGTCGGG<br>GCCACCTT<br>TCTCTTGG<br>AGTAGAGT<br>ACAGTCTA<br>CTAATTTTT<br>TGATAAGC<br>TAGTCGGG<br>TCCGAACC<br>ACTAGAGT | CCAGGC<br>TCGCCC<br>(SEQ ID<br>NO:<br>1292) | LLKAEVNTALAHRFFT<br>EKPVKSAVMSVMRQ<br>YPQSNAFVKNGKNV<br>SIAVHSWIHKARLNA<br>LHCNFNTYGENKSKV<br>CRRCGKDVETQLHIL<br>QXCEYGLPKLINERH<br>DAVLHVVRNLIRKGS<br>KKDWKLKIDETVSSC<br>NQLRPDIYMCSPDGK<br>EVIMADVTCPYESG<br>MQAMQESWNRKVT<br>KYEGGFSHFXKMGK<br>KFTVLPIVVGSLGTW<br>WKPTTNSLVQLGIEK<br>XTIRRVIPELCSMTME<br>YSKDVYWNHIFGDTF<br>RKPPMRFGVEKPKG<br>NSWKKEGSEPKGAA<br>SSD (SEQ ID NO:<br>1414) |
| NeS<br>L | NeSL-<br>1_Cjap | .<br>Caenor<br>habditis<br>japonic<br>a | GCGCCCCGGTTACATTGTCGGGGCCACCTTTCTCTTGGAGTAGAGTAC<br>AGTCTACTAATTTTTTGATAAGCTAGTCGGGTCCGAACCACTAGAGTTTG<br>CTTGAAAATGCGTCAAACCAGCATTTTAGAACTCGCCCAAAAGTTCGGC<br>CCCGACCCCCAAACAAATGGGACCTTCTTGACGATTTTCCCTGAAAATCG<br>GAGGATGGAATGGTCCCCTATTCTTGTAAATAGKACTGTGCAATACCCC<br>TTCGTCATCTGTGGGGAACAGATGACACGTGACGTCATCCGTGTAGACG<br>TCACGTTTTCCCGTGCCTGCGGGAGCCCCCAATCGAGCAATTTTTGCTCT<br>TTTGAGTGTCTGGAACGCTTGAAACCCCAGACAAATCAGGCCCAGTCGT<br>CGGAAAATTTCTTTTTTGAAATTTTTTTGGCGCCTGCGAAAAAAAATTTTTTA<br>ACCGCCACAAACCCCCGGGAGGCGCGGWTAGGGATATCGATGTCATCG<br>ACTCGTCGGTGATCTTTGATTTTCTCTCTGCGTCTCCTATTTTGGAACAGT<br>CTCGACCAAAAAAACCGGGCCTGGCAACCCACCGAATCCGGATGTCGGA | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | 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 | |
| | AGCACTGCCGACAAAAAAGCTTCAAGCCAGTCGGTTTTCTGCTTCATTTC | GTAAATAG | TGTTCCA | WCQHKSLVGYGAHD | |
| | CTAAAAAATTCCTCGATCAACGAGGTGGCGTCTCTCCGCAACGTCAAAA | KACTGTGC | CGCCTTA | LKYLTDHIKSTHSKKV | |
| | AAAWKTTTCTTGAATTTTTCTCAGGAAAACCTATCGGTGGGATGGCTTC | AATACCCC | AAAGTC | EWSYQCSICDAKAEG | |
| | TTTCAGTAGAACCAAAATAACTTTTTTTCAAACTTTGTTTGAAGAATTTTGT | 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 | ATTTTTTAA | CGCGAG | QTVPKQKDGFNCGV |
| | | TTTATGGCWGGAAACAGAAGAGATTCAATTGCAAGAAGAGAAAGTCTG | CCGCCACA | GCTTAT | HVCLVAKSVITENFW |
| | | TCTCTCGGCAAAATCTCAAACTCATTTAAAATTGAGCTGAAAAATGCTCC | AACCCCCG | GGCGGG | YTEKDVNDFRKTVKL |
| | | GGAAAAAACAACTCTTAAAAAACCGGCTGTCACTCAGAAACAAAACACG | GGAGGCG | TAACTC | WLFSEGFELYSEPYK |
| | | AGTCAGAATGTATCTAGTTCTACGGTTGTAAAAGAGAACAAAACAGGAA | CGGWTAG | GGTTGG | QIQNKNISVNSEKNQ |
| | | ATGACGTGATCACAATTGATGACACAGAAACTGTTAAAAGAAAAATAAA | 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 |
| | CATCGATTCGTGTAAAAACAAAAAACGAAAAGAGTAAACAAATCCTGAT | | 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 |
| | AAGAGTTCGCAAATATTACAATGAGAGCAATAAAAGTAATGCAAAACAT | | | VSIGAHRFVHKARLN |
| | TCCAGTAAAAGGTTCACCATTGGAGTATGTACAGTTACCCATTGGAAAA | | | LLACNYNTWQDAAT |
| | GGTGGACTAGGAGTGGCATGTCCAAAAACAACTGCGTTGATAACCTATC | | | KQCRRCGYEKETQW |
| | TGGTTTCAACAATGAAAAAATTGTGGTCCACTGATGACTATATCAGGAA | | | HILSSCPKSMGGKITE |
| | ACTACACACAGACTACCTGAAAATGGTGGCCATAAAAGAAACGAAAAC | | | RHDSVLKTVKEMIQT |
| | AAAAGAGGTCACACTAGAGGACCTTGCCTCCTACCTAAGTGATGATAAA | | | GSLKNWKLKLDHELP |
| | ACCGTCTGCAAAAAAGCGGTTGGTTATAATTCATTCACAAGGGTACGAG | | | GSTRLRPDIYLRSPNG |
| | AAATCTGCAAAACGCTATCAAAAAACAAGGAGCACTGTTAAGCCAACT | | | SEIILGDVTIPYEHGIE |
| | AAAAATCATTGCAAAAGATGGAAAGTTGGCTATTCTGGTACAGGCTSTG | | | AMQTAWQKKIEKYE |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | AAAGATGGCAAAACAAAGATTTTCACGCATGACCACGTGAAAACCTTGC<br>AAAAASTTCTTAAAAAAGAAATAAATGAAGCCCTTCTGCACAGATTCACA<br>ACTGAAAAAAGAGTGAAAAGCGAAGTGGTGCGAGTGGTCCAAGAGTA<br>CCCCCAGTGCAACTCCTTTGTCAGAGATGGAGGAAAAGTTAGCATTGGA<br>GCGCATCGCTTTGTGCACAAAGCCAGGTTGAACCTGCTCGCGTGTAATT<br>ACAACACGTGGCAGGATGCAGCCACAAAACAATGCAGAAGGTGTGGAT<br>ATGAAAAAGAAACCCAATGGCACATCCTCTCATCTTGCCCAAAAAGTAT<br>GGGAGGAAAAATAACTGAAAGACACGATTCTGTGTTAAAAAACAGTAAA<br>AGAGATGATTCAAACTGGATCTCTCAAAAACTGGAAACTAAAACTTGAT<br>CATGAATTGCCAGGATCAACCAGACTTCGCCCGGATATCTATTTGAGAA<br>GCCCAAATGGATCCGAAATAATTCTTGGCGATGTCACAATCCCGTATGA<br>ACACGGAATTGAAGCTATGCAAACAGCATGGCAGAAAAAAAATTGAAAA<br>ATATGAAGAGGGCTTCAAATACCTTCGTTCTACCGGCAAAAAACTCACA<br>ATTGTGCCAATTGTGGTCGGAGCACTAGGAAGTTGGTGGAAGCCCACA<br>ACAGACAGTCTTGTCAGTCTGGGAATCGACAAAAATACTGTAAAAAGAG<br>CTATTCCAGAAATTTGCTCTACAGTACTCGAATACAGTAAAAACATTTAC<br>TGGAACCATATATTCGGGGATTCCTACCAAAAAGTACCCATGTTTTTCGG<br>CGGTGAAAAACCAAAGGGGCAAAGTTGGAAGAAAGTGAAGCCTCCTGA<br>AGGCAAAACTGCTTCTAACCATGAGCCTCCAGGTTAAAAGCCAAAAGCC<br>ACGGAGCATCGGGAAAGAAAAATGGAAAAGGACTGAAAACGAGACTG<br>AAAAATCCCAAACAAAACAAATCCAAAACAAACTGAAAAAAAAAAAAAA<br>AACAAAACAAAAACTGGACAGACACTGGAAACAGTGTCAGGCAAAGTC<br>GCCGATTATACTGTTCCACGCCTTAAAAGTCCCGAAATGGCGCAAAACA<br>ACCTGAATCTATCTGAAAGTGCTCCAAACCACGCACAACTCGGAGAAAA<br>TCAGGGACAAGTTGCTTCACGCAACGGGCTGGGACAGGTACCCCCTCCT<br>GAAACCGCGAGGTTGAGGATGGACGGGAAGGCCGCGAGGCTTATGGC<br>GGGTAACTCGGTTGGTGTGCTAGTAGATGATTTATATCCGACAGCCCCA<br>ACTAAGAGGAATCCTGGGAAAGGAAAACTTGAAAAAGTTTTTACAGGG<br>CTGGTAATAGTTCAGCACAATTGTAGTCTACTGTCTTGCAACCACAACAA<br>ACCAGTGGTTCTGCGGGTAGATCAAACTATAATTTGTGTGTTTTCTTTTA<br>CTTGACCCGGGCAACACATTATACCACGTCCACAAGGACGAATTCATAA<br>TGGCCCCTCCCTAAATAAACTCCCTAGCAACTGGTGGTCCGGCGAAGCC<br>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) |
| NeSL | NeSL-1_CRem | . | Caenorhabditis remanei | CGCGAACCAGTCATATGACAGTCTTTATTGATCGCGGTATAGGCGAGCG<br>AGGCCAGATGGCCGTATGTAGCCTCCACCGTTATTTTTCGTTTTCACCTTT<br>TTCCCCCATCCCCCCGTATGTAAATAATGGATCGTTCGGCGAAAATGGCT<br>GTGGCACAGACAAATCACTGTTGCCCGTCATAGAAGTTGTTGTTCGTGA<br>AGTTAAGATAAATTGGTCTGAGAATATTTTGGTAGTAGAGTGTCTGATA | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 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 |
| | | TAAGCATTTCACGAATATTGAATTTGCATATAAATGTGCTACGTGTATGT | | 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 |

| | | | | | |
|---|---|---|---|---|---|
| | | TGATGGAAATGAGAGTGCAGACACCTCCAGTGATCAATGAAAAAAGAG | GAAAGG | | VHRDSDHGDDEVEH |
| | | GTAAAAAGCGAGTATCGGCCAAAGAGAAACCGAGAAAGCAAAAGGAA | GAAACC | | LKTIEQEPENEISEIEN |
| | | AAAGAGCAAAAAGTGCCAACAGGAAAACCAGATGAGCTGGTTAAAAGA | TAAAGA | | VEGSVDSVIPKLMEM |
| | | GTAAGAGTATGGTTTGAGAAAGAATTCAAATCGTATGTGGAAGATGGA | AAATAG | | RVQTPPVINEKRGKK |
| | | AAAAGTTTCCAAAGGTTGGAATGGMTAACAGATGTTCTCACTGCAGCA | TTCTCTT | | RVSAKEKPRKQKEKE |
| | | ATTCAGAAGGCGTCAGCCGGAGATGAGAAAGCAGTAGAACTGATTGAG | GCCAAA | | QKVPTGKPDELVKRV |
| | | AAAAGATGTCCACCTTTGGAAWKCGAGGAGGGTGAAATGTGTACCCAG | ATTCTGT | | RVWFEKEFKSYVEDG |
| | | ACTGAAAAGAAAAAGAAACCAAAAAGTGGTAAAGGGAATGGCGGTCA | AGAGGA | | KSFQRLEWXTDVLTA |
| | | AGAAAGTATGAAGTCCTTGATGGCCTCATACAGTGAGAACCGAGCCAA | ATACTTT | | AIQKASAGDEKAVELI |
| | | AACCTACAATAGAATAATTGGTAAGCATTCAAAGCAGTGTGAGATCCCA | GTCAAA | | EKRCPPLEXEEGEMC |
| | | ATAGCCAAAGTACAAAAGTTCTTTGAAGGGACCACTGCCGAGACAAATG | ACATGA | | TQTEKKKKPKSGKGN |
| | | TGCCAAAGGAAACACTTAAGGAAATGTGTTCACGCCTCCCGAAAGTTGA | TAGAAA | | GGQESMKSLMASYS |
| | | AGTGGGAACGTGGATTGAAGGTGAATTCAGTGAAAGTGAAGTGACTGA | CCAGTA | | ENRAKTYNRIIGKHSK |
| | | AGCATTGAAGAAGACAAAGGACACAGCACCAGGGGTAGATGGATTAA | ATCTGG | | QCEIPIAKVQKFFEGT |
| | | GGTACCATCACCTGAAATGGTTTGATCCCGAGTTGAAAATGCTGTCACA | TACGAA | | TAETNVPKETLKEMC |
| | | GATCTATAATGAGTGTAGAGAACACAGAAAAATTCCAAAGCATTGGAA | AGACAA | | SRLPKVEVGTWIEGE |
| | | AGAGGCAGAGACAATTCTTCTCTCTATAAGGGAGGAGATGAGTCMAAAM | GTAAGA | | FSESEVTEALKKTKDT |
| | | CGGATAATTGGAGGCCTATCAGTCTGATGCCAACCATCTATAAACTGTA | CCTGAA | | APGVDGLRYHHLKW |
| | | TTCTAGTCTCTGGAACAGAAGGATTAGAGCAGTGAAAGGGGTGATGAG | CTGACA | | FDPELKMLSQIYNEC |
| | | CAAGTGTCAGAGAGGTTTCCAAGAAAGAGAAGGATGTAATGAAAGTAT | AGAAGG | | REHRKIPKHWKEAET |
| | | CGGAATATTGAGAACAGCCATAGATGTGGCCAAGGGCAAAAAGAGAAA | AAGTCA | | ILLYKGGDESKXDNW |
| | | CATAGCCGTAGCATGGTTAGATCTCACGAATGCCTTTGGATCAGTACCA | GAAAGA | | RPISLMPTIYKLYSSL |
| | | CATGAGCTGATAAAAGAAACTCTGGAATCTTACGGATTTCCAGAAATAG | AATACC | | WNRRIRAVKGVMSK |
| | | TAGTAGACGTCGTAGAAGACATGTATCGAGATGCATCGATCCGTGTGAC | GCTCAC | | CQRGFQEREGCNESI |
| | | GACGCGAACGGAGAAAAGTGATCAGATTATGATCAAGTCAGGAGTGAA | AAAGCC | | GILRTAIDVAKGKKR |
| | | GCAGGGAGATCCAATCTCGCCTACTCTCTTCAACATGTGTCTCGAGAGT | TGTGAT | | NIAVAWLDLTNAFGS |
| | | GTCATCAGAAGGCATCTCGACAGATCAGTCGGCCATCGGTGCCTGAAAA | CGATTCT | | VPHELIKETLESYGFP |
| | | CAAAAATAAAAGTATTAGCCTTTGCAGACGATATGGCAGTATTAGCAGA | CTTACCT | | EIVVDVVEDMYRDAS |
| | | AAGTAGTGAACAGTTGCAAAAGGAGTTGACAGCTATGGATGCTGACTG | ACTGAA | | IRVTTRTEKSDQIMIK |
| | | CTCAGCACTGAATTTGCTATTCAAACCGGCTAAATGTGCAAGTCTGATAT | CTTGTTC | | SGVKQGDPISPTLFN |
| | | TGGAAAAAGGAATAGTAAACAGGTTAAATGAGGTAGTTTTGAGAGGGA | TCTTGG | | MCLESVIRRHLDRSV |
| | | AACCGATCAGAAACCTCATGGAAAATGAGACCTACAAGTACTTAGGTGT | CCTCGTA | | GHRCLKTKIKVLAFAD |
| | | TCAGACAGGTACGGAAACAAGGGTTTCCATAATGGATCATATAACGGA | ACCGGC | | DMAVLAESSEQLQKE |
| | | AGTGTCAAGGGAGATAGATCTAGTGAATATGAGTCAACTGGCAATGCA | TAAAGG | | LTAMDADCSALNLLF |
| | | CCAGAAACTAGATATACTCAAAGCCTTCATACTTCCAAAGATGACCTATA | GAGAAG | | KPAKCASLILEKGIVN |
| | | TGTATCAGAACACGACACCTAAACTGTCAGAACTGAAAGTGTTTGCCAA | GAATGT | | RLNEVVLRGKPIRNL |
| | | TTTGGTAATGAGGTCAGTGAAGGAATTCCACACAACATTCCCCTAAAAGGG | TAATTG | | MENETYKYLGVQTG |
| | | TCACCGTTGGAGTATGTCCAACTTCCCGTAGGAAAAGGAGGATTAGGA | GAGATA | | TETRVSIMDHITEVSR |

| | | | | |
|---|---|---|---|---|
| | | 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 |
| | | ATCCAAATGTAAACAGAAAAAACCAGTGGGGAGGAAAGAAAGACTGAT | 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) | | | |
| NeSL | NeSL-1_TV | . | Trichomonas vaginalis | GGGTGAGTAGTCTAGTGGTATGATTCCTGTTTTGGGTACAGGAGGTCCC GAGAAGCTTCCACTGCAATCGTACGTGTACTGTGGCAACACAGCTATAA CAGACAGTTTCACGCCAACCGCGAAAACGATTTTGAAGCCTGAGGAACA AAATTTAGATATCGTTTTGAAAAATATTGCAGCGTTGAATCCAGAAAATT ACTCCGACTTAATCAGGAGCCTATCGAAGATGGAGTTCAGATTAGATTA CCCGAAAGAAATAGAGAATTACTGGATTTCGGAAAAATTATTTAGCCAA TCCATCGCATCATTGCCCATCAGTTTGTTAGTCGCATCCATGTTCTCACCT GAAGACCGTGACTTGAGTACAGAACCGTTCCACTGTAACGCTGATGGCT | GGGTGAG TAGTCTAG TGGT (SEQ ID NO: 1172) | 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) | 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 | |
| | | ACATCAAAGAACATCTAAAGAAGAAACATAAGTCTACAACAGCAGGAG | CACATCC | ACRKCPEAGCNTYVT | |
| | | TAGACGGAGAAATCGCGCACACTGATGCTACATACTGCTGGATTACCAA | ATGATA | ELSDIKEHLKKKHKST | |
| | | AGAAGAACTCGACGCATTACATGCCGAGAGAGCAAGAGAAAGAGCAG | CACTCAT | TAGVDGEIAHTDATY | |
| | | AGCAAGTAGACAACACACTCCAGTACAACAGATAATTAATGCTGACAACAA | TGGAGT | CWITKEELDALHAER | |
| | | TGAAGAGAACAACGAGAACCAAGAAGACAACGGAAACAACGAAGAAG | GAAAAC | ARERAEQVDNTPVQ | |
| | | CAGATGCCCTCGACCCGCCAAATAACACAACAGAGACAGAAGATGAAG | CACCAA | QIINADNNEENNEN | |
| | | CGGTTCATGCCGTCATCATCAATCCACCAGCAACAGAAGAGGAAGAGGT | CAACAA | QEDNGNNEEADALD | |
| | | AGCCATCATCGCCGAGGCAAGAAGAAACATTCCAGAACTCCAACACAAGC | ATCCACC | PPNNTTETEDEAVHA | |
| | | AGAAGAGAGAGGCTGCGTTACACCGAAAATGACATCACTCGTCCGATTA | TAGACC | VIINPPATEEEEVAIIA | |
| | | AAACTATTGAAAGGAGGAGGAGAACTTTTCAACAAGAAACTCACTCCAT | AAATCCT | EARRNIPELQQAEER | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | TAGCCACAAGATACGCAGCTACAGGAAATACAGAAGCAGACAAAATCA | | GCCCCA | | GCVTPKMTSLVRLKL |
| | | AGGTAGATTACTTGACACTAAAATGCAATGCCGCCTTGAGAGAAATGAT | | CCTCCAC | | LKGGGELFNKKLTPL |
| | | CTACACCAATAACCACAGCGAATCAAAGTTTATGACAGCAGAAAATGGA | | CCAAGT | | ATRYAATGNTEADKI |
| | | GAAGACACAGCACACCACCGCCAAGGATATCGGAAGACACAAGAGATCGC | | 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 |
| | | TCGGCACACGGCCTCGATGGGAATTCCGAACTCAGTTTACATGCTATTCCC | | 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 |
| | | GCTGCTTAGATCCTTTATTATATGCAGTCCAGAGGAGACACTTTGAGGA | | | | PHQLIFDSLEAKKVPD |
| | | CGGTTACAGATTCCAAGACAAAGCAGGACAGTATTCAATTGCCATTCAA | | | | SIINLLKDLYKGARTAI |
| | | GCTTACGCTGACGACGTTCTAGTCATCTCTCCAACACATGAAGGAATGC | | | | YTRHAHSEIVPVRRG |
| | | AAAGAATCTTAAACACAGTAGATGAATTCCAGAAAATTGCGAAACTCAA | | | | VIQGCPLSPILFNCCL |
| | | AGTTGCACCACAGAAATGCGTCACACTTGCCAAACATCCCACTGCAATC | | | | DPLLYAVQRRHFEDG |
| | | CAACCTTTCCGCATTGGTCCAGACGAAATCCCAATCAAGACGAGCATGG | | | | YRFQDKAGQYSIAIQ |
| | | ACAACATCACATATCTTGGAATACCAATCTCTGGAACAAAGACATCAAG | | | | AYADDVLVISPTHEG |
| | | ATTTGCAGCTGCAACTGGCATTCTGGAAAAGGTCAAAGCACAGATCAGA | | | | MQRILNTVDEFQKIA |
| | | GTCGTCTTCGCGTCACATCTCGCTCTCTCAGAAGATTATCGCTCTCAG | | | | KLKVAPQKCVTLAKT |
| | | AGTCTTCATCTTGCCACAACTTGACTTTTACATGTTCCACAACGTATTCAG | | | | STAIQPFRIGPDEIPIK |

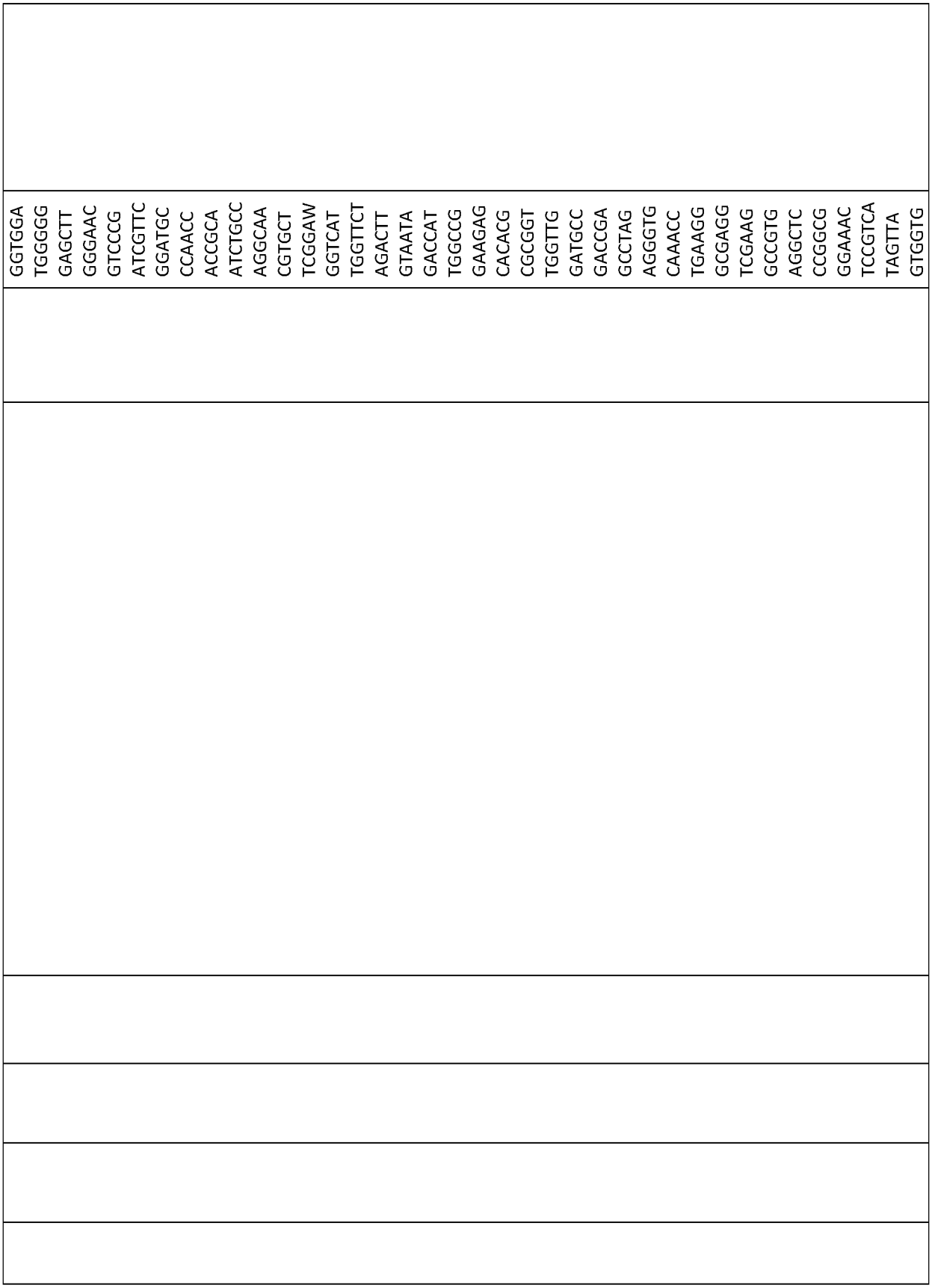
| | | | | | |
|---|---|---|---|---|---|
| | | AGTCAATGACTTGAAAGCGACAGATCAGATGATCCGAGGCCTGATCGA | | | TSMDNITYLGIPISGT |
| | | CAAAGAAGCGCCGACGTCAAACATTCCGGTTTCATTTTTCTACATGCCGA | | | KTSRFAAATGILEKVK |
| | | AGAACAAAGGCGGCTTTGGACTCGTTAAATTGGAACTTCGCCAGCCTCA | | | AQIRVVFASHLALSQ |
| | | GCTCGTTCTCACTAAATTTGCGAGGTTATGGTTAAGTCAACAAGCAGAA | | | 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 |
| | | GAAAACCACCAACAACAAATCCACCTAGACCAAATCCTGCCCCACCTCCA | | | TEQTISVDQIIEEDAE |

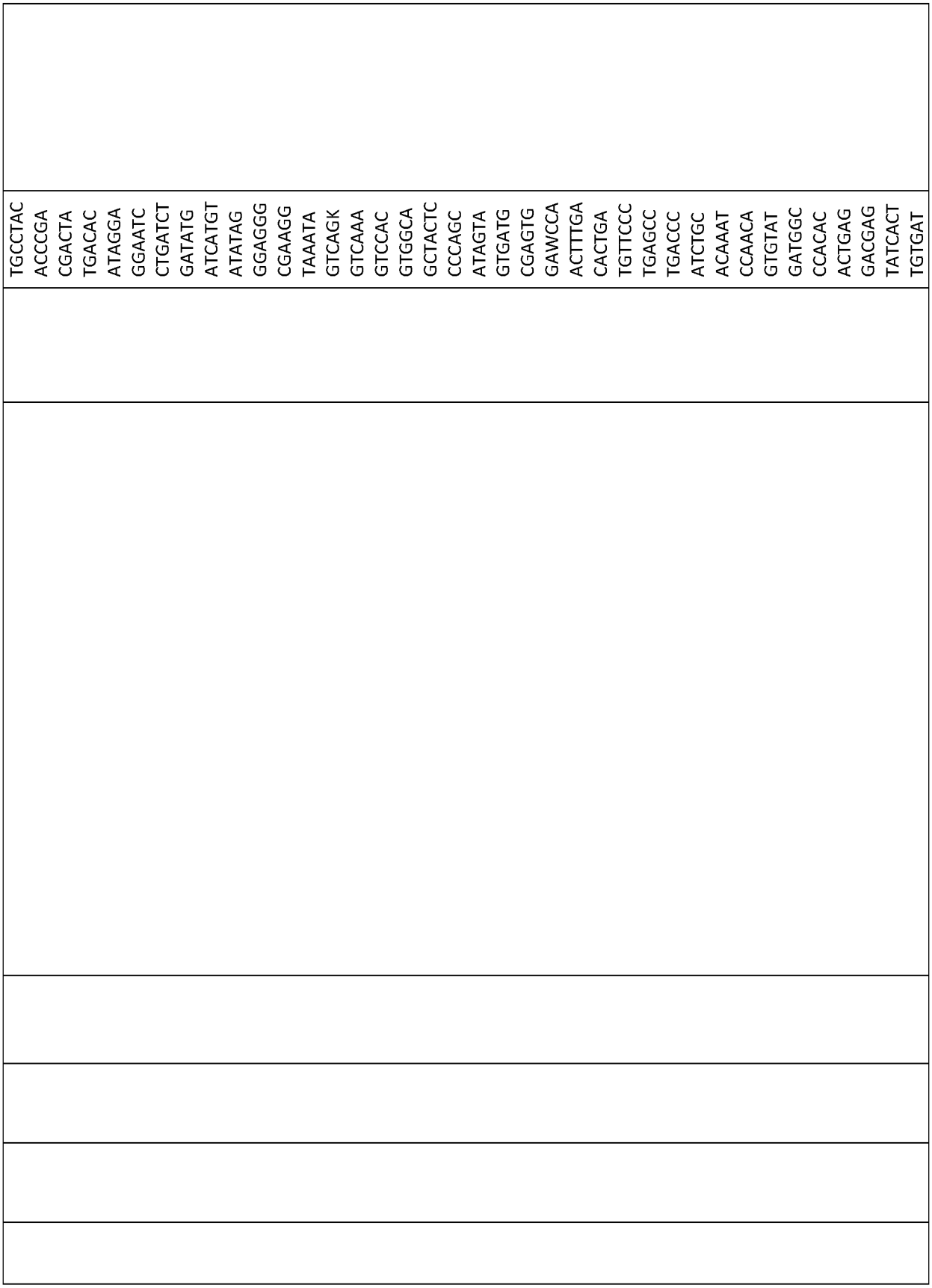
| | | | CCCAAGTAGCTCGCTTCGCTCGCTCACCTAAAACTTTGCTCGCTCGCTTC GCTCGCTCGTCTTAACCCTTTCCGAATAAACACTTACAATTCCCGGCTCG CCCCATTTTTT (SEQ ID NO: 1049) | | | NNAIEQALDQPDEDE FLN (SEQ ID NO: 1417) |
|---|---|---|---|---|---|---|
| NeSL | NeSL-2_CBre | . | Caenorhabditis brenneri | GACTCGCCTTGGGGAAGGTWTTTCAGGGGKSAATTGCCGMAGGCAAG | GACTCGCC | TAAACC | MRYHXSNXPAXRTS |
| | | | | GCAGCCCCCSMMTAGCTTACAAAGTAAGTACMCATTTTCATTTCTTGTG | TTGGGGAA | GGCTCC | DNXWRSIXKDVRRP |
| | | | | AATTCTTTAAACATATTTTTTCTTGTTTTTTGATTTCTTTTTTTCTCTACCTTCC | GGTWTTTC | TCTGGG | DPSTIEEKSRYNRSIGI |
| | | | | CCCAATTCTTCCCCTCATCTTGTGTATACATCCCCCTCCTCCAACCAATCA | AGGGGKSA | AGGAGG | PDSLKXRSSAVRSXSS |
| | | | | ATACATTGACCTCTCTCTTCTGTCAAAAAATCAATACTAGTATATTGTCCC | ATTGCCG | TATGTCA | XPPSGPQDVRLXNSP |
| | | | | TTGTATAGTATTATTTGACGTCGTCTTTGTATTAGGAGTAGGTAACAAW | MAGGCAA | GAGGAC | SLDDRRRLVDCETTL |
| | | | | CTGTGTATGGCTTCAAAAAGCATGCACACAAACWCCTGTCAAAAAGTAWT | GGCAGCCC | ATTCTCC | GSYREWTDKPMMG |
| | | | | TCCCATCMTGTGAATAGCTCAACGACWKGAAGMCCAATGATATGAGA | CCSMMTA | GTGGGC | KMTYAAVTKRAPPRP |
| | | | | TATCATMGCTCCAACMACCCAGCASCCCGCACCTCAGACAATCAMTGG | GCTTACAA | GGATGG | QTGGARLSTNLLADE |
| | | | | AGATCAATCCMAAAGGACGTCCGTCGCCCAGATCCGTCAACTATCGAG | AGTAAGTA | GAGGAG | MEIKYRDTNDIRLVID |
| | | | | GAGAAAAGCAGGTATAACAGGTCCATAGGTATTCCAGATTCGCTCAAAG | CMCATTTT | TAGGGT | LPNPHLIKCPLCKSCIS |
| | | | | AWCGGAGCAGTGCAGTCCGCAGCAKGAGCAGCCMACCTCCGTCAGGT | CATTTCTT | AACGAC | ARGRGANALKYMKR |
| | | | | CCACAGGACGTCCGTCTCWCCAATTCGCCATCTCTCGATGATAGGAGAA | GTGAATTC | CCGTCAT | HIADAHHLNADFVYK |
| | | | | GGTTAGTWGATTGTGAAACAACACTAGGGTCATACCGCGAATGGACSG | TTTAAACA | TCTGGA | CSRCQEHEPENVCG |
| | | | | ATAAACCAATGATGGGAAAGATGACGTATGCGGCAGTGACAAAAAGAG | TATTTTTCT | TGCCTA | AKWIVNHLKRVHGY |
| | | | | CGCCCCCAAGACCGCAAACGGGAGAGCCCCGGTTGAGCACCAATCTCC | TGTTTTTT | AACCAC | TLEDAVSTAKPSTRQ |
| | | | | TAGCAGATGAAATGGAGATAAAGTATCGAGACACCAATGACATCCGCCT | GATTTCTT | CACAAT | QIANAFNDSAPFIDA |
| | | | | TGTCATAGACCTTCCCAATCCCCACCTCATCAAGTGTCCGCTCTGTAAAA | TTTTCTCTA | CTGTCA | RKTSDVPEKKSREAG |
| | | | | GCTGCATAAGTGCGCGGGGAAGAGGTGCTAATGCGCTGAAGTACATGA | CCTTCCCC | AGGCAA | LEKFLAPTKSEDTREK |
| | | | | AAAGGCACATAGCCGACGCCCACCACCTCAACGCCGACTTCGTCTACAA | CAATTCTT | AGTGCC | TPPSTRKSSESSEASI |
| | | | | ATGTAGCAGGTGTCAAGAGCATGAACCAGAAAATGTATGCGGCGCGAA | CCCCTCAT | CCAAAA | QSTIQETLSESSDTLT |
| | | | | GTGGATTGTGAATCATCTCAAAAGAGTACATGGCTATACTCTAGAAGAT | CTTGTGTA | GCACAC | VQEIINISSEDEMDEE |
| | | | | GCCGTATCCACAGCAAAACCCTCTACAAGGCAGCAGATTGCAAACGCCT | TACATCCC | GCGTGG | PPKRRVNVWALIHE |
| | | | | TCAACGACTCTGCTCCATTCATAGACGCCCGGAAAACATCCGATGTGCC | CCTCCTCC | ATCGGT | NGKDAWIDSDLMVI |
| | | | | AGAGAAGAAGAGCAGAGAGGCAGGACTTGAGAAGTTCCTGGCCCCTAC | AACCAATC | TTGGAT | FLESRARGYESCSIIDP |
| | | | | AAAGTCCGAGGACACAAGGGAAAAAACCCCGCCCTCCACCAGAAAGTC | AATACATT | GCCGAC | LNFICTDMSYLTTIVR |
| | | | | CTCTGAAAGTTCAGAGGCATCAATCCAATCGACCATCCAAGAGACTCTTT | GACCTCTC | TGAGCC | RRMEEGYKKIIFPLCA |
| | | | | CGGAGTCGTCAGACACATTGACCGTCCAAGAAATAATCAATATCAGCAG | TCTTCTGT | AGAGGG | NDHWTLVTITGSTAT |
| | | | | TGAAGATGAAATGGACGAGGAGCCACCGAAACGGCGTGTCAATGTCTG | CAAAAAAT | CAAAGT | FYDPMGNEPTETVKK |
| | | | | GGCCTTGATCCATGAGAATGGCAAGGACGCCTGGATAGACTCAGACTT | CAATACTA | CGAAGG | MIDELDLEMQLAPS |
| | | | | GATGGTCATATTCCTGGAATCAAGAGCAAGAGGATATGAATCATGCAGC | GTATATTG | CCGGTA | NSPRQRDSWNCGVF |
| | | | | ATCATAGACCCTCTGAACTTCATTTGCACTGACATGTCCTATCTGACCAC | TCCCTTGT | GGCTCC | VMKMAEAYIKDTQ |
| | | | | AATAGTCAGAAGGCGCATGGAAGAAGGCTACAAGAAAATCATATTTCC | ATAGTATT | CGGCGG | WDLTDVDTDVKTFR |
| | | | | ATTATGTGCAAATGACCACTGGACACTGGTCACGATAACAGGTAGCACG | ATTTGACG | GTTGTC | RSLLTELKAKFNIFAE |
| | | | | GCCACCTTTTACGATCCAATGGGAAATGAGCCAACTGAGACTGTCAAGA | TCGTCTTT | CGTCAT | 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 |

| | | | | | |
|---|---|---|---|---|---|
| | | AAGGTGCGTACAATGAAAAGCACAAAAGAAGAGGGTAGAAGTAGATC | | CAAGGC | EEGRSRSNAKKGPN |
| | | GAATGCCAAGAAAGGTCCGAACTGGCGAAGATCAAAAAGCGAATCGGA | | ACCGTG | WRRSKSESDGRSVSK |
| | | CGGAAGGAGTGTGAGTAAAGGCAGATACTGGCGAGATCCGTCGAACAA | | CTCCAA | GRYWRDPSNKPPHS |
| | | GCCGCCACACTCGAAGATGACCCAGTCGGCTTTAGCTAAGCGCTAAACC | | AAGCAC | KMTQSALAKR (SEQ |
| | | GGCTCCTCTGGGAGGAGGTATGTCAGAGGACATTCTCCGTGGGCGGAT | | ACGCGC | ID NO: 1418) |
| | | GGGAGGAGTAGGGTAACGACCCGTCATTCTGGATGCCTAAACCACCAC | | GGGTTG | |
| | | AATCTGTCAAGGCAAAGTGCCCCAAAAGCACACGCGTGGATCGGTTTG | | GTTTGG | |
| | | GATGCCGACTGAGCCAGAGGGCAAAGTCGAAGGCCGGTAGGCTCCCG | | ATGCCG | |
| | | GCGGGTTGTCCGTCATAGTCAGTGGTGCGCCTACACCCAACTGCTATGA | | ACTGAG | |
| | | CACACAAGGACAACCCAAAATAAATAAGCCAAGGCGGCGTTAGCTTCG | | CCAGAG | |
| | | AGCTAACAAGCTCCCCGAGAGGATGGTTGCCACAGGGCACCATCCTGG | | GGCAAA | |
| | | GGAACGACCCGATCTTTCGGATGCCCAACCACCGCCAATCTGTCAGGCA | | GTCGTA | |
| | | ACGTGCCCCAAAAGCACACGTGCGGAGCGGTTGGATGCCGACTGAGCC | | GGCCGG | |
| | | AGAGGGCAAAGTCGTAGGCCGGTAGGCTCCCGGCGGGTTCTCCGTCGT | | TAGGCT | |
| | | AGTCAGTGGTGTGCCTACACCTAACTGCTATGACAAGCGTATAGGAGGC | | CCCGGC | |
| | | CCGGAAAAACAAGCCAAGGCGGCGTTAGCTGAGAGCTAACAAGCTTCT | | GGGCTC | |
| | | CGTGGATGGGTGCCAGAGGGCACCATCCTGGTGGGTGGATGGGGGGA | | TCCGTCA | |
| | | GCTTGGGAACGTCCCGATCCTTCGGATGCCCAGACCACCGCAATCTGTC | | TAGTCA | |
| | | AAGGCACCGTGCTCCAAAAGCACACGCGCGGGTTGGTTTGGATGCCGA | | GTGGTG | |
| | | CTGAGCCAGAGGGCAAAGTCGTAGGCCGGTAGGCTCCCGGCGGGCTCT | | TGCCTTC | |
| | | CCGTCATAGTCAGTGGTGTGCCTTCACCCAACTGCTATGACATGCGTACA | | ACCCAA | |
| | | GGAGGCCCGGAAAAATAAGCCAAGGCGGCGTTAGCATAGGGCTAACA | | CTGCTAT | |
| | | AGCTTCTCGTGGATGGGTGCCAGAGGGCACCATCCTGGTGGGTGGATG | | GACATG | |
| | | GGGGGAGCTTGGGAACGTCCCGATCGTTCGGATGCCCAACCACCGCAA | | CGTACA | |
| | | TCTGCCAGGCAACGTGCTTCGGAWGGTCATTGGTTCTAGACTTGTAATA | | GGAGGC | |
| | | GACCATTGGCCGGAAGAGCACACGCGCGGTTGGTTGGATGCCGACCGA | | CCGGAA | |
| | | GCCTAGAGGGTGCAAACCTGAAGGGCGAGGTCGAAGGCCGTGAGGCT | | AAATAA | |
| | | CCCGGCGGGAAACTCCGTCATAGTTAGTGGTGTGCCTACACCCGACGAC | | GCCAAG | |
| | | TATGACACATAGGAGAGGAATCCTGATCTGATATGATCATGTATATAGGGA | | GCGGCG | |
| | | GGGCGAAGGTAAATAGTCAGKGTCAAAGTCCACGTGGCAGCTACTCCC | | TTAGCAT | |
| | | CAGCATAGTAGTGATGCGAGTGGAWCCAACTTTGACACTGATGTTCCCT | | AGGGCT | |
| | | GAGCCTGACCCATCTGCACAAATCCAACAGTGTATGATGGCCCACACAC | | AACAAG | |
| | | TGAGGACGAGTATCACTTGTGATACTCAGAGGTGTCCCCCATGATCAAC | | CTTCTCG | |
| | | CAATATCACAGCTAGCGGACCTACCGTGAGGTAGACCCCCGCCGCTGTA | | TGGATG | |
| | | GCAGGCTCGCCTC (SEQ ID NO: 1050) | | 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

| | | |

| | | | | | |
|---|---|---|---|---|---|
| | | CTAAAACTGGGAAATACTCCTCAGACGTCGAAAAGGTCGCCTCCAAAAC | TGGACCAC | GCGACC | NVWSLINETGKDAYI |
| | | CATCGAGAGGAGGAGCTCGAATCAGCAGTGTGTATACTCTGACTGATG | GAAGTGG | CGGAAG | DTDIMMAFLKMRVE |
| | | AGCTGGAGATTACGCACAGAGAAGAAGGTAAGATCACATTCGCGATAG | ACAGTTGT | TATGGA | NCDSVNIIDPLNYQF |
| | | ACCTTCCAAACAAGAATAACATCTTGTGCCCGCTGTGTCGGGAGTGCAC | CGGCGGA | TGCCCA | PARVDLVPLIQRNLE |
| | | CCAAACCCGTGGGAGAGGGTCCAGTTTTACCAAGCATATGAAACTCCAC | CTTCCAGA | ACCACC | DGKKRVVFPICADEH |
| | | GTGAAAGAAGCACCAACTTGATGCCACGTTCATCTACAAGTGTAGTA | GTGGAGA | GCAATC | WTLLTISNGIAAFYDP |
| | | TGTGCAACGAGTACGAACCGGAAAAAAAATGCGGTACGAAGTGGATCC | GAAAAGG | TGATCT | TGSRMSSYIEELVNEL |
| | | AGACCCACCTTCAAAAAGTGCACAACTACAAGTATGACGAGTCTGCAAT | TGTGAAGA | GGCATT | GLIIPKEQDEQPRQR |
| | | AGTTGTCCCAGTACCACCCAACACAAGACAGCAAATAGCTAATGAGTTG | GAGGAGG | GTGTTTC | DSYNCGVFVMKMAE |
| | | AACAATGCTGCCCCCATTCGTTGACATCAGAAAACCGAAAGCTGCTGCTG | TCTAGAAA | GGATGG | AFIQDTEWEMEEVE |
| | | TTGAGGAGAAGAAGACTGAAAATGGTGCTCTGTTAAAATTCCTGACCAA | CACTTCGG | TCTCTGT | EDVKNFRRNLLEELK |
| | | GTCCAATAAGGACGATCAGGTAAAATCCCCATCGGAWGATATTCCAGA | CTGTCTAG | CTCTAG | PNYEIFAEKIKYYNSP |
| | | TGCGGAAAGCCCTGAAAAAGAAACTCAGGCGCTCACTATCGATCCGAA | GACCAGTT | ATCTGA | GKSFAQSRPTSRSSQ |
| | | AGGGAACAACTCACCATCAAAAAGCTCAATAAGATCGAGCCAGTCCTCA | CCTGAGTG | AATAGA | CAVCPTCSRSATPM |
| | | GCTTCCTCCGTTTGTCAAGAAATCCAGGAAATCATCACGTTGAGTGAGG | GAAAGAG | GCTCTG | MDVGNMEVDPVPQ |
| | | ATGAAGACCCAAAAGGGGCTCGTCCAAAACCAGGAATCAACGTGTGGA | GAAGGTCT | GCCTGA | QQETPKSREPEQDEG |
| | | GCTTGATAAATGAAACGGGGAAAGGATGCATACATTGATACAGATATCAT | 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 |
| | | AGACAATTCACTGGTCCAGAGATCAAGTCGTCTCACCTGGGAAGTTTC | TTTCTGTTT | CCGTCAT | VVENHFRQTLKAQP |
| | | ACCCACTTGTGGGCGAAACTGAGGAGATGGAGGTGACGTGTGACAGCC | GACCTCCT | AGTCAG | VTEEALNTVCSGIKKA |
| | | CACCAACGAAAGAGCCCCACTACGGAACCGAAAGTGACTCCAAGCCTGC | GGAATTTG | TGGTGT | KVDPSIEGPISSGEVK |
| | | CAGCAATGAAAATTGCTAGCCCAGAAGTGACGAAAAAGCAAACGTCAA | CGAGGAG | GCCAAC | AILAKIKDTSPGTDGV |
| | | AGAAGAAGGGAAAGTATGGCAAAAAGAAACAGSAGACAAAGAAAGCT | TGCGCGAG | ACCCGA | KYSDLKWFDPEGERL |
| | | CAGCCGCCGAAAGGGGGAGCCAACAAAGAAAGCTCAGCCAAAAGGAGA | 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 | IRRSLAAFGYPESVINI |
| | | 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 |
| | | CGAACGCTTTTGGGAAGCGTGCCCCACGAGTTGATCAGAAGATCACTAGC | 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 | | | |
| | | AGAGAGCGTGAGAAGAAATCCGAGGAAGAAAAAGAGAAGAAGATCAAA | | | |
| | | GTCCAAAAAAGGCAAAACTTGGCGAGGATCCAAAAAACAAACTGATTC | | | |
| | | CCGGCAATCCGGCAAAAGCAATCAGAATCAGGGCTTCCAAAGAAGCGT | | | |
| | | TGGACAAGGCGTATCACGGTGAATACCGTCAGATAAGCCCCCAACATAA | | | |
| | | AAATAAAAGTCGGCGTTAGCTAACCACTAAACCGGCTCCTCATTGGGGG | | | |
| | | AGAGTATCATTCCGGTGCTCTCCGTTTGGGCGGTAGGGAGGAGTTGGG | | | |
| | | TAGCGACCCGGAAGTATGGATGCCCAACCACCGCAATCTGATCTGGCAT | | | |
| | | TGTGTTTCGGATGGTCTCTGTCTCTAGATCTGAAATAGAGCTCTGGCCTG | | | |
| | | AAGAACACACGCGCGGACCGGTTGGATGCCGACTCGATCTGGAGGGTG | | | |
| | | CAAACCTGAAAGGGAAAGTTGAAGGCCGTGAGGCTCCTGGCGGGAAAC | | | |
| | | TCCGTCATAGTCAGTGGTGTGCCAACACCCGACGACTATGACATAGTTG | | | |
| | | GAGGAATCCTGATCTGATAATAATCATTGTTCATATAAGGGAGGGGGAT | | | |
| | | GGTAAATACCCAGGGTCCGAAACCATCAAAGCAGCTACTGACCAGCATA | | | |
| | | GTAGTGATGAACACATAGACCCTGGGGTTCCCTGAACTCGACCCATCTG | | | |
| | | CACAAACCCACTTTGTACAAATGAACCAAACTGATGAAGAGTTTAATGA | | | |

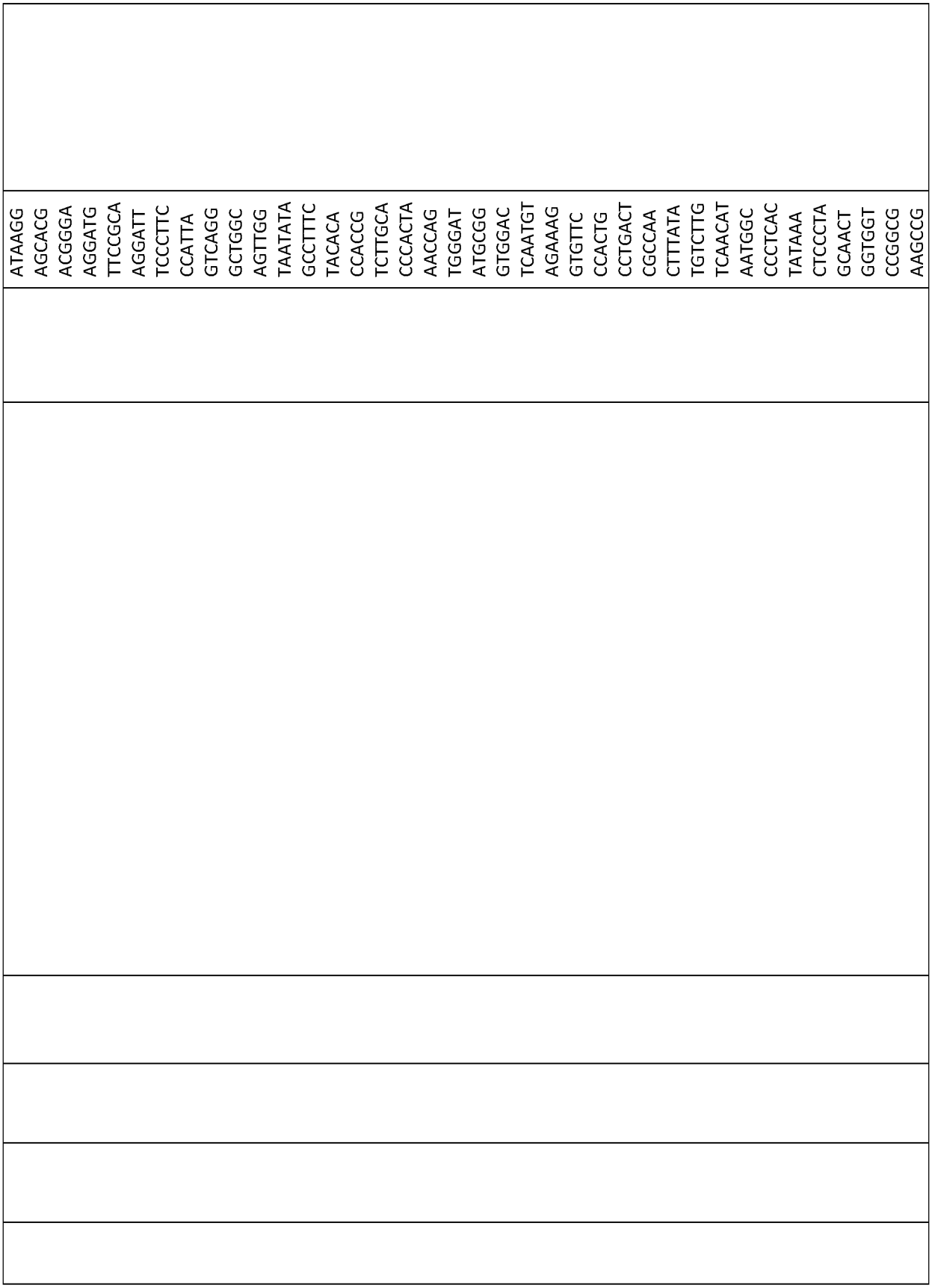
| | | | | TTTCTTACATCACAGCTAGCGGACCTACCGTGAGGTAGACTCCCGCCGCT GTAGCAGGCTCGCCATTG (SEQ ID NO: 1051) | | | |
|---|---|---|---|---|---|---|---|
| NeSL | NeSL-4_CRem | chrUn | Caenorhabditis remanei | CCCTTTTCTATCGTATTAACTACGATAACCGCTCATTTGAGTGTAAAAAA | CCCTTTTCT | TAACAT | MTKTEWSWRHRSRS |
| | | | | GGTTCCCCCCTCCTCGCCTGCCTTACCCACGCATCTCTGCCTCTGGGAAG | ATCGTATT | GCCTTG | RSVGIVVKIDTSDYAN |
| | | | | GCGGAGGGTCAACTTGCGGGTCTGTGGATTTCCTTTCCTATCCACCGCCC | AACTACGA | GAAGGC | VRVHVAADLSNEDG |
| | | | | ATATTCTCTGTCGAAAGCCTACCTAGATCAGCCGGGAGTTTTTCCTATCC | TAACCGCT | ACCACG | HTSHNNGIILPIPMKP |
| | | | | CATTCAGGCGATCGCTCAAGGCTGTTTTATCGACACTCCTTCTTGACAAG | CATTTGAG | CCAAAA | SVDRFCQIQYPPRGY |
| | | | | TATTTATTTCTTGACAAATTCTATTTTCCTTTTATCGATTTTCTCTTATTTA | 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 |
| | | | | TGGAATGAAGAGGCACAACCTCCCTACTACCACACAACAATCATGGGA | 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 |
| | | | | CTCCGCATCACTTTGTTTGCATGATGACGTATCCAACCGACACAACATCTAT | CACTCCTT | ATATCA | GAGKHISKWIKPHM |
| | | | | AGATGCTTCGTCAAAGGTTGCACAGCAACATCACAAGGTGGTTGGGGA | CTTGACAA | AACAAA | RKKHNRDAPTNFKM |
| | | | | GCAGAGGACCTTAAGTACCTGACTGTCCATATCAGACAAGAACACAAAA | 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 |
| | AGTGTCACGAAGGCTCTCAGCGTCCTGAAAGAGTCACGACACAAAAAGAG | TTCTTGTG | CAAAAC | MITPGGEHKGTKIPM |
| | CTGGAAGTGTTGAGAGAAGAAGAAAAGGCTAACGCTAAACAAAAGTCT | AAAAT | ATCAAG | VHTKRGLIQKINRKAK |
| | AAACTTCATCCTTTCTTCACCAAAGCCCCCTCATATAGATGGTGTGAAACC | (SEQ ID | CACGAC | KAKPMHLDESTIIEAS |
| | AACAGTACGGAGAGAAGAACTATCAAAAATGATTACTCCCGGAGGAGAACA | NO: 1175) | GCAAAA | QLDVITIDDDDEDDN |
| | TAAGGGAACAAAGATACCAATGGTCCACACCAAGCGCGGTCTCATCCAA | | AGGGGT | MTPMRRRFNTWCL |
| | AAGATAAACAGAAAAAGCTAAAAAGGCTAAACCAATGCATCTTGACGAA | | AACTTTG | DHETTQEAWLTDDVI |
| | AGTACCATCATAGAAGCGTCACAGCTCGACGTCATCACTATTGACGACG | | GGCAAC | NWYLKDLCFGNEQY |
| | ACGACGAAGACGACAACATGACACCAATGCGAAGAAGATTCAACACTT | | TAATTAA | MLVDPLVWLIYKMG |
| | GGTGTCTTGACCACGAGACGACTCAAGAAGCATGGTTAACTGACGACG | | CGGATA | GMAGVEQRFKSKKT |
| | TAATCAATTGGTACTTGAAAGACCTATGCTTTGGTAACGAACAATACAT | | CCTCCGT | CLFPICEADHWILLVF |
| | GCTCGTAGACCCACTAGTATGGCTGATATACAAGATGGGAGGAATGGC | | GTATCA | DETNLCYANSLGSQP |
| | AGGCGTCGAACAAAGGTTCAAAAGCAAGAAGACGTGCCTATTCCCAATC | | 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 |
| | | AAGATTTCAAACAGTGTGACATACCGATCGCGATACTAGAAGAATTCTA | | | 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 |
| | | GATATGGCCGTTCTAGCAGAATCAAAAGAGCAGCTACAAAAGGAGCTT | | | 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 | | |
| | | GGTCGAGTAGAGCACACAAGCCAAAATATCAAGTATGACGCAAAAATGGG | GGGTAA | | |
| | | TAACCTTGGGCATCCAATCAACGGATACCTCTGCGTATCAGGCAAAGTC | TTTAGG | | |
| | | GCCACCAAACTGTACTACTCCGAAAAAACCAAGAAACATGATTTTCCCAC | CAACCA | | |
| | | TCCGTTAAAGCATCTCAACCAAGCTAAAGCGGTAAGGTTATCATGTCAA | ATCAAC | | |
| | | AAGGTGTAGCTACAGCAACCTAAAGCCCGAAAGGTAGGGCCGTATAAA | GGATAC | | |
| | | AAGACCTACACCCTCCAAGACCTAAACCCACGAACTCGAACGACCTACA | CTCCGT | | |
| | | GGAAGTCCGTGAATGGAGAGAAATATCTCACCAAATCTCTTCCATTCAC | GTATCA | | |
| | | AAAGGCTAACTGGTCAAGTAGAGCACAAGCTAAGCCTCCAAGCACGAA | GGCAAA | | |
| | | GTGATATGGGTAATTTAGGCAACCAATCAACGGATACCTCCGTGTATCA | GTCGCC | | |
| | | GGCAAAGTCGCCACAAACACTGTACTACTCCGTTACTCCCAAACACATG | ACAAAC | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | GATCTCCTTCTCTCACCAAAAAGCTTTATAACCAAGCTAACGGTGGAAAG<br>GACATCATGTCACGAGGAGTAGCTACAGTAACCTCTCTCTTGAGACTGC<br>AAAGTCGAGGATGGATTGGGAAGGCCGCGAGGCAAAAGGCGGGTAAC<br>TCGGCCAGACGCTAGTGATCTTCGGATCCGACAGCCCTGGCCTTAGAGG<br>AACCCTGGGATAAGGAGCACGACGGGAAGGATGTTCCGCAAGGATTTC<br>CCTTCCCATTAGTCAGGGCTGGCAGTTGGTAATATAGCCTTTCTACACAC<br>CACCGTCTTGCACCCACTAAACCAGTGGGATATGCGGGTGGACTCAATG<br>TAGAAAGGTGTTCCCACTGCCTGACTCGCCAACTTTATATGTCTTGTCAA<br>CATAATGGCCCCTCACTATAAACTCCCTAGCAACTGGTGGTCCGGCGAA<br>GCCGGTTCTTGCCACTATTGCGCCCCAGGCTCGCC (SEQ ID NO: 1052) | | | ACTGTA<br>CTACTCC<br>GTTACTC<br>CCAAAC<br>ACATGG<br>ATCTCCT<br>TCTCTCA<br>CCAAAA<br>AGCTTTA<br>TAACCA<br>AGCTAA<br>CGGTGG<br>AAAGGA<br>CATCAT<br>GTCACG<br>AGGAGT<br>AGCTAC<br>AGTAAC<br>CTCTCTC<br>TTGAGA<br>CTGCAA<br>AGTCGA<br>GGATGG<br>ATTGGG<br>AAGGCC<br>GCGAGG<br>CAAAAG<br>GCGGGT<br>AACTCG<br>GCCAGA<br>CGCTAG<br>TGATCTT<br>CGGATC<br>CGACAG<br>CCCTGG<br>CCTTAG<br>AGGAAC<br>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 | GTTCTTG CCACTAT TGCGCC CCAGGC TCGCC (SEQ ID NO: 1298) | TGAGTG | MNVDLDATIKSIGM |
| | | | | ATATCAATGCTAAAAAAAAAAATCGTGCKCATCAGGCGCACGAAAATAATG | TTTTTAAAT | | TGCTAC | NTKETTYPNSQLRVE |
| | | | | GACACAACTCGTCGACCTGCTGTCGACTCACAGAGAACCTCAATTTGGA | GTGTTTGA | | GAGGCA | TTPCTSTTIMHASCN |
| | | | | AGAATGGGAAGCCTATAATGCTACAATTCCGCCAACCCCTATTTGAATG | ATATCTTA | | GCGCTG | TTSTISYSPLPSAVSLP |
| | | | | ACAGATAGTCAAATATCAAAAAATATACAAACTGCTGTCAAGCGTGACT | AATTATCA | | GTAATT | ESPASSITITTTDDNC |
| | | | | CACTTCCTTCCAATCGAAAAATAGGAAKATGTAAGAAACATGAAAGTCA | AATCATAT | | GCATCG | DIIETPYPLPQTNGDL |
| | | | | AGCTGAAAAACCAATAATATGTCCTAAAATAAAACAATTTGAAAATATG | 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 |
| | | | | TATTAAATTAATATGTTTTTTATTAATTATTAAATTTTTATAGTGGGGGGAAA | ACCTCAAT | | AAGCCG | IPFNEQNNVDCQPC |
| | | | | TTAATATACTTGATCCCAAGAATCAACTGATGATGAAGAATATGTTATTT | TTGGAAGA | | AAAGGC | HPKKGKDVHTIVLIKY |
| | | | | CAAAATACATACAAGAAGCTGGAAAAAACAAATCAATCGCTACAATGAA | ATGGGAA | | AGCATA | ADIFNHIEAHSHVVQ |
| | | | | TGTGGATCTCGATGCAACAATTAAAAGTATTGGAATGAACACAAAAGAG | GCCTATAA | | TATATG | TAITDNMKTYLRLTKE |
| | | | | ACGACCTATCCAAATTCACAACTGCGAGTTGAGACGACTCCCTGTACCTC | TGCTACAA | | AGACAA | NXFYCSYRNNKKKNK |
| | | | | AACGACTATTATGCATGCATCTTGCAACACAACCAGCACTATATCTTACT | TTCCGCCA | | TTTAAAA | CKKAFNLESNMMDI |
| | | | | CTCCATTACCATCGGCTGTGTCACTTCCCGAAAGCCCTGCCTCGTCAATC | ACCCCTAT | | AAAAA | TEHMKTHTGYSFDX |
| | | | | ACAATAACCACACAACAGACGATAATTGCGATATTATAGAGACCCCTTACC | TTGAATGA | | (SEQ ID | NLNILCYCGIWKPFTE |
| | | | | CATTACCTCAAACAAATGGTGACTTGAGTGAAATATTAAAGGATATAGA | CAGATAGT | | NO: | LIAHIKTEHLQEYINSI |
| | | | | AGCTAATAAGGACACCACCATGTCGAATAAAGTATTGGACTGTGACTCT | 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 |
| | | ATATTAACTCAATACCAAACAAAGAAAATATCCATAAATACTACTACCATA | ATACCTAT | | QKKKSAELLCVKRLKF |
| | | GTTTCCCCTCTAAACTTTGCTGGGATACTTGCATCTGGCGAAACTCAAAA | AAAATCAC | | QISQKTEIGKIWEND |
| | | TATCCCCGATGAAGAAATAATTAAACCCAGAGATCTGCCAGAAAATCTT | AGCCGAAT | | DVDHRPPMARFLKT |
| | | GCCTTCAACCGAAACATCGAAAATGAATTAAGTTGATGGTCGCAGCACT | AAATTCCC | | FASQDCPVSNTSSINL |
| | | TGGTCAAAGCATATATCTTCTCTTATGCTGTTAAAACATCAACCATCTTCA | ATCCGTTC | | PYYMDTDTDXCTDC |
| | | TCAATCCTTATACTTGCAATGCTTTGATCCAGTGCAACTACAAAACTTTCT | TAAGCAGA | | ENLSHIMKNLDSSAP |
| | | TTGAAACCTTCCCTTTCAAAGACTTTGCCAAGTGGAACGAGATAGTCCTG | AACCGCTA | | GMDLITGGDWKKISP |
| | | CCAATTCACAACAACACTTCTTCTTCTTCTTCTTGAACAAGAAA | 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 |
| | | ACCAAAGACGGATATATAATACATCTGCAAAAGAAAAAGTCGGCGGAACT | 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<br>TCAATACCAACAAAGCTGAAACCTGACATGGTAATTAGGTCTAAGAGGA<br>ATAAAGATATACATGTCCTAGACCTAAAAGTGCCATATGACTCAGGAGA<br>AGGCTTTGAAAAAGCGCGGGAAGACAACTATGTTAAATACAAAGATCT<br>AGCCATTGAAATTGGAAAGGCATTTAATCAAAAAGCGACTATATCTGCT<br>GTGGTGATTGGATGCCTGGGCACATGGGACAAGAAGAACAATGCCGCT<br>CTTTCCAAAATCGGGTTGACTAAGACCGAGATCATATCTCTTGCCAGGAT<br>AGCATGCCCAAATGCGGTAATCGCATGCTATCACATATACCGTGAGCAC<br>GTCTCATTTACAAAGAGTGCCATGGCCCTCCCCTTCAGCCTTGCATGAGT<br>GTGCTACGAGGCAGCGCTGGTAATTGCATCGGCGTTGCAGATTTGTGTA<br>CGATAGATAAAAACCAATAGTAATAAATGCTGAGCCTAGCTCGCATATC<br>TAAGCCGAAAGGCAGCATATATATGAGACAATTTAAAAAAAAAA (SEQ ID NO: 1053) | GTAGGTAACTATGACTGCAAAATAATAATTCTACACCTATTGTTGATAAC<br>TCATCTCGTGCGCAAACGGAGCATGTTATTTCTAATCATTTCGTCACACA<br>GGATTCTTCTAATTCTGATAGTAATATTATAGATAGAGATAGGAACCTTG<br>TTGATTTAGATGCGTCAATAACTTCTCCTACTATTATACAGCCAGAGGAT<br>AGTAAGATATCTGAGGATGAGGACTTCATCTTAGTCAATAGGAAAAAGA<br>GCAAAAATAAGAAAAAATCTAAGAAAACAACTGAAAATAAAAATGAAA<br>TTCCTATTCAAAAGAGTAAAGATAAGAAAAAGAAGTCTAAAATTAATAC<br>CGAAAAACTAACTGAAAATATTACTACTTCTGAAATACCACTTGAAATTG<br>CTCCTTCCATACCTTTACCTTCAGCAAGTACCTCGGGTTCTCAACAACCG<br>GCCAATCCTCCAGAAGACGCTACTCTAAGTGATACGGATCTCTTCCTTAC<br>ACAGGATGATCCCGATAGTCTTATTCTTTCTGGAAGTACTCAACCAACCT<br>TTGTTGACCTCAACCCTTCACAGCAATCGGAACTTCCTTCAAATACTGAC<br>AGCCAAAGATTTGAGGCGGGTGAAACACCCAAAATCATAACTTCTTACA<br>GGGATGACCTTTTCTACTCTACAGTCCTTCACTACAACTCAGATACAGGT<br>TACGGTATAAGTGTTGACAATGGGGAGCAGAGGTTTCGAATTCTTGCTA<br>GGAATCTTGTCAGGAAAACCAAGGATAAGTTCCCCTCTTTATATGCTGG<br>ACAAGTAATTAGACACACAGTCTTCTTCAATCACTTCAACCAGGCATACT<br>ACGCCAATAATATAACTGATAGTAAAGGTAATCTAATTGAGTTTTCTGAT | GTAGGTAA<br>CTATGACT<br>GCAAAATA<br>ATAATTCT<br>ACACCTAT<br>TGTTGATA<br>ACTCATCT<br>CGTGCGCA<br>AACGGAG<br>CATGTTAT<br>TTCTAATC<br>ATTTCGTC<br>ACACAGGA<br>TTCTTCTA<br>ATTCTGAT<br>AGTAATAT<br>TATAGATA<br>GAGATAG | TAAATTAA<br>TATGTTTTT<br>ATTAATTA<br>TTAATTTTT<br>ATAGTGGG<br>GGGAAATT<br>AATATACT<br>TGATCCCA<br>AGAATCAA<br>CTGATGAT<br>GAAGAATA<br>TGTTATTT<br>CAAAATAC<br>ATACAAGA<br>AGCTGGAA<br>AAAACAAA<br>TCAATCGC<br>TACA (SEQ ID NO: 1176) | TGATCC<br>GTGTGT<br>TTGTGTC<br>GTATGA<br>TTGTTTC<br>CGTGTG<br>TGTCTAT<br>ATTTTTC<br>TTTTTTA<br>TACTTTC<br>AATTACC<br>TCGTTGT<br>AATGTT<br>ATAACTT<br>CATATG<br>GAATAT<br>ATGTAA<br>TTTAGTT | TTGRNLGQWSCYSR<br>SIQQSNYSFKLSSTEV<br>GELVEQSPAPLQSPQ<br>FSNNYNNLNINNNLY<br>YSLNTFNQSNNLCCL<br>VNIEFFPTQHLLGDIV<br>NSGCINYMNNYNNF<br>DNINLYINSNVLSYNN<br>YNHSFLASPYTTNITE<br>HADINMHVQEVNM<br>QQDNNTQHAITQQV<br>SLQATSLQHTLDEMI<br>VQFNTAVRLKKKHKV<br>AKIFRGHNHRKDLPT<br>LPAREQYKTKPKLAIR<br>EVLHRKTTATSSPSEN<br>AIKAFFSSYSRPAELFT<br>GQELLESSWFPVHPE |
| NeSL | R5 | AY216701 | Girardia tigrina | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | 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 |
| | CAACAACAACTTATACTATAGTCTCAATACTTTTAATCAGTCTAATAACCT | 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 |
| | CCATTTTGAATATATAGTACAAAATAACATATGCCCATCCGAGGGGAA | AGTCTAAA | | QKIHAYETFIHSQLPF |
| | GGCTAGTAGGACCATTATGATTCCCAAACCGGGAAAAAGTGACTATTCA | ATTAATAC | | HLRHSRIPFSDFITNR |
| | GATCCTTCTTCTTGGCGGCCCATTACAATTACCAGCGCAGTATACAGACT | CGAAAAAC | | KTNKTTNNSNDSEKS |
| | TCTCATGAAATATCTTACATGGGAGCTGTATAACTGGATTCTTCTTAATC | TAACTGAA | | IQKAYDPESGQLFLN |
| | AGATGCTGTCCAGGAGCCCAAAAGAGTTTAGGGAAGTTTGAGGGATGTC | 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 |
| | | GAAGCTGGATATAAAGGAGCTCCTCCAGATGATGGATGAAGACTAA | 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 |
| | | CACTTAAATTTAAAGGGTTTTTGTAAATATTAATTTTTTCTCAAGCCATTTCA | 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 | | |
| | | ATCAATTCCTCCTGCGAGCAGTGGAAAAGATTTGTGCAATATGCACCCA | CGGTATAA | | |
| | | CAATTACACCTGGGCCGATTCCAGACGCGGAGAGCGAGGACGATCAGG | GTGTTGAC | | |
| | | GGACGAGCGACAATGGGCCAACAGCTGCTACAGTGCAAGGACCGGTGA | AATGGGG | | |
| | | TTGGCGATGAGGAGGAGGAACTTCAAATCTACGATTCCGGCCTTGACG | AGCAGAG | | |
| | | AGTCCAGCGATGATGAACCCGACCCAGATGATGCTGAATTACTTTTCAC | GTTTCGAA | | |
| | | AATTGACATAGAACAATATTTGAATTCTGTGATAACAGACTGATCCGTGT | TTCTTGCT | | |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| | | | | 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 | |
| | | | GTTTGTGTCGTATGATTGTTTCCGTGTGTGTCTATATTTTTCTTTTTTATAC TTTCAATTACCTCGTTGTAATGTTATAACTTCATATGGAATATATGTAATT TAGTTTAGTTTAGTTTAGTTTAGTTTAGTTTAGTTTAGTTTAG TTAGTTTAGTTAGT (SEQ ID NO: 1054) | | | |
| | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | 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 |
| | | | | TGAGCGCCGCCCTGCTGCTCCTGAGAAGGCTGCTACCAGAGGAGCCCCCC | | 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 |

| | | | | | |
|---|---|---|---|---|---|
| | | | | CTTGCAC | GPASAPQPASPEPQX |
| | | | | CCCAATC | VRGDQPPTEGSATPA |
| | | | | TGGGTC | SRTDDATKRTSPASRI |
| | | | | TATGCC | PTLDPAVRGITATSQ |
| | | | | GGTTAT | VSDLTRCLSDLIKTIRH |
| | | | | GCGATA | NTDTRRXSAPPQVTS |
| | | | | TGTATGT | CRPAVGATSTAPQA |
| | | | | ATCTCTT | ARRDPANGGASRSP |
| | | | | CATCCTT | QIPRPDPAPGRPNTS |
| | | | | GCAACC | SKVTQRDSDRQKPH |
| | | | | GATACC | APPRTPQPDTTRRRT |
| | | | | TGTAATC | RTIPSASKHDRAPTKP |
| | | | | CCTCATA | NTGVSRTPLPPGRSS |
| | | | | ACCCAA | AASETPRAAPPPPHQ |
| | | | | GCCTGA | DPRLKTHLNTAPQSE |
| | | | | CCCCAG | GQQGHRLSPQHLNP |
| | | | | ATGTAC | RRQRSRRNDGREQR |
| | | | | AGTACC | VATPWQSAWMEEL |
| | | | | TTCCCTC | AKAEDFETFDTLMDR |
| | | | | TTAACTC | LTAELSAEITARRREP |
| | | | | GTGTAT | QEASRATRRFPAPTR |
| | | | | ATTTAAT | NNTAREGRRGDVGR |
| | | | | TTTAAAC | RYDPAAASRIQKLYRT |
| | | | | ATTAACT | NRTKAMREILDGTSS |
| | | | | TTAATAA | YCAIQPERLYSYFKDV |
| | | | | AATTTTT | FDHEAQTNLQRPECL |
| | | | | AAA | LPLPRINLTEDLERDF |
| | | | | (SEQ ID | SPQEVQARLMRTKN |
| | | | | NO: | TAPGKDGIRYHLLKK |
| | | | | 1301) | RDPGCLVLAAIFTKCK |
| | | | | | QFHRVPRSWKKSMT |
| | | | | | VLIHKKGERDDPGN |
| | | | | | WRPISLCSTIYKLYAS |
| | | | | | CLAARITDWSVCGG |
| | | | | | AVSSVQKGFMSCEG |
| | | | | | CYEHNFLLQTAIQEA |
| | | | | | RRSKRQCAVAWLDL |
| | | | | | TNAFGSIPHHHIFATL |
| | ACAGCANGTAAGGGGGGGACCAACCGCCAACCGAGGGAAGCGCAACCC | | | | |
| | CCGCCTCGAGGACTGACGATGCCACCAAAAGGACCAGCCCCGCCTCCAG | | | | |
| | AATCCCCACGCTGGACCCTGCCGTGAGGGGGATCACCGCCACCTCSCAG | | | | |
| | GTCAGCGACCTCACCAGATGCCTCAGCGACCTCATAAAAACCATCCGGC | | | | |
| | ACAACACGGATACGAGACGGKGCAGCGCTCCCCCACAGGTMACCTCAT | | | | |
| | GCCGCCCTGCCGTAGGAGCAACTAGCACTGCCCCCCAGGCCGCACGGC | | | | |
| | GAGACCCAGCCAACGGAGGAGCCTCCCGCAGCCCCCAGATCCCACGGC | | | | |
| | CGGACCCCGCCCCCGGGAGACCCAACACCTCCTCCAAGGTTACCCAACG | | | | |
| | AGACTCCGACCGCCAAAAACCCCATGCCCCCACCGAGGACCCCCCCAGCCG | | | | |
| | GATACCACCCGCAGAAGAACCAGAACCATCCCCAGCGCTTCCAAACACG | | | | |
| | ACCGCGCCCCCACAAAGCCCCAACACCGGTGTTTCCAGAACCCCACTCCCT | | | | |
| | CCCGGAAGATCCAGCGCTGCCTCGGAGACACCGAGAGCTGCCCCCCCCCC | | | | |
| | CACCGCACCAGGACCCCCCGCCTCAAGACCCACCTGAACACCGCTCCCCA | | | | |
| | GTCCGAGGGACAACAAGGCCACAGACTGTCTCCGCAGCACCTGAACCC | | | | |
| | GCGGAGACAACGCAGCAGGAGGAACGACGGCCGCGAGCAGAGGGTCG | | | | |
| | CCACGCCGTGGCAATCCGCCTGGATGGAGGAGCTGGCAAAGGCTGAGG | | | | |
| | ACTTCGAGACCTTCGACACCCTGATGGACAGACTGACCGCAGAACTGTC | | | | |
| | TGCGGAAATTACAGCCAGAAGGAGGGAACCCCAGGAGGCCTCACGGG | | | | |
| | CCACGCGCAGATTCCCCGCGCCGACCCGTAACAACACTGCCAGAGAAG | | | | |
| | GCAGGAGAGGGGACGTCGGCCGCCGCTACGATCCGGCAGCTGCATCCC | | | | |
| | GTATTCAGAAACTATACAGGACGAACCGGACGAAAGCCATGAGGGAGA | | | | |
| | TCCTCGACGGGACCTCCTCCTACTGTGCCATCCAGCCCGAGAGACTCTAC | | | | |
| | TCCTACTTCAAGGATGTGTTCGACCACGAGGCCCAGACCAACTTGCAAC | | | | |
| | GCCCAGAGTGCCTTCTCCCGCTACCCCGGATCAACCTCACGGAGGACCT | | | | |
| | GGAGCGAGATTTTTCCCCGCAGGAGGTGCAGGCGAGGCTGATGAGGAC | | | | |
| | CAAAAACACTGCCCCTGGAAAAGATGGCATCCGCTACCACCTGCTGAAG | | | | |
| | AAGCGAGACCCCGGCTGCCTGGTGCTTGCTGCCATCTTCACCAAATGCA | | | | |
| | AGCAGTTTCATCGCGTTCCCCGCTCCTGGAAAAAGTCCATGACCGTGCTC | | | | |
| | ATCCACAAAAAAGGCGAGCGAGACGACCCCGGCAACTGGAGGCCCATC | | | | |
| | TCCCTCTGCTCCACCATCTACAAGCTGTATGCCAGCTGCCTCGCGGCAAG | | | | |
| | GATCACGGACTGGTCAGTGTGCGGGGGCGCCGTCAGCTCAGTACAGAA | | | | |
| | GGGTTTCATGTCCTGCGAGGGATGCTACGAGCACACAACTTCCTCCTTCAG | | | | |
| | ACGGCCATCCAGGAAGCCAGGAGGTCCAAGAGGCAGTGCGCCGTAGCA | | | | |
| | TGGCTTGACCTGACCAACGCCTTCGGGTCCATACCCCACCATCACATCTT | | | | |
| | TGCCACCCTGGGAGAGTTCGGGATGCCAGAAACCTTCATCCAGATCCTC | | | | |
| | CGGGACCTCTACAAGGACTGCACCACCACCATCCGCCGCCACGGACGGA | | | | |
| | GAGACGGACGCCATCCCCATCCGCCGGCGTGAAACAAGGATGCCCC | | | | |
| | CTCAGCCCCATCATCTTCAACCTGGCCATGGCAACCGCTCATCCGAGCCAT | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 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 CCACCCCACCCCCGCCCGGACAACCACGGACCCTGCTCCGCAGCAGGACC CCACGCACGATGGCCGCTAAATCCGTCGCCTGCCCTCACGATGGATGCG CCAACAAGTACGCGTCGGAAGCCTCCCTCCGAAGACACATTAAGAACAA ACACGCTACAGATGAGGAAGGAGATGAGACCTCACACTCCTGTCCCCAC TGCCACCGACCTTTCTCCACCGCCCCGCGGGCTCAGCGTCCACATTGGCA AATCGCACCGTCAGGCCCCCCCTGAGCCGACGCGCCCCCCCCGGCCCC GGCCCCTGCCGATCCCGGCCTCGATCCCGACCCCGGCCCCCACCGTGACG CCCCCCAGCCGTGATGACGAAGACCGCGAGGAACCCGACGACGACCCC GTGGAGATCGCGGACCTAAGCTGCCCTCACTGCGCCCAGGCCCTCCCGT CGGCCCACGGCCTCGCCAACCACCTTCGCGCCTGCAAGGACCACAGGGT CCCCGCCCCTGGAGCACCCCGCTCGGGTCCGCCCAGCTCCAGGTACTGG ACTGCTGTCGAGCACCACCGCTATGTGGAGGCCATGGCGCGCTTCGCG GATCACCCCGACCTACTTGCGCGCGCGGCTGCCCACATCGGGACCCGCA CGTACAAACAGGTTGACTCCCACCGCACCAAGGTGATCGCGGCGGAGC GCGAGGGCCGCCCTGTCCGCACGCTCGACCCCACGATGGACTGGCGCA TGCGGCCCTACTGCGCCAGCACCACGGCCCGGTGGCTGGCTGAGCAGG GGCGTAGCCCAGTAGCGCCCCGCTCGCCCTGCCCCGAGCCCCACGCCCC GCCGCCTGCAGCCGCGCTGCTGTACATCCCGGCCACGCCCCCCGCGCCA ACGCCCCGTGCCCCAGTGGCGCCTCCCAAGCTTGCGCCTCCCGCCGAGA GCACCGTGCCCGCCACGCCCGATGGGAATCCGGAGGCGCCAGCACCCC CGTTTAGCGCCCCCGGACCTCCCACCCCCAAGGCATTGCCGCCCCCGCCC CCGTCCCGCCGCAACCTGCGCCCTCACCTCGTGCCCAAGGATGCTTGGC AGGGGGTCGCCGATGCCGTCGCCCCTGCCGCCTCGCGCCTCCTGCGCAC GCCCCTTGCGCACCTCTCCACCGAGCAGTGGGCCACGTTCGAAGCCGCC CTCGCCGGCCTCGAGGCTACGCTCCACCATGCCGCCCGCAGTGCAGAGG CGGTGCCCACACGCTGCGCTAGCCGAGCAAGGGAAGACGCCGAGCGCC AACTCCGTGAAGCCCGAAAGACGCGTGAGATCTTTGGCAAGGCCGCTG CCCTCTACGCAGCCGGCAAGGACCCCACTGCCACCATCGAGCGCATCCC CCCAGAAGTCCGCCTACACCTGCCAACCCCTGGCTCGGCTGAATGGCCC GCCAGGGCGGCCGCCGCCCGCAGGGTGATCCGCCGTGCAGTCGCGCGA GCGGACCGGTTGCGCAAGCGCATGGGCATCCTCGATAGCGACCGCGAC CTCCAACGCCTCTTCAACGCTAACCAGAAGAAGGCAGTTCGGCAGATCC | | TAACAA CCATGT ATGGTG AACCAC ACCTCTC TCGATCT TGTATTC TGTGATT GGACAT CAGAGT TCCTGC GAAGGG ATACACT CTGCCA ATCTCGT GGGTTG TAATAA ATCCAC ACCTTCA ACA (SEQ ID NO: 1302) | CCCGTCAA GGGTGCTC CACGAGAT CCCTGTCG CTAGCCGA CCGGTTTT ACCACCCC ACCCCGCC CGGACAAC CACGGACC CTGCTCCG CAGCAGG ACCCCACG (SEQ ID NO: 1179) | DWRMRPYCASTTAR WLAEQGRSPVAPRS PCPEPHAPPPAAALL YIPATPPAPTPRAPVA PPKLAPPAESTVPATP DGNPEAPAPPFSAPG PPTPKALPPPPPSRR NLRPHLVPKDAWQG VADAVAPAASRLLRT PLAHLSTEQWATFEA ALAGLEATLHHAARS AEAVPTRCASRARED AERQLREARKTREIFG KAAALYAAGKDPTAT IERIPPEVRLHLPTPGS AEWPARAAAAARRVI RRAVARADRLRKRM GILDSDRDLQRLFNA |

| | | | | | |
|---|---|---|---|---|---|
| | | TCGCCCCGTCCACCAAGGCGCCGCGGTGCCAGCTAGACCCAGCCGCCGT | | | NQKKAVRQILAPSTK |
| | | CGAGGAGGCCTACATCCAGACCCTCGCCAAGCCGCCGCCGATCGACCCC | | | APRCQLDPAAVEEAY |
| | | AGCCCCCCGTGGAAGAACTCCGTCCAGTGGCCCCGCCCGCCCACTGCCG | | | IQTLAKPPPIDPSPPW |
| | | CCGATGACGGAGGCAGCCCCTTCAGCGTCGCCGAGGTCCGGGCCCAGC | | | KNSVQWPRPPTAAD |
| | | TCCGCCGACTGCCCAACGGGTCCGCCCCAGGGATCGATGGCATACCGTA | | | DGGSPFSVAEVRAQL |
| | | CGAGGCCTACAAGCGTACGAAACTGGACGCCACGCTCGCCCATGTCTTC | | | RRLPNGSAPGIDGIPY |
| | | GAGGTCGTGCGGCTGAATGCGCGCCTGCCAGCTCGATGGGATGTGGCG | | | EAYKRTKLDATLAHV |
| | | CGCACGGTCCTGCTCTACAAGAAAGGCGACCCTAACGACACCGGCAACT | | | FEVVRLNARLPARW |
| | | GGCGACCGATAAGCCTCCAGGTCACCATCTATAAGATCTTCACGGCCGC | | | DVARTVLLYKKGDPN |
| | | CCTGTCGAAGCGGCTCATCTCCTGGGCTGGCAAGCACACAACACTTTCTCC | | | DTGNWRPISLQVTIY |
| | | GCATCGCAGAAGGGATTCCTACCGGCCGAAGGCTGCCACGAGCACGCG | | | KIFTAALSKRLISWAG |
| | | TTTGTCTTGCGAAGCGTGCTTGACGACGCCCGTCGGCACAAGCAGAACG | | | KHNTFSASQKGFLPA |
| | | TGTACCTTGCCTGGTACGATCTGCGCAACGCCTTCGGATCGGTGTCGCA | | | EGCHEHAFVLRSVLD |
| | | CGACCTCATCGCCTGGTGCGCTGCCATGTTGGGCCTGCCCCGCTACCTCC | | | DARRHKQNVYLAWY |
| | | GGGATGCCATCGGCGCAATCTATCGGCACTCAGCGCTCTTCGTCCAAGT | | | DLRNAFGSVSHDLIA |
| | | TGGGGATCAGGAGACCACCGGCGTCATTCCTATGCGCTGCGGCGTCAA | | | WCAAMLGLPRYLRD |
| | | GCAGGGCTGCCCTCTCAGCCCCCTCCTCTTCAACCTGTGCGTCGAGCCG | | | AIGAIYRHSALFVQV |
| | | GCCCTTCGCTGCCTACGCCGCACCACCGGGTACAAGTTCTACGGCACGT | | | GDQETTGVIPMRCG |
| | | CGATCACCGTCGAGGGCCAGGCCTACGCCGACGACCTGCTCACTGCCGC | | | VKQGCPLSPLLFNLC |
| | | GCCCTCCGCCTACCATGCGGCCCGGCAGGTGGCCACGATCGAGGAATG | | | VEPALRCLRRTTGYKF |
| | | GGCCAACTGGGCGGGAGTCTCCTTCGTCGTCCAAGCCCTCTCCCTGGAT | | | YGTSITVEGQAYADD |
| | | GCGCCGGCCGGCAAGTGTGCCGCCCTCGCGATCAACTTCGAAGGTGGT | | | LLTAAPSAYHAARQV |
| | | CTAATGCACTCTATCGACCCTGCCCTCAAGGTCCAAGGCGCAGCCATCCC | | | ATIEEWANWAGVSF |
| | | GGCCATGTCAAGAAACAACGTGTACCGCTACCTCGGAGTACATGTCGGT | | | VVQALSLDAPAGKCA |
| | | CTCACAGATGCGCTCGGCCAAGCGAACGAGCTCCTCGAGAAGGCCTCA | | | ALAINFEGGLMHSID |
| | | CGCGATGCACGCACGATCTGTGCCTCTGGCCTCGAACCCTGGCAGAAGG | | | PALKVQGAAIPAMSR |
| | | TGGTCGCAATCAAGACCTTCATCCTCTCCCGGCTCCCCTTCTTCTTCCACA | | | NNVYRYLGVHVGLT |
| | | ACGGGAAGATCCAGAGGGGCCGATGCCAGCAATTCGACCGCGAGCTTC | | | DALGQANELLEKASR |
| | | GAGAAAACCTGCGGGCCGCCCTCCGACTCCCCGTCTGCACCACGAACGC | | | DARTICASGLEPWQK |
| | | CTTCTTCCATTCCCGCGTGGCCTCAGGCGGCCTTGGCATCCTGCCCATCG | | | VVAIKTFILSRLPFFFH |
| | | CGGAAGAACAACAAGTCTACCTGGCAGCCCACGTGTTCAAGCTCCTGAC | | | NGKIQRGRCQQFDR |
| | | TTCGCCAGATCTGTCGATCCGCGCCATCGCCCGACACCAACTTGCCGAG | | | ELRENLRAALRLPVCT |
| | | GTCACCCACGCGCGACACACCACGCCAGTCCAGGACGGCGAAGCGTCA | | | TNAFFHSRVASGGLG |
| | | CCCTTCTTCGGATGGCTCATGCGGGGGCAGGAGGTCGCATCAACTACCC | | | 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 |
| | | | CGGCCACCTCCCTTGATCTCCCCCCCAAAACTTGGAATGTGCTTCTTAAA | | | | PLRPDIYCSRPGQCAI |
| | | | CGTTGTGTTGCTGCCTCAATCAAAGGCAGTTACACAGCGTGGAGAAGAT | | | | IEVAVSYEDAFNASM |
| | | | TCCGGCGCTCTACTCCATAACAACCATGTATGGTGAACCACACACCTCTCTC | | | | EGRAKQKTDKYAGLA |
| | | | GATCTTGTATTCTGTGATTGGACATCAGAGTTCCTGCGAAGGGATACAC | | | | ATVEEQLRLQTRHAA |
| | | | TCTGCCAATCTCGTGGGTTGTAATAAATCCACACCTTCAACA (SEQ ID | | | | FVVGFSGVVLPASVT |
| | | | NO: 1056) | | | | ATATSLDLPPKTWNV |
| | | | | | | | LLKRCVAASIKGSYTA |
| | | | | | | | WRRFRRSTP (SEQ ID |
| | | | | | | | NO: 1424) |
| NeS | Utopia | . | Acromy | GGTGCACAACGGATGCATCATACGTGTACCGGAGCATACGGGCTGTCA | GGTGCACA | TAAATTA | VCSVRGCRREDSRRF |
| L | -1_AEc | | rmex | CGGCGGCTGCATGCGCGATCTAGCTCGGAGATTTTATTTATTTATTTATT | ACGGATGC | TTTTGTC | YKFKFPLNFVKVPKTI |
| | | | echinati | AATTTATTTATTTATTCATCGAGTGTGAGTGTTCGCGTTTTGCCGAGAAG | ATCATACG | TTTGTCT | VIGSAFQKSSVSARS |
| | | | or | CGATTTTCGTTAAGTGATACGCGCCGCCGTTCATAGGTTAGGTGTGCAGT | 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 |
| | | | | AGCAAAACATCTGAAAAAAGTGCCACCCGGGAGACACATTAAATTATAAA | 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 | ERVSPHPPPKGIDILS |
| 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 |
| GGGGAGGCTGATAACACCWCCGAGCCTCCCCCAACGAACATCCTCCC | | TCTTTTT | TPPSLPQTNILPTIGEE |
| GACCATCGGGGAGGAAGGAACGTCACCGTGCGTGGCGGTCGTCACCAC | | TTCTTTT | GTSPCVAVVTTHPRP |
| CCATCCTTAGGCGGCAGGAGCACGTCTCCGCTCATCCTACCAAGGCCAA | | CTGTTGT | TGEDAPCEAPQPVSD |
| CGACACCGGAGCCTGAGCGGGGACAAGAGGAGCGGCGGCTAGAAGGC | | GGGGCC | HQRQSIGEPRRDTDT |
| GCGGCGCAGCCACCTACCACACCCGTCGTCGAGGGGGGACAACCAGTGG | | CTGACC | HLACDVATGNAHHL |
| GATGGCCAGTGGACGGTGAGCGTGAGGAGAAGAGCGAGGAGGCAACA | | GTCCGA | HGDDDHLWKPHNIH |
| ACTGAACGATACATCCCCCTCCAACTCCGAGTCCCCGCCAACCGCTGGA | | GTGTGA | GLHRWRGEADNTXE |
| CCATCGCGTTCGCCGCGCATAGCCCCACTATCTGCGCTGATAGCGGCGT | | ATGCCG | PPPNEHPPDHRGGR |
| CGACGAGCCGCCATGAGACCAGCTTAAATCTCAATTGCACGAACGGCAA | | CGAAAA | NVTVRGGRHHPSLG |
| TATTTGCATGGACCGAACTCCGCCCCGTAACATTTTGCCGGTARGGGCG | | ACAATA | GRSTSPLILPRPTTPEP |
| GAGCGCCGTCGCGAGACATCGCCACAGGATCGCGTGGAGGGAGACATC | | TTATGTT | ERGQEERRLEGAAQP |
| GGTTATGGTGCTGGAAAGGTAAGTGCCGAACACCCGAGTGCTCCCGTA | | TTATACG | PTTPVVEGDNQWD |

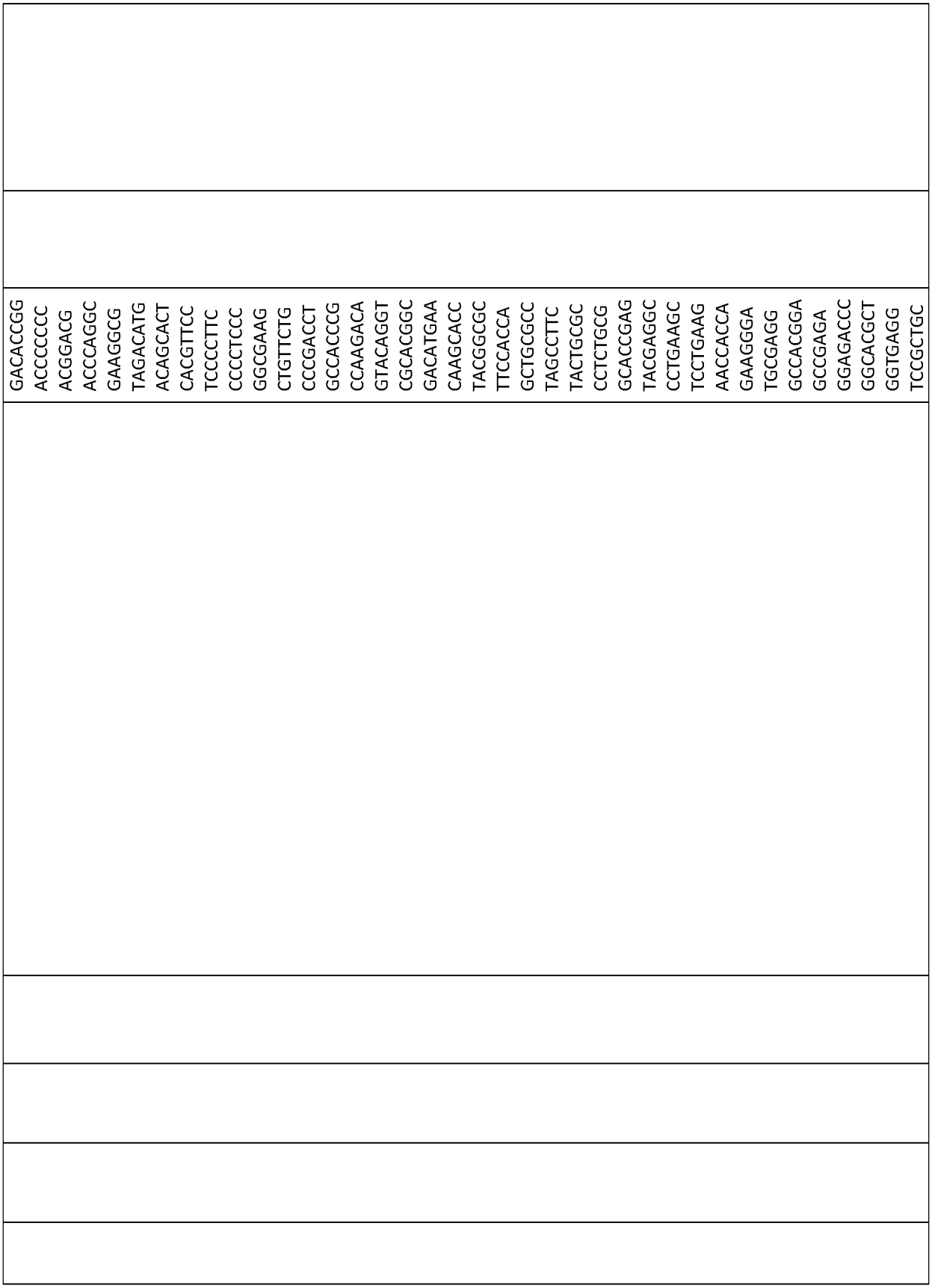
| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | AATGTCCGTGGTGTGATGTCTCGAGGGAGAGCAACCGCGTCATCCATCG | | | AGTGTG | | GQWTVSVRRRARRQ |
| | TGCCACCGCGAGCCAACCGTGGGGAGGGCGGTCGGCAGCATCATAGTC | | | CATGTG | | QLNDTSPSNSESPPT |
| | GGCGGCGTCCGGACGCTCCTGTCGGTCAGCCGTCGCGGGATCACCCGG | | | CGTGAT | | AGPSRSPRIAPLSALI |
| | CGCCTGCGACTGTCGCGAGGCAGCGTAGGCGTGAGCGGGTGGCCGCCC | | | ATATTTA | | AASTSRHETSLNLNC |
| | GCGACGCGCTGCTCGATCGGGCTAAGGACGTCGCTACGATTGCGGATC | | | TCTATTT | | TNGNICMDRTPPRNI |
| | TGGAGGCGTTCGCGGCTTCGGTCGCGGCGTTCTTCGGGGAGGATGCAT | | | TATTTTA | | LPVXAERRRETSPQD |
| | CGGCCACTGGTGCTGCAGCCCGCGCTCGCGATCGTTCGGTACGCTCACG | | | TTTATTT | | RVEGDIGYGAGKVSA |
| | GGAGGCGGGTGCGCGTCGGGGGGTGAAGGGAGGTGAGCGTCCGGAG | | | ATTATAA | | EHPSAPVNVRGVMS |
| | AGAGAGGGCGCCGGTAGGCCGGGTCAGCGCCGGCTGACCCCGGAGC | | | TTTATTG | | RGRATASSIVPPRAN |
| | GTCGGGAGAAGCACGCGGGGGACTGGGTGCGCGAGGCCAAACGCTTGC | | | 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 |
| | CATCGTCTCGGACCTGCGCTGTCGACCGCTCACTCCTTGCCCCGTGGC | | | CTTTTGT | | MYDGCTTRVRTAEG |
| | AGAAGATAGAAATGCTGGGGACCTTCATCCTATCCAGGCTTGACTTTCT | | | TTGTATT | | HTAPIPIRSGVRQGC |
| | GCTCCGGGGGGCCAGAGTGTTCAAGGGTCCCCTCACGGCCGTGGACCT | | | ACCCCA | | PLSPIIFNLAIDSVVRV |

| | | | | | |
|---|---|---|---|---|---|
| | | TAACATCCGWAGGCATGTTAAATCCTGGCTTAACCTCCCTCAGCGAGCA | GAGGGG | | AAEXNDGYSLHGNT |
| | | AGCGCGGAGGGGAGTCTACATGCCGCCCCGTTGGGGGGGATGTGGACTC | AATTGTC | | WSALAYADDIALLAQ |
| | | CTGCCGCTCTCTGACCTCGCCGACGTCCTCACGGTTGCCCACGCGTACCG | CCTCTG | | TPEGMERMLASVEA |
| | | TATGTTAACGGTGCGCGCATGGCGCCCGTGAGGGAGTTGGCGTGGGAATC | GGGAAA | | EAASVGLRFNPAKCA |
| | | GCTGAGGGGGAGTGGTTGGGCGCAGGATCGGCCACGCCCCTAGTTGCGA | AAAAAT | | TLHVGAGNGGRVLP |
| | | GGATATCGCCTCCTTCCTATCCGGCTCGCTGGATGGAAGGATGAGGGGC | GATTGG | | TSFQIQGETINPLAQ |
| | | GGAGGGGAGGCTTCGCTCTGGTCGAGTGCGCGGAACGCTGCGCTCAGA | AAAAAT | | GESYTHLGVPTGFSV |
| | | CAGTCCGAGAGGTTGTCCCTGCGTTGGCGGTGGGTCGAGGCCACGGAG | AAAGTG | | DQTPYAAVGDIVSDL |
| | | GAGATGACGTTGGAGTGTCGAGGGCCCAGGGGGGGCAGCGATTAAGAT | AGCTAA | | RAVDRSLLAPWQKIE |
| | | TCCGCCTGAAGCGCGCGGTCAGGTAGTGAATCGGTTGCGCTCAGCTGTA | (SEQ ID | | MLGTFILSRLDFLLRG |
| | | GCAGAGCACTACGCAAGTAGGTTGCTTAGCAAGCCTGATCAGGGTAAG | NO: | | ARVFKGPLTAVDLNI |
| | | GTCTTCGAGGTGTCGTCGCGGAGCCGAGTGAGCAATCACTTTATCCGCG | 1303) | | RRHVKSWLNLPQRA |
| | | GCGGCAGCTTCACTCGCTTCGCCGACTGGCGCTTTATCCATAAGGCCCG | | | SAEGVYMPPRWGG |
| | | GTTAGATGTTCTTCCTCTCAACGGCGCACGACGTTGGGAGGCCAACGAC | | | CGLLPLSDLADVLTVA |
| | | AAGCGCTGTCGGCGATGCGGTGAGGTATCGGAGACATTACCCCATGTG | | | HAYRMLTVRDGAVR |
| | | CTCTGTCACTGCGGCATCCACTCCGCCGCGATACAGCTGAGGCACGACG | | | ELAWESLRGVVGRRI |
| | | CTGTCCTGCACCGCCTTTGGAAGGCCACTCGCCTTCCAGGGGTAGTGCG | | | GHAPSCEDIASFLSGS |
| | | GGTTAACCAGCGGGTGGAGGGCGTCAGCGACGAATTGGGGGCGCTAC | | | LDGRMRGGGEASL |
| | | GACCTGATCTCGTGGTCAGGCACGAGCCCTCCAAAAGTGTCGTCATCTG | | | WSSARNAALRQSERL |
| | | CGACGTCACGGTGCCATTCGAAAACCGCTGGACCGCTTTCGAGGACGCC | | | SLRWRWVEATEEMT |
| | | AGGGCGAGGAAAATCGCCAAATACTCGCCTCTGGCAGAGGAGCTACAG | | | LECRGPRGAAIKIPPE |
| | | CGGCGAGGGTACCGTGTCGTCGTGACGGCCTTCGTCGGCGCCCTCG | | | ARGQVVNRLRSAVA |
| | | GCTCGTGGGATCCGAGGAATGAGGCGGTGTTGAAGACTGCTGCGGGTTG | | | EHYASRLLSKPDQGK |
| | | GCAACCAGTATGCAGCTATGATGCCGGCGCCTCATTGTCTCGGATACCAT | | | VFEVSSRSRVSNHFIR |
| | | TCGCTGGTCACGCGACATATATGTGGAGCATGTGTCCGGCACCCGCCAG | | | GGSFTRFADWRFIHK |
| | | TACCTGGCTCCTTCCCGTCCCTCTGGGGATCTCGCGACGCCGCCGAGAG | | | ARLDVLPLNGARRW |
| | | CGGTTCGTCGACGCTGGCTCGCCGAGGAGAAGCGCACACGGACGCG | | | EANDKRCRRCGEVSE |
| | | GCGCGTCGCGGTTCGGATAGTGTGAGTGTCGCGTAAATTATTTTGTCTT | | | TLPHVLCHCGIHSAAI |
| | | TGTCTTGGCCCCCCCCTTTTTAAACCAAGCAGGAGAGAGTGGCCCAATGC | | | QLRHDAVLHRLWKA |
| | | CCAACTATTATATTAACTATTTACTGTGATATTTATTATTTGACTGTTG | | | TRLPGVVRVNQRVE |
| | | GGCGGGCCCCTCTCTGCTGGTTTTATTTATATATATTTTTTACTCGCGTAC | | | GVSDELGALRPPDLVV |
| | | TTTTTGTACTACTCTATTTTTCTTTTTATTTTTAGCTATGCTATTTTTTATCTCT | | | RHEPSKSVVICDVTVP |
| | | TTCTTTGTCTCTATTTTCTTTCTTTTTTTCTTTCCTTTTCTTTTCTTTCTTTTAT | | | FENRWTAFEDARAR |
| | | TCTTCTTTTATTTATCTTTTTTTCTTTTCTGTTGTGGGGCCCTGACCGTCCG | | | KIAKYSPLAEELQRRG |
| | | AGTGTGAATGCCGCGAAAAACAATATTATGTTTTATACGAGTGTGCATG | | | YRVVVTAFVVGALGS |
| | | TGCGTGATATATTTATCTATTTTATTTTATTTATTATAATTTATTGCC | | | WDPRNEAVLRLLRV |
| | | GCGCGCGCTCCTCCGGGACTTTTATTCGTTGACAATACTGTGATATTTTT | | | GNQYAAMMRRLIVS |
| | | CTGCKCAGGCTGGGGGGGGCTTGCCCCCCAGCCCCTTAGTTTTAATTGCCT | | | DTIRWSRDIYVEHVS |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | ATGCGGGGGGGGGCTTTTGTCCCCCGCAAATGTATATATATATATATTTA GCGCGCGGCTTAGCCGCTTTTGTTTGTATTACCCCAGAGGGGAATTGTC CCTCTGGGGAAAAAAAAATGATTGGAAAAATAAAGTGAGCTAA (SEQ ID NO: 1057) | | | | GTRQYLAPSRPSGDL ATPPRAVRRRWLAEE RSAQDAARRGSDSV SVA (SEQ ID NO: 1425) |
| NeSL | Utopia-1_AMi | . | Alligator mississippiensis | TGCTGGAAAGACGGAGAACCGCTTCCTTTTTCCCTGCGCCTGGCCTGGT | TGCTGGAA | TGAACC | CHHAGLRPGTPNRT |
| | | | | ATTGCAGTACCTCCAGGATTAGCGCCAACTAGTCCGGCAGACTGTCGGA | AGACGGA | CCCCCTC | RRPDQTAPLPDPRG |
| | | | | ATACAGCAATAGAAAGWGAGCTGACTAGCAGCTTGCTTTCCTTCCTCCG | GAACCGCT | TGCACC | HPMPPNRRGSRSRP |
| | | | | GTGCAGCATGGGTTCTCGTCAGTCMTGACGGGCTAGGGAAGGCGGTG | TCCTTTTTC | AGATGG | EEPSRREPPXPRACQ |
| | | | | CTGCCAGTACGTCCGAAAGAAGTGCCGGTTGCGCAAGCGACCGCGCCAC | 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 |
| | | | | ACCCCGGCTGACGAGAGACCGCCTCCCGGCGGAGGACCCCGGAGCCTGA | 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 |
| | | | | GCCGCCGGGAGCCCCCCGKCCCCCGAGCGTGCCAGGGTCTCCGGGTGT | 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 |
| | GTCCTCATCCACAAGAAGGGAGACCCGACCGACCCGGGCAACTGGAGA | 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) |
| | CCACGCCGCCTCGGGAAGCGCATCGGCTGCGCCTGGACCTGGACCG | 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 | | |
| | | TTCTCTCTCAGCGTCGCGAACCCCCTCCCTCCCTTCCCCTCCCCTCCCC | 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) | TGAGCC GGTACG ACATCG TGCATC AACTAT GAGAAA GGGACT GAGAGA CTTTTTC CATTGG ACCATAT GAACTG GAACCA TAAACTC ACTGAA CATTAA ATCTCAC CAAATG AGGGTA AATCCAT CCTCATC | MTTKKVLGASTTLQT SSTKGKNSGCSKDPL RDAVPGRSWILRPAC RDITTRRNIPPAPQQ QQPPMESPPTLQLQ DALRRPSPTPAAAQV ADAGGALAALHTIKR GISVDWTSISPKXXQ RXTSASPDACPASET TQRDXRXLLDARPAG PLDPTRPHQDEPASD TADAAGTPLLQGNE DTIYLQYPLAADMLIC PICSPPQSFHLLGVVT RHLKRCHSKRVAFSC ALCSLPFETQKQCKM HQVACRKCLKGTTQS PAPAPSPPAARRPAA PEPQRRKXTSQAAVK KPAPVARPAERDAAI EKVPAASGNITQVLA |
| NeSL | Utopia-1_CMy | . | Chelonia mydas | CTCTTCTTATGAATACTTGCAACACCTGCACTGAAGATGGATTCTCCGGC TGCTATTTTTGAAAAACTGATGCTGCTTTGAAGGTGTATTCTGCTGCTGC TACCTTGGAAGGAAATTCTCTCTCTGCTCCTGAGACATCCCCAGCTGCAC CGTGTACCACCACCACCACTGCTGCTCCACAGAAGGTTTCTCGGAC AATGACTACAAAGAAGGTCCTAGGTGCCTCCACAACATTACAGACCAGC AGCACGAAGGGGGAAGAACAGTGGCTGCTCAAAGGACCCCCTCCGAGAT GCTGTTCCAGGAAGATCCTGGATTCTGAGGCCGGCCTGTCGGGACATCA CAACCAGAAGGAACATCCCCCCCGCCCCCCAGCAGCAGCCGCCAAT GGAGAGCCCCCCCCACTCTGCAGCTGCAAGATGCTCTCAGGCGACCATCT CCCACCCCCGCAGCTGCCCAGGTCGCTGACGCTGGTGGTGCTCTCGCTG CTCTACACACCATCAAGAGAGGAATCTCCGTAGACTGGACCAGCATCTC TCCAAAGASCMCCCAGAGGSTCACCAGCGCCTCGCCGGATGCCTGCCCA GCCTCAGAAACCACCCAGAGGGACCSCAGGWSCCTGCTGGACGCCCGC CCAGCCGGACCTCTCGACCCTACCCGCCCACACCAGGATGAACCAGCCA GCGATACCGCTGATGCTGCTGGAACCCCCCCTGCTGCAAGGTAATGAGG ACACCATCTACCTGCAGTATCCCCTCGCTGCGGACATGCTCATCTGCCCC ATCTGCTCTCCGCCCCAAAGCTTCCACCTCCTCGGTGTCGTCACCAGGCA CCTGAAGAGATGCCACAGCAAGCGGGTTGCCTTCAGCTGTGCCCTCTGC AGCCTGCCCTTCGAGACGCAGAAGCAATGCAAGATGCACCAAGTCGCCT GCAGGAAATGCCTCAAGGGAACMACACAGTCTCCTGCCCCGGCTCCCA GCCCTCCTGCTGCACGCCGGCCCGCTGCTCCTGAGCCTCAACGAAGAAA | | | |

| | | | | | |
|---|---|---|---|---|---|
| | | 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 |
| | | GAAATCGCTCCCAGGAGGAGTTCGAACCAGGAGAACGCCCCGCCTGCC | | 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 |
| | | TCACGGAGTGGTCGGTGAGCGGGGGAGCCATCAGCTCCATCCAGAAAG | | TTCTATG | SCLASRITEWSVSGG |
| | | GCTTCATGTCCTGCGAGGGCTGCTACGAACACAACTTCGTCCTCCAAACC | | CTTAAAC | AISSIQKGFMSCEGCY |
| | | ACCATCGAAACGGCCAGAAGGGCGCGGAGGCAGTGCGCGGTAGCGTG | | TCTGTAC | EHNFVLQTTIETARR |
| | | GCTCGACCTGGCTAACGCCTTTGGGTCCATGCCCCACCACCACATCTTTG | | CGTTTTT | ARRQCAVAWLDLAN |
| | | CCACGCTCCAGGAGTTTGGGATGCCAGAGAACTTCCTTCGTGTGATCCG | | TTTTATT | AFGSMPHHHIFATLQ |
| | | 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 |
| | | GGAGCTGGGAGTCCTGGTGCCGCAGATCAGGTCCAACGACACAACCAT | | 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 |
| | | GAAGGGCTACGAGGTGCAGATGGATGCCCTGATCGTCGGAGCCCTGGG | | GTPANHFLAGGGFT |
| | | CGCTTGGGACCCCTGCAATGAGCGTGTGCTGCGGACCTGTGGGATCGG | | RFADWRFIHRARLNC |
| | | TCGACGCTACGCACGGCTCATGCGGCGCCTCATGGTCTCGGACACCATC | | VPLNGAVRHGNRDK |
| | | CGATGGTCCAGGGACATCTACATCGAACACATCACCGGCCACCGACAGT | | RCRKCGYSNETLPHV |
| | | ACCAGGAGGTGTGAGCCGGTACGACATCGTGCATCAACTATGAGAAAG | | LCSCKPHSRAWQLR |
| | | GGACTGAGAGACTTTTTCCATTGGACCATATGAACTGGAACCATAAACT | | HNAIQNRLVKAIAPR |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CACTGAACATTAAATCTCACCAAATGAGGGTAAATCCATCCTCATCATCG<br>TATCCACTCATTATACTCCACACCTGAACATAGCCATTATATGAACAACA<br>TACCCCCATATCTCAATGTCTGTACTTTGACCCGTTAACCTTTTACCCCCA<br>ATCGGGGATATTGCAGATTATGTATTCCTTACGCCACCCGATCCTAAACC<br>GAATTTCGCACCCCTTGATAATCTGTACCTTATTCCCTGATAACCAGAAA<br>CTTCTATGCTTAAACTCTGTACCGTTTTTTTTTTTATTTCAACATCATCTTAAT<br>AAAATTATTAAA (SEQ ID NO: 1059) | | | | |
| NeSL | Utopia-1_CPB | . | Chrysemys picta bellii | TTTTTTCTGATGCTTGACTGCAAACACCCATCCAGAAGATGGAATCTCCT<br>GCAGCCATTTTTGAAAAAATTGATGCTGCTTTAAAGATATACTCCATTCT<br>CCTWKTTTGKAAGAAAACTCTTTTTCAGCTTCAGCTATTCTGTCATCGGC<br>TGCTGCTGTTCCTGCTTCCCAGAAAGCTCAGCMAAAACCTATCCTGAAG<br>ACCWCCCTTGGTGCCTCACGGAAGACCCGGASCACCTGCAAGAACCAA<br>AACATTAGGAGCTGGCTGAAGAAACCCCCCGTGGATACCTCWGCAGGG<br>AGACCTGGSTCCAGMAGGACAKCTCTTCGGGACCTCMCATCSAGGAGC<br>AAGAATATCTCAACAGCTCTTCAGGAGGGGACCCCCGGAGAACCCTG<br>CCCGCTTCCCAGAACCAGGATGCTGATCGCCGCCCCACCGGGAAGGAT<br>GCCACCGCAGGAGCCCCCCCCAATGACCCAGGACCAGGATGCTGATTGCT<br>GCCCCGCCGGGAAGGATGCCACCAGAGGAGCCCCCCCGATGACCCAGG<br>ACCAGGATGCTGATCGCTGCCCCGCTGCTCCAGAGAGGGATGCTCCGG<br>AAGGAACCACCTCCTCAACCCCAGACCCCAAAACTACTTACCACCCGGCT<br>GTCCGGAGGAGGGCCGCTCGAAGGGGAATGCACCTCAGAGCCCAAGA<br>TCTCGATGCCGCACGCTGCCCTTCCGGGCAAAGAGACAACGTGGCCAGT<br>GAGTCCAGCGCCCCCCCAAGAGCGACTTCACCTCCCCAAGCTTCTCTACC<br>AGACCCAGAGGAATCACCTGGCGAGTCTGCAGGCACAACAGAGATCCG<br>CCCCACGGAGGGTGAGGCAGGGGAAGAAGACCGCATCTACCTCCAGTA<br>CCCGCTCCCTACAGGTCTCCTCCTCTGCCCCTTCTGTCTCCCCGTCCATGG<br>AGTCCAGACCCTCGCGGCCCTCAGCAAACACGTTCGTAAGACCTACAAC<br>AAACGGATTGCTTTCCGGTGTAGCCGCTGCGATCTCCCCTTCGAGACCC<br>AAAAGAAATGTAAGTTTCATCAAGCCACGTGCAGGGGACCCCCCACGAC<br>CGCGAAAGTGAATCCCACTGACATCCTCCGGGTTCCAACCCTGACCCCC<br>ACCGATGATCTGGCTTCAGCACCCCAGCCAGCATCCCCAGAGTCACAGC<br>AGATAAGGGGGGGACCAACCGCCAACTGAGGGAAGCGTAACCCCCGCCT<br>CGAGGACTGACGATGCCACCAAAAGGACCAGTCCCGTCTCCAGAATCCC<br>CACGCTGGACCCTGCTGTGAGGGGGACCACCGCCACCTCTCAGGTCAAC<br>AACCTCACCAGACGCCTCAGCGACCTCATAAAAACCATCCGGCACAACA | TTTTTTCTG<br>ATGCTTGA<br>CTGCAAAC<br>ACCCATCC<br>AGAAGAT<br>GGAATCTC<br>CTGCAGCC<br>ATTTTTGA<br>AAAAATTG<br>ATGCTGCT<br>TTAAAGAT<br>ATACTCCA<br>TTCTCCTW<br>KTTTGKAA<br>GAAAACTC<br>TTTTTCAG<br>CTTCAGCT<br>ATTCTGTC<br>ATCGGCTG<br>CTGCTGTT<br>CCTGCTTC<br>CCAGAAAG<br>CTCAGCM<br>AAAACCTA<br>TCCTGAAG<br>ACCWCCCT<br>TGGTGCCT<br>CACGGAA | TGAGCC<br>AGAGTG<br>ACATCG<br>TTCTCCC<br>ACTACG<br>AGAAAG<br>GGACCA<br>AGTGAC<br>CTTCTCC<br>GTTGGA<br>TCATATG<br>AACTGG<br>AACCAT<br>AAACTC<br>CCTGAA<br>CATTAA<br>ATCTCAC<br>CAAATG<br>AGGGTC<br>AATCCAT<br>CCTCATC<br>ATCATAT<br>CCACTCA<br>TTATAM<br>TCCACAC<br>CCGAAC<br>ACAGCC<br>ACTCTAT | LGEVAVNCAIPGTDS<br>QLRPDVVVTDEAQK<br>KIILVDVTVSFENRTP<br>AFREARARKLEKYAPL<br>ADTLRAKGYEVQMD<br>ALIVGALGAWDPCN<br>ERVLRTCGIGRRYARL<br>MRRLMVSDTIRWSR<br>DIYIEHITGHRQYQEV<br>(SEQ ID NO: 1427)<br>MTQDQDADCCPAG<br>KDATRGAPPMTQDQ<br>DADRCPAAPERDAP<br>EGTTSSTPDPKTTYHP<br>AVRRRAARRGMHLR<br>AQDLDAARCPSGQR<br>DNVASESSAPPRATS<br>PPQASLPDPEESPGE<br>SAGTTEIRPTEGEAGE<br>EDRIYLQYPLPTGLLL<br>CPFCLPVHGVQTLAA<br>LSKHVRKTYNKRIAFR<br>CSRCDLPFETQKKCKF<br>HQATCRGPPTTAKV<br>NPTDILRVPTLTPTDD<br>LASAPQPASPESQQI<br>RGDQPPTEGSVTPAS<br>RTDDATKRTSPVSRIP<br>TLDPAVRGTTATSQV<br>NNLTRRLSDLIKTIRH<br>NTDTRRCSAPPQVTS<br>CRPAVGATSIVPQAA<br>RRDPANGGASRSPQI<br>PQPDPAPGRPNTSSK<br>VTQRASDRQKPHAP<br>PRTHQPDAARRRTRT<br>IPSASKHDRAPTKPST<br>GASRTPLPPGRSSAA |

| | | | | | |
|---|---|---|---|---|---|
| | CGGACACGAGACGCTGCAGCGCTCCCCCACAGGTAACCTCATGCCGCCC | GACCCGGA | GAACTT | SETPRAALPTTPGPPP |
| | TGCCGTAGGAGCAACTAGCATCGTCCCCCAGGCTGCACGGCGAGATCC | SCACCTGC | CATACCC | QDPPEHRSTVRGTTR |
| | AGCCAACGGAGGAGCCTCCCGTAGCCCCCAGATCCCACAGCCAGACCCC | AAGAACCA | TCATATC | PQTVPAAPEPAETTQ |
| | GCCCCCGGGAGACCCAACACCTCCTCCAAGGTTACCCAAAGAGCCTCTG | AAACATTA | TCAATGT | QEERRPRARVATPW |
| | ACCGCCAAAAACCCCATGCCCCACCGAGGACCCACCAGCCGGATGCCGC | 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 |
| | GGACGTCGGCCGCCGCTACGATCCGGCGGCTGCATCCCGTATTCAGAAA | GGAGCAA | TCTTAAA | DGIRYPLLKKRDPGCL |
| | CTATACAGGATGAACCGGACGAAAGCCATGAGGGAGATCCTCGACGGG | GAATATCT | CCAAAC | VLAAIFNKCKQFHRV |
| | ACCTCCTCCTACTGTGCCATCCAGCCCGAGAGGCTCTACTCCTACTTCAA | CAACAGCT | TTTGCAC | PRSWKKSMTVLIHKK |
| | GGATGTGTTTGATCACGAGGCCCAGACCAACTTGCGACGCCCAGAGTG | CTTCAGGA | CCTCGAT | GXRDDPGNWRPISL |
| | CCTTTCCCCGCTACCCCGGATCGACCTCACGGAGGACTTGGAGCGAGAT | GGGGGAC | AATCTGT | CSTIYKLYASCLAARIT |
| | TTTTCCCCGCAGGAGGTGCAGGCGAGGCTGTCGAGGACCAAAAACACC | CCCCGGAG | ATGTTAT | DWSVCGGAVSSVQK |
| | GCCCCTGGAAAAGATGGCATCCGCTACCCCCTGCTGAAGAAGCGAGAC | 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 |
| | | TCGCTGTACGTGGGAAACCCTGAAGCGTAAACCGGACCAGGGTAAAGCC | | | NRDKRCRKCGYPNET |
| | | TTTGAGTTGACCAGCAAGTGGGACGCCAGCAACCACTTCCTCGACGGG | | | LPHVLCSCKPHSRAW |
| | | GGCGGCTTCACCCGTTTCGCCGACTGGCGGTTCATCCACCGTGCCCGGC | | | QLRHNAIQNRLVKAI |
| | | TCAACTGCGTCCCGCTCAACGGAGCCGTCCGCCACGGGAACCGAGACA | | | APRLGEISVNCTIAGT |
| | | AGCGTTGCAGGAAGTGCCGGCTACCCCAACGAGACCCTGCCCCCACGTCCT | | | 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 |
| | | | | AATATCTCTGAAATTCAATCAATCAATTAATCATGGCGTCTCAGCGAGTG | TGTAGTGG | AAAAAT | ETVSAEPCDDESQRT |
| | | | | CACGTATTTGCCTACCCCTTCGTGGGACCATTCCGGTGCTCCGTATGCAT | GAAGTCTG | AAATAA | RYASPHKQARTLHDA |
| | | | | GGATGCGTCCGGGATGCATCCCACTAGGTCGCTGGGCGAATACGGCAC | TTTCTTTTT | ATATATC | EPRDASREHAPSCAE |
| | | | | ATACGCTGCGGCATATCAGCACATAACCCGGCGCCACCCACAAGTGGTT | 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 |
| | | | | TGAGAAGCGGCAACCAGCCGCCGATCTCGTCCATACCGCCCCGTTCGCC | 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 |
| | | | | AAGAAAGGGSGACAGAACTGACCTGTCAAATTGGAGGCCTCTTTCCATGG | 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 |
| | | GCCCAATCCTCTTTAACATCGCAATAGAGCGGTACTTCGCAAAATCAA | CTGCGGCA | NO: | LTGKKSSRRAVQTGF |
| | | AACCGTCAACGCGGGGGTACCTGCTCTATGGGAGCCGCATTAGCCCGCTG | TATCAGCA | 1307) | LVRGTPIPAMTEGDA |
| | | GCGTACGCCGATGACCTGGTGCTAATTGCGAGCTCCCCAGAGGAGATG | CATAACCC | | YEYLGIPLGLKKNQTP |
| | | AGGTCCTTGCTGCGTGCTGCGGACGACGCCGCAATAGAAGCCGGTCTG | GGCGCCAC | | RAAMEAIVGDIAKID |
| | | CACTTCAACCCCAAGAAGTGCGCGACCCTACACCTCACGGGGAAGAAAT | CCACAAGT | | DSLLAPWQKIDAART |
| | | CCTCGCGGAGGGCAGTGCAGACCGGCTTCCTCGTCCGTGGCACGCCAAT | GGTTATTA | | FVAPKLDFVLRSGATL |
| | | ACCGGCCATGACAGAGGGGGATGCCTACGAATACCTCGGCATCCCCCT | CATACCGT | | RAPLRHLDTVIKKHIK |
| | | GGGTTTAAAAAAAAAACCAAACACCCAGGGCAGCGATGGAAGCGATAGT | TGCCGGGT | | KWLYLPQRASAEVVY |
| | | TGGGGACATAGCCAAGATAGATGACTCGCTGCTCGCCCCGTGGCAAAA | CTGTGGCG | | TPLKKGGAGILPSSILA |
| | | GATCGACGCGGCCCGCACCTTCGTGGCACCGAAGCTTGACTTCGTGCTA | CTGA (SEQ | | DVLTIAQAHRMVSCP |
| | | CGAAGTGGCGCCACCTTGCGGGCCCCGCTGCGTCATCTGGATACAGTCA | ID NO: | | GEVVSRIASEGLREAV |
| | | TTAAAAAACACATTAAAAAAATGGCTGTATCTGCCGCAGAGGGCGAGCG | 1184) | | KRKINREPSGDEMAH |
| | | CGGAGGTAGTATACACCCCGCTGAAGAAAGGTGGAGCGGGCATACTAC | | | 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 |
| | | GCGAGAAAGAAGAAAAATACCGCCCCCTAGCTGAAGCCCTGCGCGCCA | | | RVLKTLGVSRFYAGL |
| | | TGGGATACCAGGTAAAACTGGAGGCATTCATTGTCGGAGCCTTGGGCTC | | | MRRLMVADTIRWSR |
| | | GTGGGACCCTAAAAACGAAAGGGTCCTTAAGACTTTGGGTGTCTCCAG | | | DIYVEHVSGIRQFTLP |
| | | GTTTTATGCTGGCCTGATGCGCAGACTGATGGTGGCCGACACCATCAGG | | | SGAPSN (SEQ ID |
| | | TGGTCCCGGGACATTTATGTGGAGCATGTATCCGGGATCAGGCAGTTCA | | | NO: 1429) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CCCTGCCAAGTGGAGCTCCCTCCAACTAAAAAATTAAAATGCCTTAAAA<br>ATAAATAAATATATCAAAATTTAAAAAAAAAAAAACGAGGAACAAATAAAC<br>ACAAATTCTGAAAGATTTATATAATTTAAAAWATAAATCGAAAATAAAT<br>GTTGAAAACAAAAAAAAAAAATAATAATAATAAAAAACACAATAACACT<br>CACCCGGCCTGCCCCAGAGGCAGGTAAACATTTACTGGCCATATGGCTT<br>TTTTTTTAA (SEQ ID NO: 1061) | | | | |
| NeSL | Utopia-1_Gav | . | Gavialis gangeticus | CGCTGGAAAGACGGAGAACCGCTTCTTTTTCCTGCGCCCGGCCTGGTAT<br>TGCACTTCCTCCAGGACCAGCGCCAACCTAGTCCGGCAGACTGCCGGAA<br>TAATAGCCTCAGAAAGAGAGCTGGCTAGCAGCCCTCTTTTCTTTCCTCCG<br>GTGCAGCGTGGGTTCTTGTCAGTCCTGATGGGCTAGGGAAGGCGGTGC<br>CGCCAGTACGTCCGAAAGAGCGCCGGTTGCGCGAGCGACCGCGCCGCT<br>CAGGCGAGTAGCCCAAGGGTCTTACGGTTCGCCGGACCCGATAACGCG<br>AAAGCCCCGACTCGGGCCAGTAGCCGAAGACCNTGGGCCTCCCTCCCCA<br>GGTCGGAGTAGGCGAACGCCCGTGCTCGGAGGACGGAACGTGGACAA<br>AACACCCCCAGGTCCCAATGACGCCCTGATCCACTGACAAGAACGCTCG<br>AGGCACNCCAGGAGACCCCCAGCTAGGGCAGACCGCCGACCACGGGTC<br>GCGGAGGACCCTCCCAGGAGGGTGGACCAGCGAACCCGAGTCGGCGA<br>CGAACCCCGACGCACCCCCCCCGCGATGTCGGGACCCCGACAGGCGGC<br>GGCGGACCCCCCGGCCATCGACCGACCCCCGGAGGCAGAGAGACTCTCA<br>GAGCCCGGAACCCCGGCTGACGAGAGCCGCCTCCCGGCGGAGGACCCC<br>GGACCCCGAGGACGCCCCCCGGACGACGGCGGAGCACCCCGAGCGAC<br>GGCGGACTCCTCCGGACCCCCGCGGNCCCTCGGCGACGACGGCGGGCC<br>CCGAGCGGCGACGNCCCCCGGACCCCGGCGGTCCCGAGGACGACCCCC<br>CCGAGGGCCTCCCCACNCTGGTGGAGGAGCCCCGAACCCCCCCGACAC<br>CGGACCCCCCCCGACGGACGACCCAGGCGAGGGTGCAGACGCGGCAGC<br>GCTCACGTTCCTCCCCTTCCCCCTCCCTGCGAAGCTGCTGTGCCCGACCT<br>GCCACCCGCCAAGGCAGTACAGGTCGCACAACGACATGAACAAACACC<br>TACGGCGCTTCCACCAGCTGCGCCTAGCGTTCTACTGCTCCCTCTGCGGC<br>ACCGAGTACGAGGCCCTGAAGCTCCTGAAGAACCACCACACAAGGTATGC<br>CAGGGCCCCGGGGCCGAGAGGAGACCCGGCACGCTGGTGAGGTCCGC<br>CGCCCCACGACCCGGGCCGCGGCGCGAAGGCCCGCCAGACCGGCCGCC<br>CCGCCGGCGACCCCACCGGACCAGACCTCCGGGGACCGCCCGACGGAG<br>AGACCCGCCCCGGCGACGCCACCACGCAGGTCTGCACCCAGGGACCCC<br>CGACCGGACGTGACGCCCCGACCGGACGGCCCCCCTCCCGGACCCCCG<br>GGGCCGCCCGACGCCCCCGACCCGCCGAGGATCCCGGAGCCGCCCGGN<br>GAGCCCGAGCCGCCGGGAGCCCTCCAGCTCCCGAGCGTGCCGGGGTCT<br>CCGGGTGCGGAGAGGAGCTCTCCCGGGCCACCGCCTTCGAGGCCTTCG | CGCTGGAA<br>AGACGGA<br>GAACCGCT<br>TCTTTTCC<br>TGCGCCCG<br>GCCTGGTA<br>TTGCACTT<br>CCTCCAGG<br>ACCAGCGC<br>CAACCTAG<br>TCCGGCAG<br>ACTGCCGG<br>AATAATAG<br>CCTCAGAA<br>AGAGAGCT<br>GGCTAGCA<br>GCCCTCTT<br>TTCTTTCCT<br>CCGGTGCA<br>GCGTGGG<br>TTCTTGTC<br>AGTCCTGA<br>TGGGCTAG<br>GGAAGGC<br>GGTGCCGC<br>CAGTACGT<br>CCGAAAGA<br>GCGCCGGT<br>TGCGCGAG<br>CGACCGCG<br>CCGCTCAG<br>GCGAGTA | TGAACC<br>GCCCCC<br>CCTCCGC<br>GCCAGA<br>CGGACC<br>TTCACTT<br>CACTCC<br>GAGAGG<br>ATTCTTC<br>GACCAC<br>GGACGA<br>CCCCGCT<br>CCACCC<br>GAAGAG<br>GACCCC<br>CGCGAT<br>GAGACT<br>CTATAC<br>GGACTG<br>AGGCAC<br>TTCCTTC<br>GAACCA<br>CTTCCTC<br>CACCATT<br>GCGGAC<br>CATTGTA<br>ACGGGT<br>TTGTGT<br>GTATCTA<br>TCTCCTT<br>TCTCTCT<br>CAGCGT | MSGPRQAAADPRPS<br>TDPRRQRDSQSPEPR<br>LTRAASRRRTPDPED<br>APRTTAEHPERRRTP<br>PDPRGPSATTAGPER<br>RRPPDPGGPEDDPPE<br>GLPTLVEEPRTPPTPD<br>PPDGRPRRGCRRGS<br>AHVPPLPPPCEAAVP<br>DLPPAKAVQVAQRH<br>EQTPTALPPAAPSVLL<br>LPLRHRVRGPEAPEE<br>PPQGMPGPRGREET<br>RHAGEVRRPTTRAAA<br>RRPARPAAPPATPPD<br>QTSGDRPTERPAPAT<br>PPRRSAPRDPRPDVT<br>PRPDGPPPGPPGPP<br>DAPDPPRIPEPPGEP<br>EPPGALQLPSVPGSP<br>GAETSAQQRMPTPR<br>QALWLEELSRATAFE<br>AFEASVARLTEELSAA<br>ARPGQPRRGADNGP<br>TTRRDHRPQPQRRP<br>RRQRYDPAAASRIQK<br>LYRANRPKAAREILEG<br>PSAFCQVPRETLFNYF<br>SRVFNPPAEAAAPRP<br>ATVEALTPVPPAEGF<br>EEAFTPREVEARLKRT<br>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 |
| | GGCCGTCAGCCGGAGCCAGAAGGGCTTCATGTCGACGGAGAGGGCTGCTA | 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 |
| | CATGGAACCGCTCCTTCGAGCCGTGGCGGGCGGCCCCGGCGGCTCGA | 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 | CGCG (SEQ ID NO: 1185) | | | FLHRARLNCLPLNGA |
| | | | GCCTGGACCTGGACCGAGGAGTGCCGGGAGCTGGGAGTCTCCCTGCAA | | | | VRFGHRDKRCRRCG |
| | | | CCAGCCCCGCACGCCGACCGCGTCACCGTGACGCCCCGCACGAGGACCT | | | | YAAETLPHVLCSCKP |
| | | | TCCTGGAGAGGTTCCTGAAGGACGCCGTCCGAAACAAGTACGCCGGCG | | | | HARAWQLRHNAVQ |
| | | | ACCTGAGGGCCAAACCCGACCAGGGCAAGGTCTTCGACGTCACCTCGA | | | | DRLVRAIPAAAGEISV |
| | | | AGTGGGACGCTAGCAACCACTTCATGCCCAGCGGGAGCTTCACGCGCTT | | | | NRTVPGCESQMRPD |
| | | | CGCGGACTGGCGCTTCCTCCACCGCGCCCGCCTCAACTGCCTGCCTCTGA | | | | IVITNEEAKKVVIVDV |
| | | | ACGGGGCCGTGCGCTTCGGCCACCGGGACAAGAGGTGCCGACGGTGC | | | | TIPFENRRQAFTDAR |
| | | | GGCTACGCGGCAGAGACCCTCCCCCACGTGCTGTGCAGCTGCAAGCCG | | | | ARKREKYAPLADTLR |
| | | | CACGCCAGAGCCTGGCAGCTCCGCCACAACGCTGTCCAGGACCGCCTG | | | | GRGYDVTVDALIVGT |
| | | | GTGAGGGCCATCCCGGCCGCGGCGGGGGAGATCTCCGTGAACCGCACC | | | | LGAWDPSNESVLRA |
| | | | GTCCCGGGCTGCGAGAGCCAGATGCGACCCGACATAGTCATCACCAAC | | | | CRVSRRYAKLMRCL |
| | | | GAAGAGGCCAAGAAGGTCGTGATCGTGGACGTCACCATCCCCTTCGAG | | | | MVSDTIRWSRDIYVE |
| | | | AACCGGCGCCAAGCCTTCACCGACGCCCGGGCTCGCAAGCGGGGAGAAG | | | | HITGHRQYSDPTRRA |
| | | | TACGCCCCGCTGGCCGACACCCTGAGGGGCCGCGGCTACGACGTGACG | | | | AAGPDPEGTA (SEQ |
| | | | GTCGACGCGCTCATCGTGGGAACGCTCGGAGCCTGGGACCCCAGCAAC | | | | ID NO: 1430) |
| | | | GAGAGCGTCCTGCGTGCCTGCCGCGTCTCCCGCCGCTACGCCAAGCTGA | | | | |
| | | | TGCGCTGCCTCATGGTGTCCGACACCATCCGTTGGTCCCGCGACATCTAC | | | | |
| | | | GTGGAACACATCACGGGCCACCGCCAGTACTCCGACCCCACCAGACGA | | | | |
| | | | GCCGCCGCCGGACCGGACCCGGAGGGACCGCCTGAACCGCCCCCCCT | | | | |
| | | | CCGCGCCAGACGGACCTTCACTTCACTCCGAGAGGATTCTTCGACCACG | | | | |
| | | | GACGACCCCGCTCCACCCGAAGAGGACCCCCGCGATGAGACTCTATACG | | | | |
| | | | GACTGAGGCACTTCCTTCGAACCACTTCCTCCACCATTGCGGACCATTGT | | | | |
| | | | AACGGGTTTGTGTGTATCTATCTCCTTTCTCTCTCAGCGTCGCGAACCCC | | | | |
| | | | CTCCCCCACCCCCCACCCCCGGGCTTAGTTGGCTAACATTGTATCTCCTG | | | | |
| | | | TAACCTAGTCGCGTTCCCCTCCTCACCCCCATCCCTCTATTGTTAGTCCCT | | | | |
| | | | CGCTCGGGCGATCTGTATTTCCCTATCGGCTTTGTCATCTTTTTTCTGGAT | | | | |
| | | | TCCCGATCCTAAACATTTACTAATAAAAGTCAATCTGTTCTTT (SEQ ID NO: 1062) | ATCTACTA TC (SEQ ID NO: 1186) | TGAATA GCATTTA TATTGTG TTCCAAA CAACAT ACTCATT ATTATAT CTAAAC ATTTTTT | | MSCPREGSDHLGPD |
| NeSL | Utopia-1_LV | AGCV013581 06 | ATCTACTATCATGTCTTGTCCAAGAGAGGGAAGCGATCACCTCGGTCCT | | | | PETPALHQGSDIRVT |
| | | Lytechinus variegatus | GATCCTGAGACACCCGCCCTCCATCAGGGTTCTGACATCCGGGTTACCA | | | | SSRLRGSRGKSSRQP |
| | | | GTTCTCGCCTTCGAGGTTCCCGCGGAAAGAGTTCTCGCCAACCAAGCTC | | | | SSRHQVPASEASATA |
| | | | CCGACACCAAGTTCCTGCCAGCGAGGCTTCCGCCACCGCCCAGCAGACT | | | | QQTAANECQVCGSS |
| | | | GCCGCCAACGAGTGCCAGGTGTGTGGATCTTCCTTCGCCACCTCCAGTG | | | | FATSSGLRRHMARLH |
| | | | GACTCCGCCGCCACATGGCCAGGCTTCATCGAGCTGCCTCTGCGGATCC | | | | RAASADPEGAAPASI |
| | | | TGAGGGTGCTGCGCCGGCTTCCATCACAGAGATTTTCGACTACCCCTTG | | | | TEIFDYPLPSRWKCSA |
| | | | CCTTCCCGGTGGAAATGCTCGGCATGCTCGGAGAACTTTTTCAACCAGC | | | | CSENFFNQQTLKRH |
| | | | 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 |
| | | ACTCCCAAGGCTCACAAGTCTACAGGAGAAGCTCTAGGAAGCTGTCGTCA | | 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 |
| | | GGTCGTGGCACGGGGACAACGAGCCGTGTTCTGGGACGTCTAGGTATCT | | ERLVKAIHYGDIYQD |
| | | CTCGCCGATACAGAACACTGATGAGGCGCCTGTGCTGCATAGATGCGAT | | QHVPGDPNPRERPDI |
| | | CAAAGGCAGCCGGGATATCTATATCGAACATGTTACCGGTCACCGGCAG | | TVVEGNKVTIVDITIP |
| | | TATTGAATAGCATTTATATTGTGTTCCAAACAACATACTCATTATTATATC | | FDNGPDALSTAANA |
| | | TAAACATTTTTTTTTCTGTTCCTGACAATCTACGTAAAGTCTGCTAACCAA | | KVMKYDTLRQELASR |
| | | CTGGCATGATGAAATAAGATAAAATCCCCTTACACATTAATTTCTTGTCA | | GMDVEVHAFVIGSL |
| | | CATCATAATGCTTTGTCAAAGCAATGTCCTACATAATATCTCGATGTCAC | | GSWHGDNERVLGRL |
| | | CCCAATTAATTTTACATCCTTCGGTAACCTTTATACCGTTGGATCAACATA | | GISRRYRTLMRRLCCI |
| | | TATGATTTGTAAAACTGTTATTTCTGAGTTTTTTCTATGCTAATAAA (SEQ ID NO: 1063) | | DAIKGSRDIYIEHVTG HRQY (SEQ ID NO: 1431) |

| NeSL | Utopia - 1_NVit | . | Nasonia vitripennis | CCATTCCTCGTACGGGTTTTCGTGCCGGCATAGCCGGGTGGGAGACTC | CCATTCCT | TAGTGG | SGXTGREVKCITVNV |
|---|---|---|---|---|---|---|---|
| | | | | GCGCGGGGGAGGTCATATCTCACCACCATCCTCGAGCTTTTTGCTGCAC | TCGTACGG | GGCCAT | LMEQQPHTKAIREG |
| | | | | CTGATCCGGGMTAACTGGCCGAGAAGTTAAATGTATTACTGTTAATGTA | GTTTTCGT | AACACC | DFIVILLPQSDDETLC |
| | | | | CTCATGGAGCAACAACCACACACCAAAGCGATCCGGGAGGGGGGACTTC | GCCGGCAT | TAGGCC | CPLCVGRGRYSGKTR |
| | | | | ATCGTGATACTCCTCCCGCAAAGTGACGACGAGACACTTTGCTGTCCACT | AGCCGGGT | CCACAG | VECLNRHVKEVHPDL |
| | | | | CTGCGTGGGACGCGGACGGTACTCCGGCAAGACGCGGGTTGAGTGTCT | GGGAGAC | TGTGGC | TTTFRCWGCGFAAP |
| | | | | CAATAGACATGTCAAAGAGGTGCACCCAGACCTAACCACCACCTTTAGG | TCGCGCGG | GATGTC | GDKKYPRKIVTQHCA |
| | | | | TGCTGGGGGTGCGGGTTCGCGGCTCCCGGGGATAAGAAATATCCCCGG | GGGAGGT | CATGTG | TCVPEVSSAPSGRVD |
| | | | | AAGATCGTGACCCAGCACTGCGCCACCTGCGTCCCCGAGGTTTCGTCCG | CATATCTC | TGTCGT | GERRVNTRRRLGIAA |
| | | | | CGCCCTCGGGCAGAGTGGACGGAGAGAGGAGAGTTAACACCCGCCGTC | ACCACCAT | CCTTACT | ATEASPVRRTRRNGL |
| | | | | GGTTGGGTATTGCCGCGGCGGCAACCGAGGCCAGTCCCGTGAGACGTACTC | CCTCGAGC | TATTTAT | ASPPVEQNISQSAAP |
| | | | | GTCGTAACGGCCTGGCCTCACCTCCAGTCGAGCAGAACATCAGCCAAAG | TTTTTGCT | TTATTTG | PEPARVPQHPEIVAL |
| | | | | CGCGGCTCCTCCGGAACCCGCAAGGGTGCCACAACATCCGGAGATAGT | GCACCTGA | TCCTGTC | GESADDEVFRSPVNS |
| | | | | TGCGCTGGGAGAGTCAGCGGATGACGAAGTATTTCGATCTCCGGTGAA | (SEQ ID | GAGCTG | PPRDWRAAAPQQA |
| | | | | CTCTCCCCCAAGAGACTGGAGGGCTGCGGCGCCCCAACAGGCGGCTTC | NO: 1187) | TTTAATC | ASSSPXAVPGITAATP |
| | | | | CTCCTCTCCTCSGGCGGTGCCCGGTATCACTGCTGCAACCCCGTCCAACA | | TATTCAT | SNTTRTGNGSAXSILA |
| | | | | CCACCAGGACAGGGAATGGCTCGGCANACTCCATCCTAGCCGAGCACC | | ATGTAT | EHPIPAPPPTNTTEA |
| | | | | CGATACCTGCACCTCCGCCAACCAACACCACGGAGGCGAATGGTCGTGC | | GTGTTG | NGRADIPRSGRAPPP |
| | | | | CGACATACCCCGTTCTGGGCGGGCTCCGCCGCCCGGTCNACAAGCCGCC | | TGTGTG | GXQAARRRAPTTEQ |
| | | | | CGTAGAAGGGCTCCCACGACAGAACAACGTCGCATTGTGGGACTCCTC | | TCGTCCC | RRIVGLLEAATGREQL |
| | | | | GAGGCGGCCACTGGTCGGGAGCAGCTTGAGGAGGCAACGACTCAAGC | | TCGTAG | EEATTQAMLFLARLT |
| | | | | TATGCTATTCCTGGCCAGATTGACCGGGCGGCGGCCTGAACCCCGGAAC | | CAGTTTT | GRRPEPRNAIRPGXR |
| | | | | GCCATTCGGCCGGGAMCGCGGCAACGTCATCCAGCTCAGGGGGATGTC | | ATTCCGT | QRHPAQGDVQAQA |
| | | | | CAAGCGCAGGCGCCTGATCGGATTCASGAGGCCAAGAAACTCCAGCGT | | CCAACC | PDRIXEAKKLQRLYRT |
| | | | | CTCTACCGAACATCCAAGAAGCGGGCAGTMCAGAAGATTTTGGCAGGT | | AGAGGT | SKKRAVQKILAGPXM |
| | | | | CCTNATATGAACTGCCAAATCGACAAGAATACCATCACAGCCCACTTTGT | | CGACCA | NCQIDKNTITAHFVEL |
| | | | | GGAGCTGGCTGCCCGCAGAGACGGTGGAGAGGATTGGCCCGATGTCTT | | ATATTAA | AARRDGGEDWPDVF |
| | | | | CGATCGGGAAGAGCCCACGGCGGCGAGCGGCGAGGCCCTTTGTACCCC | | AATAAG | DREEPTAASGEALCT |
| | | | | TATCACAAGGGAGGAAGTGTTTCGACGGCTCAAGGGGAGGAACAACAC | | CATGGC | PITREEVFRRLKGRNN |
| | | | | CTCACCTGGGCCGGACGGAATCACCTACAGGGACCTCGCCAAGGCCSAT | | TTGAAG | TSPGPDGITYRDLAK |
| | | | | CCAGGGGCCCATGTGCTCGCCGCCCTCTATAACASCATCTGGCGCATTG | | CAGGCC | AXPGAHVLAALYNXI |
| | | | | AGGCGACGCCTGCCTTGTGGGGGGTGTCCAACACMACCCTKATATACA | | AAGCGC | WRIEATPALWGVSN |
| | | | | AGAAAGGCGACGCCATGGACATCTCCAACTGGAGACCGATCTCCTTGG | | CGTGTT | TTLIYKKGDAMDISN |
| | | | | GAGATACGGTCCCCAAGCTATTTGCCGCGATCCTGGCCGATAGAATCAA | | CTAACCC | WRPISLGDTVPKLFA |
| | | | | AAGGTGGGCCGTGGCCAACGGGCGCTACAGTGCCTCCCAGAAGGGCTT | | CGTTTTA | AILADRIKRWAVANG |
| | | | | CTTGGAGTTCGAGGGATGCTATGAACATAACTTCGTGCTCCAGGAGGCC | | GGGGAA | RYSASQKGFLEFEGC |
| | | | | ATTCGGGAGGCAAAAGGCGGGAGAAAGGAGTTGGTGGTGGCCTGGCT | | GTTACTT | YEHNFVLQEAIREAK |
| | | | | GGACTTAGCCAGTGCCTTTACATCCGTGCCCCACTCCTCTATCCTGCAGG | | AACCTA | 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 |
| | | AAGGTCTTTGAGGTTACCCGGAGAGGAGCGGACAATCCAACCACTTCCTCC | | | APGTRVSPQTKHLVT |
| | | GGAGTGGCAGCTTCACTCGCTTCTGCGACTGGCGCTTTATCCACCGCGC | | | TSLRRCLNRHYAESLL |
| | | CCGTCTCGACGTTCTCCCCCTCAATGCTGCCAAACGGTGGCAACGAGGG | | | AKKDQGKVFEVTRRS |
| | | ATGGACAAGAGATGTCGCAGATGCGGGTCAGACCTGGAGACTCTCCCG | | | GQSNHFLRSGSFTRF |
| | | CATGTGCTGAGTCACTGCGGCCCGCACTCGGCGGCGAGACAAAAGAGG | | | CDWRFIHRARLDVLP |
| | | CACAACAACATCCAGGACAGGTTGGTGAAGGCAGCCTCGAGATGTCCG | | | LNAAKRWQRGMDK |
| | | GGGACCATTAGTGTCAACCAGACCGTTGTGGGCGTCCGGCGCCCCGAC | | | RCRRCGSDLETLPHV |
| | | GCGGCATTAAGACCKGACATTGTCGTTCGCGACGATGTCAACCGCCGCG | | | LSHCGPHSAARQKR |
| | | TCACCATCGTCGATGTGGCCGTGCCGTTCGAAAACCGCCTCGAGGCGTT | | | HNNIQDRLVKAASRC |
| | | CGACGGCGTTCGKGAGGCCAAGATTGCCAAATATACACCTTTGGCCAGG | | | PGTISVNQTVVGVRG |
| | | CAACTCACGGACAGCGGTTACACCGTAACGGTGGAGGCCTTTGTAGTG | | | PDAALRPDIVVRDDV |
| | | GGCGCCCTCGGCGCGTGGGACCCGCGAAATGAACGGGTCCTGAGCCTC | | | NRRVTIVDVAVPFEN |
| | | CTCAGCATTAGTCGTTATTATGCGATACTGATGAGGAGGCTCATGGTTA | | | RLEAFDGVREAKIAKY |
| | | GTGACACAATTCGCTGGTCCCGGGACATCTACGTGGAACATGTGTCGGG | | | TPLARQLTDSGYTVT |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CATTCGGCAGTACCGGGAATAGTGGGGCCATAACACCTAGGCCCCACA<br>GTGTGGCGATGTCCATGTGTGTCGTCCTTACTTATTTATTTATTTGTCCTG<br>TCGAGCTGTTTAATCTATTCATATGTATGTTGTGTGTGTCGTCCCTCGT<br>AGCAGTTTTATTCCGTCCAACCAGAGGTCGACCAATATTAAAATAAGCAT<br>GGCTTGAAGCAGGCCAAGCGCCGTGTTCTAACCCCGTTTTAGGGGAAGT<br>TACTTAACCTAAAAATACAACTTTTCC (SEQ ID NO: 1064) | | | | VEAFVVGALGAWDP<br>RNERVLSLLSISRYYAI<br>LMRRLMVSDTIRWS<br>RDIYVEHVSGIRQYRE<br>(SEQ ID NO: 1432) |
| NeSL | Utopia-1_PCa | . | Phytophthora capsici | GCCCGCCGGTGGAGTAGCCATGTTGGCCACCACCGCCCAAGTCTCCGCC<br>GCAGCTGCGACTGCTGCTCATCGAGTCGCAGTAGTCGCAGCTGCAC<br>AAGCACCACCTCCTGCCGGACGCGCCGCCGTGGAGCACCACGCGCGCG<br>CTGAGCCGTACCAAGACCAAGGNTTCCAGGCTCGCGCGCGCGTGGMG<br>CTCCCAGCCAGCGACGCGTTCAGCAGCAGGGTCGGCTGCGGCAGCTCG<br>ACGCTGCACGGGGACGGCAGCGACGGCAGGCACCAAGACCACGACGA<br>CCAAGCCGTCGCTGCCGTCCCCATGGACGTCGACCAAGGTGCTCGGTGG<br>CCGACAGCGGATCAMCACCCGCTGTCGCCGCCACCGGAGCTCCGAGTC<br>GGCGGCAAGCGCCGCCGCCTGAACGACGGCGACGACGAAGACATCCGC<br>GAGCTGGCCGAGCTTCTGCTGCCCGACGAGGAGGAGGCCGACGACCAC<br>AAACCAGCGCCCAGGTTACCCGCGACCAGCGCTCATCCGGCCTCCGTCC<br>TCGCTGTGTACGCGCACAACGCGCAGCGCTTCAACTGCACGTTGTGCGT<br>GTACACGGCTGCCAGCTTCGCTGCTCTTACGCGACACACGGGACTCTCGG<br>CACCGGCGCGTGACCTTCCTGGACAGGTTCTCGGCGGGTTGCGCGTGC<br>GGCAAACCTTTTGCCTCGAGGCTGGCCGCAGCAAGACACGCACAAACG<br>TGCGCCAGCCTCAGCACCACACTGGTCGCGGTTTCGACGACAGGAGGA<br>GCATCAAGCCACACTGTCGTCGGAGCCAACACCACCGTCGCCACGGCG<br>GTCACCGCCGAACCCCCCCCTGCTCCACCATCAAGCCTCGGAACTCACTGT<br>GCCCCCCCCCACGTGTGAGTTCCCCCGACGTCGATGTGCAGCTGCACAG<br>TCCGCCACAGGAAGATCAGCACGAGGACGCCACCCANCACCCGGAGAG<br>CACGCAACACCAACCTCCTGAGGCAACCCGCTGGGGTTCGCCGCTCGCG<br>CCCACGCTCGTTGCCTCCAGGATTGCTCAGCGACTCGGCGAGCTGGAAC<br>CTCCACGCTGGGGCCCACCATTACCCCGCGCGATGGTTGTTTCGCGAAT<br>TACAGCACGGCTTGAAGCTACACCTGCTCCTCGATGGGATCCGCCACTC<br>CCACGCCGCGTCATCGCCTCCCGCATTGCCGACCGGCTTGTACCTGCGA<br>CGGCACCGTGTCGCTCGGCCCTGAACGCTGCATTCCCTTCTCCATCGCGA<br>GACACGGTGACGGAATCGTTCACCCAGGAGGACAGGCAACTGGAACCC<br>CTGACGCGGCACGTGGACGAAGAGACGAAGGACAGCGAGCTGCCGGG<br>CAGGGCGCCGACTGTACTCGACGAGGAGTCCAAGGACAATGACGCCAC<br>TGCCGGTGAGTGGCTCCTCCGGTTCGACGGAGCTTGTCGAGCGAACCCA<br>GGCCCAGGCGGAGCAGGTGCCGTGCTATTCAACCCTAGCGGTGCAGCG<br>TCATGGACCTGCTCGCACTTTATGCCGGGCGCCACGGAGACAAACAACA | GCCCGCCG<br>GTGGAGTA<br>GCCATGTT<br>GGCCACCA<br>CCGCCCAA<br>GTCTCCGC<br>CGCAGCTG<br>CGACTGCT<br>GCTGCTCA<br>TCGAGTCG<br>CAGTAGTC<br>GCAGCTGC<br>ACAAGCAC<br>CACCTCCT<br>GCCGGAC<br>GCGCCGCC<br>GTGGAGC<br>ACCACGCG<br>CGCGCTGA<br>GCCGTACC<br>AAGACCAA<br>GGNTTCCA<br>GGCTCGCG<br>CGCGCGTG<br>GMGCTCC<br>CAGCCAGC<br>GACGCGTT<br>CAGCAGCA<br>GGGTCGG<br>CTGCGGCA<br>GCTCGACG<br>CTGCACGG | TAGACG<br>GCACAG<br>TTCTGG<br>CCCACG<br>TAGGCC<br>GAAAGG<br>GCCCCA<br>CCCATGT<br>AGGGAA<br>CCGCCCT<br>CGGGAA<br>ATCCATT<br>CCGGTG<br>TTCGACT<br>GAAGAG<br>ATGCTCC<br>TTCGCCT<br>TGACGG<br>AGGTAC<br>ATCTCG<br>ACAGTC<br>GAACTT<br>CAACTC<br>GCAACA<br>TATCCG<br>ATACAG<br>TTACAA<br>ACCACA<br>GTTAGA<br>TATCAG<br>ATAGGA<br>ACCTTCC | DEETKDSELPGRAPT<br>VLDEESKDNDATAGE<br>WLLRFDGACRANPG<br>PGGAGAVLFNPSGA<br>ASWTCSHFMPGATE<br>TNNTAEYTAFLLGAR<br>AAADHGATLLRVQG<br>DSQLVLRQVKGIYGA<br>KSTRLRRLRDAVRAEL<br>ARVGQFSLHHIDRQD<br>NAHADRLANRALDM<br>KSTLVECATHPGRNA<br>CTTTLTTSAAAESPAS<br>PPPVGARDTPMADA<br>GEERLADVDDGEVY<br>AAMRLGPGEVPERR<br>PRLRLRQLSDEELEAA<br>SEMVERLGAALSAKI<br>TDAEDWASAEGYITA<br>LPYMLYDKLQSYSQA<br>PRGPQQPVLTRSPRG<br>DDRPASSEPNASSTT<br>GGVASEHQPRRRRR<br>RGRRKGRRQRRNPR<br>RSGREGATGGHQQH<br>KKHKPRPPRETQHHR<br>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 |
| | ACAACACAAGAAACACAAACCACGCCCACCCCGCGAGACGCAGCATCAT | AGCGCCGC | | DNARRKHRPLYEVW |
| | CGGGAACACCGTCTCGACGAGGCTCTCGATGAGCTCCACGCCCTCGAGC | CGCCTGAA | | YDFRNAFGSVPFALL |
| | GTACCGATCCTCACAACCGGCCAGCAATTGCCAAGGCCCCGCCGCCGCGT | CGACGGC | | WDSLQRLGVPPDYV |
| | CGGTCGCATCCGGTCTGCCATCAACCAGCAGCTACTCCGCCACAAGTTC | 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 |
| | GGCATCCTCGCGCGGCGCGACTAACCCGGTGGATGGAACGCGAATGACCGC | ATCCGGCC | | AVKVYLYPRVEYALR |
| | CACGCACCCGGTCAGAAGGGATTCCGAGCGGTGAACGGCTGCGGGGA | TCCGTCCT | | HLRPDDQHLESFDLH |
| | ACACAACTTCCTCGCAGCTACGCTCATCGACAACGCCCGCCGCAAGCAT | CGCTGTGT | | LRRGLRHLLRLPKNAT |
| | CGGCCACTGTACGAAGTCTGGTACGACTTCCGTAACGCGTTCGGCAGCG | ACGCGCAC | | NEFFYAPVSRGGLGL |

| | | | | |
|---|---|---|---|---|
| | TGCCGTTCGCGTTGCTGTGGGATTCGCTACAGCGCCTCGGTGTCCCGCC | AACGCGCA | | LPLVELHAALQIAHG |
| | CGACTACGTGGACATGTGCAAGGGACTCTACAACCAAGCCTCGTTCGTC | GCGCTTCA | | WQTLHSPDPAIRRVA |
| | GTCGGGAACGCAGTGGATGGCAGCACAGCGCCGGTAGAGCAGCGCGT | ACTGCACG | | REQLYQIADARHRLD |
| | CGGCGTTTTTCAAGGCTGTCCGCTCAGCCCCCAGCTCTTTAACGCCGCCA | TTGTGCGT | | KDHWPHRREELCELL |
| | TCAGCCCACTGCTCTATGCGCTGCCCGGCTACCAGATACCGGCGTGCA | 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 |
| | CGTGCCTCACCACGCCGAGTGGCTGGACCACGGCAACGTGCTGAGGCA | GCACCACA | | EHLGLLKRDAKRLDR |
| | CGTGAAGCTACACATCAAGAACCTCCACTGGCAAACCTGGTGCGCCCTA | CTGGTCGC | | QLSSACIQSSRRIWN |
| | AGTGACCAAGGGAAGACCGCACGGGTGCACGGCGGTGTTGGCAGCGC | GGTTTCGA | | FHCAKHRARQHEHQ |
| | ATTCCTGACCCGACCTCGCGGCATGTGGGAAAGCGACTACCGCTTCGCG | CGACAGG | | APPSQATRGRRVTET |
| | GTGGCCGCTCGCCTCAACGTGGTCGACACCGTCAACACACACTGTCGCGCC | AGGAGCAT | | GGNPSRTDRR (SEQ |
| | GGCGACTCCGCGCCCACGACCGGTGTCGGTACCCGGCCTGCAGGTGGA | CAAGCCAC | | ID NO: 1433) |
| | AGGAGACTCTGGCACACGTCCTCAACCACTGCCCGGGTACCATGGACGC | ACTGTCGT | | |
| | GGTACGTGGCCCGGCACGATGACGCGCTCAAGGAAATCGAACACACGCT | CGGAGCCA | | |
| | CCGCGCGTCGTCGGGCGACCGCCGCGAGCTGCGTGTTAACCAAACGGT | ACACCACC | | |
| | GCCTGGCCTGCCCGGGCCACCCCTGCGCCCCGACATCCAGGTCTACAAC | GTCGCCAC | | |
| | CATGACAAACGCACGGTCGCAGTTGTAGACCTGGCTGTGGCGTTCGATG | GGCGGTCA | | |
| | AGCAGCCGAGTGAGGACCCAGAGAGCTCGGGGTTGGCCAAGGCAGTC | CCGCCGAA | | |
| | CAGATCAAGAAAGCCAAGTACGCGGGCATCAAGGAGCATCTCGAGAAC | CCCCCCCT | | |
| | CAGGGGTGGAAGGTTCATCTCTCGGCCATCGTGTACGGCTCGCTTGGCT | GCTCCACC | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NeSL | Utopia-1_PI | . | Phytophthora infestans | AGCTCGGCCTCGCGGCTGCCTTCCCAGGCGCCGCCGACTTCGCGCTCTG<br>GCGCGGCCCACACGCCGCCGCCGAGCCTCCAAGCGCGCCCGTTGGCTTT<br>CGCAGACGCAGGGCTCGCGGCGACGGCCCTCGCCAGCCCCCAAGACCC<br>CCCCTACGATGTGGCACCACCCGGCAGGGCGGCCGGCAGGCTGCCCGA | AGCTCGGC<br>CTCGCGGC<br>TGCCTTCC<br>CAGGCGCC | TAAGCT<br>GGTCAT<br>CATGAC<br>CGACAG | MLADPAALAAGLAR<br>APPPPSAPQDPSPAF<br>PAGPAGQNPRAAAP<br>ARVEVHTVVAPPGR |
| | | | | CGGTGGCGGCCAGCAACCACACAAGGTGTACACTGAACACCTTGGGTTGC<br>TCAAGCGCGATGCAAAGCGGCTGGACCGGCAGCTCTCGTCGGCATGCA<br>TCCAGTCCAGCCGCCGCATCTGGAATTTCCACTGCGCCAAGCACCGAGC<br>ACGCCAGCACGAGCACCAAGCACCACCAAGCCAAGCGACACGGGGCAG<br>GAGGGTGACGGAAACCGGGGGGGAATCCGTCGCGCACCGACCGCCGCT<br>AGACGGCACAGTTCTGGCCCACGTAGGCCGAAAGGGCCCCACCCATGT<br>AGGGAACCGCCCTCGGGAAATCCATTCCGGTGTTCGACTGAAGAGATG<br>CTCCTTCGCCTTGACGGAGGTACATCTCGACAGTCGAACTTCAACTCGCA<br>ACATATCCGATACAGTTACAAACCACAGTTAGATATCAGATAGGAACCT<br>TCCTTTAGGAAGCTAACGGGTACACTGGATGGTAAATACACATACATTT<br>C (SEQ ID NO: 1065) | ATCAAGCC<br>TCGGAACT<br>CACTGTGC<br>CCCCCCCC<br>ACGTGTGA<br>GTTCCCCC<br>GACGTCGA<br>TGTGCAGC<br>TGCACAGT<br>CCGCCACA<br>GGAAGATC<br>AGCACGA<br>GGACGCCA<br>CCCANCAC<br>CCGGAGA<br>GCACGCAA<br>CACCAACC<br>TCCTGAGG<br>CAACCCGC<br>TGGGGTTC<br>GCCGCTCG<br>CGCCCACG<br>CTCGTTGC<br>CTCCAGGA<br>TTGCTCAG<br>CGACTCGG<br>CGAGCTGG<br>AACCTCCA<br>CGCTGGG<br>GCCCACCA<br>TTACCCCG<br>CGCG (SEQ ID NO: 1188) | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | CTCGGTATCGCCGGGTGCTACACTCTCAGCCGCCACAGCTCGGGCCTTG | GCCGACTT | GGCACT | AGGMLPDPGLVEEPI |
| | | GCGGTCCGCCATTGGCCCTTGGAGCTCGACAGCGACAGCAGCGACGAC | CGCGCTCT | ACCCAG | QATYAHDAAQFECA |
| | | GAGGACGCTCAAGACCCCCACGCCCGCCCCAGAACCCCCAGAAGAC | GGCGCGG | GTAGGG | LCPYVAESMAVLVQ |
| | | GTCGCGAGTGTGCTTGCCCCACCCGGCAGGGCAGGCAGCATGCTCGCC | CCCACACG | AACCGC | HRRSAHRGTRFKDIF |
| | | GACCCAGCAGCGCTGGCAGCCGGTCTGGCTCGCGCTCCGCCCGCCCCCAT | CCGCCGCC | CCTTAAA | TSGCQCSLVFYARIVA |
| | | CGGCACCCCAGGACCCAAGCCCCCGCCTTCCCTGCTGGCCCCGCCGGCCA | 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 |
| | | CCGCCAAGACCACTGGGCTTCAGCTCGTCCCCCCGAGCTGGAGCCTGC | 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 |
| | | CCATGTCACCGCGCGATCGGGGGCCGCAGCCATCGCACGACTTGTCACGG | 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 |
| | | CCGAGTGCGTCACGCGCTCCGCTCCGTGGACACGCACACAAACTCCGGCAC | 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 |
| | GACGTGTGGGCCGGGGTGCGAGCCTCAATGGCACAACAGGAGCTGCGCC | | CGGTTG | HPETCPIDAATLHAH |
| | ACGAGTTTGCCAAGGACGAGTCTAAGTGTGTGGCCAAGATCCTGAAGA | | TCGGGC | FTGVNAPRTDFDYDA |
| | CAGCGTCGACAGAGACGGCAGCGGAGGACGAGCACCCGGAGACGTGC | | TTTTTTC | 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 |
| | CCCAGAAGGGGTTCAGGGCCTTCAACGGGTGCCACGAGCACAATTTCAT | | 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 |

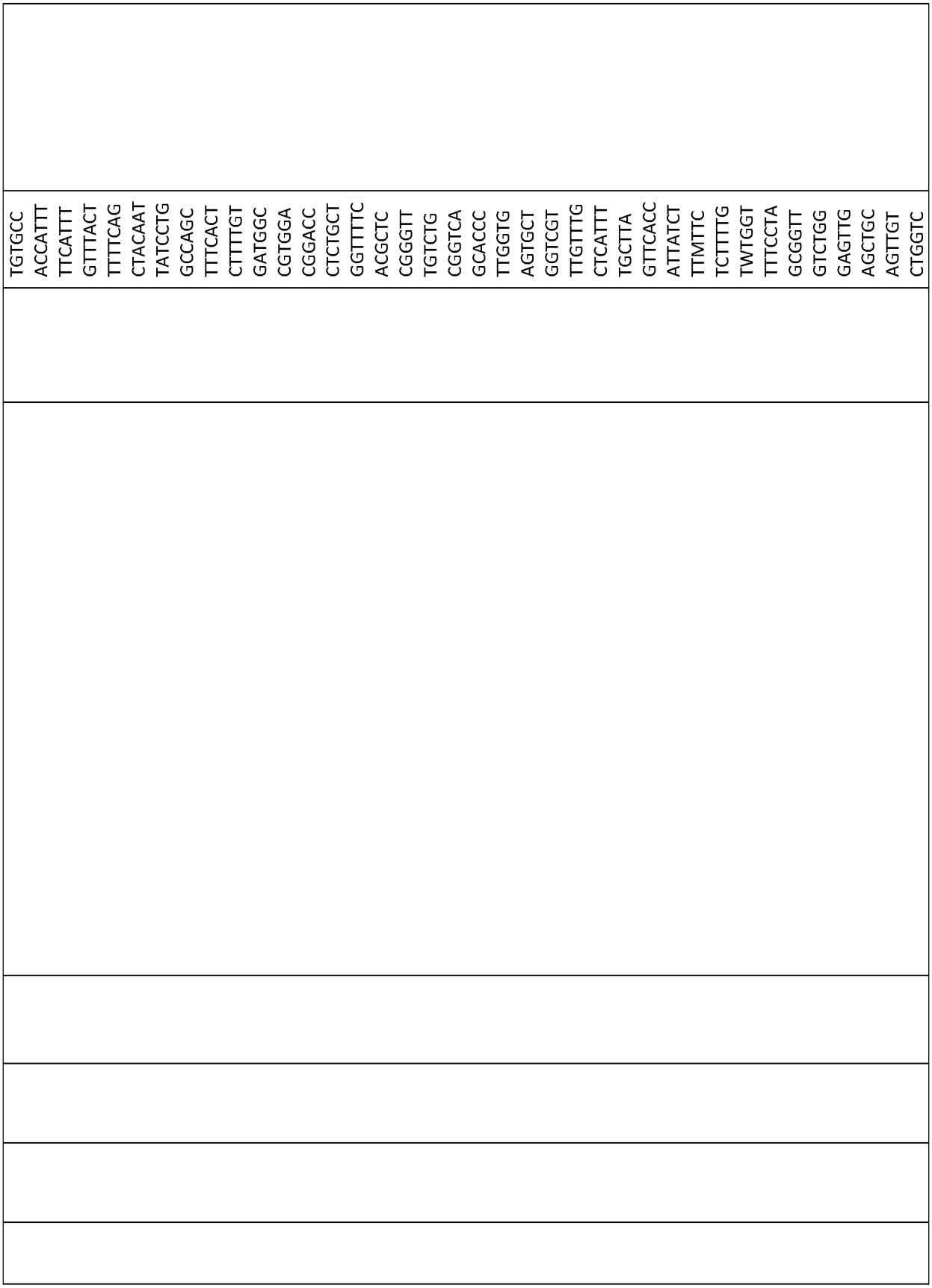
| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | AGGCCG | TLSLHESYRYLGVGD |
| | | | | | ATCTGG | GYDHVRHRLQLEPKL |
| | | | | | CGGCCC | KQLKREAVALLTSGL |
| | | | | | CCGATT | APWQVVRALKVVYY |
| | | | | | GGTCCG | PKVEYALRHLRPLQS |
| | | | | | ACTTTCG | QLQAFDRVVSKGLRH |
| | | | | | GCCAAT | LLSLPRSATSEVLYAP |
| | | | | | CAGCGA | TSSGGLGLQPLVELH |
| | | | | | CGACGA | RALQLAHAWQMLH |
| | | | | | GGGGGC | SKDPAIQAVARAQAC |
| | | | | | AGGGGT | QVVRKRYRLQEDHW |
| | | | | | TTACACT | RGRDDELVRSFLNSE |
| | | | | | TTTGCCC | LAASPHAEVLRRNGD |
| | | | | | CCGTTTC | IASLWSDVQRWLRIY |
| | | | | | GACTTC | HLRFEHCDETEAHGP |
| | | | | | AACTTCA | LSFRVPHHNKWLTH |
| | | | | | GGCCAA | KTVLRHVKLHLKIRH |
| | | | | | AATGGC | QTRWKGMVDQGKT |
| | | | | | GATTTG | VRVHGGVGAKFMTT |
| | | | | | GACCCT | GAGLSDDDYRFGVK |
| | | | | | CCACGC | GRLNQVDTNSVLKRK |
| | | | | | GCCGTG | RLRAHTTCRDPTCSS |
| | | | | | CCACTG | AETLAHVLNHCESN |
| | | | | | CTCGGC | MDAIRQRHDDALEQ |
| | | | | | ACCGGC | IGSKIRGALDRAKSPT |
| | | | | | GGCGAT | ELRLNQTVPEYTGAA |
| | | | | | TCAGCG | LRPDIVLRNVAAKTM |
| | | | | | GGTGCA | VIADLAVTFEDQAAR |
| | | | | | ACTTCG | ARHSSLQLSHDHKTL |
| | | | | | GGCACG | KYQPIVAELQHKGW |
| | | | | | TGTGCA | RVQTAAIVYGTLGSV |
| | | | | | ACACAT | QPSNFKAYTEKFKLH |
| | | | | | GCAGCG | KREARQLDLQLSSHCI |
| | | | | | CCCATTG | RASHRIWGWHCRQ |
| | | | | | CACGCC | HRDRQRSGTASRASR |
| | | | | | AAGCGG | GSGGAPRRTSQAPA |
| | | | | | CATCGC | RR (SEQ ID NO: |
| | | | | | GGGACG | 1434) |

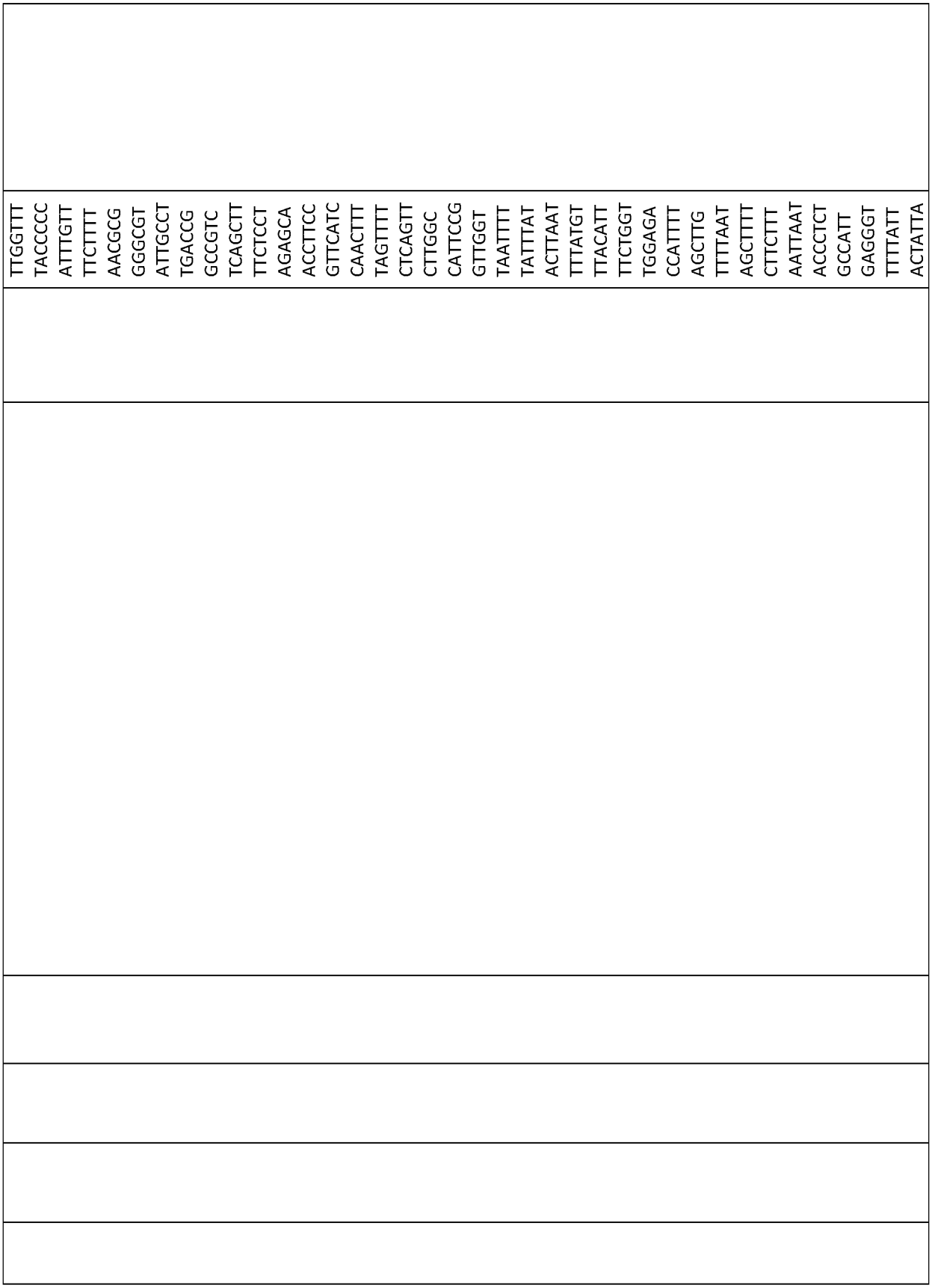
| | | | | | |
|---|---|---|---|---|---|
| | | | GGGCAATCGTCTGCGGCTAGCTTGCTAGCGGCGGACGAGCGGTCTCCG<br>GGGGCGTTCACCTTTCCCCCGCGAGGCCAACTACACCGATCTTCTCTACA<br>CTTTTCTAATTCGCCTCCGTCTTCGGTCTTCGGTTGTCGGGCTTTTTTTCTTT<br>TTGACCAATCAGAGCGCGCCATGCGCCTCTTCTGGCCAATCAGAGACCG<br>GGCCCTGTCCTCGGACAGCGAGGCCTCCACGGCCAGCCAATCGAGTCTC<br>GGCAGCGACGCGTCTTTCTATAGCGCAGCTGACGAGGCCGATCTGGCG<br>GCCCCCGATTGGTCCGACTTTCGGCCAATCAGCGACGACGAGGGGGCA<br>GGGGTTTACACTTTTGCCCCCGTTTCGACTTCAACTTCAGGCCAAAATGG<br>CGATTTGGACCCTCCACGCGCCGTGCCACTGCTCGGCACCGGCGGCGAT<br>TCAGCGGGTGCAACTTCGGGCACGTGTGCAACACATGCAGCGCCCATTG<br>CACGCCAAGCGGCATCGCGGGACGACGCCTCGGCCGCCCAAGCGCAGC<br>CCCGCCCTTCCAGCACGACCTCGCGCCGTTTGGCGGATCGCCATCAAGA<br>CGTGCGAGAGCCAGGCGGGGTCGGGCAAAATATACTTACTCTAAGTAT<br>GCCCGAATCCCTGCCCTCTCAGGCTGAACGCGGCCCCATACTTGATCTAA<br>GTATGGGAGGATCCCTGGCCTCTCAGGCTGTACGCGAGACCC (SEQ ID<br>NO: 1066) | | ACGCCT<br>CGGCCG<br>CCCAAG<br>CGCAGC<br>CCCGCC<br>CTTCCAG<br>CACGAC<br>CTCGCG<br>CCGTTT<br>GGCGGA<br>TCGCCAT<br>CAAGAC<br>GTGCGA<br>GAGCCA<br>GGCGGG<br>GTCGGG<br>CAAAAT<br>ATACTTA<br>CTCTAA<br>GTATGC<br>CCGAAT<br>CCCTGCC<br>CTCTCAG<br>GCTGAA<br>CGCGGC<br>CCCATAC<br>TTGATCT<br>AAGTAT<br>GGGAGG<br>ATCCCTG<br>GCCTCTC<br>AGGCTG<br>TACGCG<br>AGACCC<br>(SEQ ID<br>NO:<br>1312) | |

| NeSL | Utopia_1_PMi | . | Patiria miniata | CTGATGTGGATACCTTGGAATTACTCAACCGTGTCGGAGTCTTTGTCTT<br>TTGCGCCCAACACCTCATGGATACCATGCTTGTCGCTGGAGCGACGTTA<br>CAAGCGTGAGGGCGCCCTCCATGCCGGACAGCTGGTCTGTGCCATGTG<br>CTTGAAGTCGTTCTCCTCGACTTCGGGCTTGCGGCGGCACATGGCCCGG<br>CTGCACCGGCAGCCTTCTCCCGACGCCTCCACGCCCTCGACTATGACGG<br>AAGTGTTCCCGTACCCCCTTCCGAAGGTGTGGCCGTGCGTGGTGTGCCG<br>CGAGAACTTCTATCACAATCAAACCTTGAAGCGTCACCAGAAGAATTTTC<br>ACCCCAATGTTGACCTCACCACCGTCTACCAGTGCTCAGTGTGCGGGCA<br>GGAGTTTGTCACCGGCAGGAAAGCATCCTTCCATTTCAAAGTTCATCGG<br>CGGATGTCAGCCAGCGCCATCCCTACCCCGTCTGCAATGCCCTCCTCACC<br>GATGGATTTGATCCGAGGCCTCGTCGGGGAACCCCTGCCTCCTAGCCCT<br>GCCCGTACACCGCCACCACTTGCCCGATACATCTCCCCAGCGCCCCGGTC<br>TTCTTTTCACCACCCTGGAATCCCTCCCCACCGCCCCGGTCGCCCACCCC<br>ACTGCCACGGCCACTCACTCCACCGCCACGATCACCGAGCCCGCCGCCC<br>CGGTCGCCAACACCACCTCCCCCGGTCACCCTCACCACCGCCCCGGTCAC<br>CGAGCCCGCCGTCCCGGTCGCCCTCACCACCGCCCAGGTCACCGAGCCC<br>TCCGCCCCGGCCGTCCACACCGCCGCCCCGGTCACCCTCTCCAACGCCCC<br>GGTCACCGAGCCCGCCACCCCGGCCACCGATCCTGCCACCCCGGTCACC<br>AGACTCCACTCACCGGTCACTCACATCTCATGCTCGATCTCCTTCACGGC<br>CAGCCACGCCCCCTATTCCTGTGCAGCCCCGACATCTCCGTCCGTCTACG<br>CCTGCTCACCCAGGCGTAGACAATGCAGTTCCACCATCGCAGCAGTCTG<br>CAATTGACGTGTGGCTGGCCGAGCTGTCACGCTCGGCTGACTTCGAGTC<br>TTTTGAGGATGTGTGTGATCGGTTTGTGGAGTTTGCCGCCGCTGAGGGT<br>CGAAACAACGGACGCCCAGCCCGCCACCAACCGCCCCCGTGAC<br>CGCGGGAACCAGGGACCTCGCCCCCAACGTCCCCCTCGTCCCCACCGCC<br>CCGGGCTTGGCCCCGGAGTTTGATGCACAGGAGGCCAGTCGTCTGCAAA<br>AGCTGTATCGGACCAGCAAGAAGCGCGCCATCAGGACTATCTTGACCG<br>GCTCTGATGTCCGATACAGCGGCTACCGAGGTCCTCACTAGTGGATTTC<br>TTTCACTTGAAGTATTTTCCGCTCGTGAGGTCGACACCGACACCATTGCC<br>ACGGATACCAGCCTCCTGTTTCCGAACTCCGCCCAAGCCCGTGAGTCCG<br>GCCAAGACCTGCTTCGCCCAGTCACCCAGCGTGAAGTCTCCCTCCGTCTT<br>GGCCGGATGTCCAACTCCGCACCAGGGAAGGACCGGTTGGAGTACCGT<br>CACATCCGCCAGGTTGATGGCGCATTCCGCGTCACCCTGGAAATCTTCA<br>ACCGGTGCCTGCGTGAAAGCCGGGTCCCATCCTCCTGGAAGACAGCGA<br>CCACCGTGCTGATCCACAAGAAGGGGGATGCCACCGACCCGGCCAACT<br>TCCGCCCCATCGCCCTCCAGTCCTGCCTCTACAAGCTGTTGATGGCGATC<br>TTGTCCGACCGTGTCACCACATGGGCCCTGGACAACGATCTCATCAGCA<br>GCTCGCAGAAGAGCGCCCGCCCCGGTGAAGGTTGCTACGAACACACAT | CTGATGTG<br>GATACCTT<br>GGAATTAC<br>TCAACCGT<br>GTCGGAGT<br>CTTTTGTCT<br>TTTGCGCC<br>CAACACCT<br>CATGGATA<br>CCATGCTT<br>GTCGCTGG<br>AGCGACGT<br>TACAAGCG<br>TGAGGGC<br>GCCCTCCA<br>TGCCGGAC<br>AGCTGGTC<br>TGTGCC<br>(SEQ ID<br>NO: 1190) | GATTAG<br>CGAACA<br>CTAATAT<br>CCTGCC<br>MTAGAC<br>GTGATT<br>GCTAAT<br>CCGCAA<br>ACCAAC<br>CGGATC<br>TACAGC<br>CTGAAC<br>ACTGAA<br>CTTTAAT<br>CTTCACC<br>CATGTC<br>ACATCT<br>GGACAC<br>TAGGTTT<br>TTGCTCT<br>GTTTGT<br>GTTTTCC<br>TGCCTTT<br>WCMTT<br>GGAWCT<br>TTCGCM<br>CTGGAA<br>TTTATTT<br>GTCGCT<br>TGGATT<br>ATTTTTT<br>TTCTCAC<br>AATTTG<br>GATCTAT<br>TTTCGTT<br>TGTTCAC<br>TTCGAA<br>CTCTAGC | MCLKSFSSTSGLRRH<br>MARLHRQPSPDAST<br>PSTMTEVFPYPLPKV<br>WPCVVCRENFYHNQ<br>TLKRHQKNFHPNVDL<br>TTVYQCSVCGQEFVT<br>GRKASFHFKVHRRM<br>SASAIPTPSAMPSSP<br>MDLIRGLVGEPLPPS<br>PARTPPPLARYISPAP<br>RSSFSPPWNPSPPPR<br>SPTPLPRPLTPPPRSP<br>SPPPRSPTPPPPVTLT<br>TAPVTEPAVPVALTT<br>AQVTEPSAPAVHTAA<br>PVTLSNAPVTEPATP<br>ATDPATPVTRLHSPV<br>THISCSISFTASHAPYS<br>CAAPTSPSVYACSPR<br>RRQCSSTIAAVCNSE<br>ASSGNPCLLALPVHR<br>HHLPDTSPQRPGLLF<br>HHPGIPPHRPGRPPH<br>CHGHSLHRHDHRAR<br>RPGRQHHLPRSPSPP<br>PRSPSPPSRSPSPPPR<br>SPSPPPRPSTPPPRSP<br>SPTPRSPSPPPRPPILP<br>PRSPDSTHRSLTSHA<br>RSPSRPATPPIPVQPR<br>HLRPSTPAHPGVDN<br>AVPPSQQSAIDVWL<br>AELSRSADFESFEDVC<br>DRFVEFAAAEGRNN<br>GRPARPAHQPPRDR<br>GNQGPRPQRPPRPH<br>RPGLGPEFDAQEASR<br>LQKLYRTSKKRAIRTIL |
|---|---|---|---|---|---|---|---|

| | | |

| | | | | | |
|---|---|---|---|---|---|
| | | 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

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NeSL | Utopia-1_PS | . | Phytophthora sojae | AGACGAGCAACGCGCTGGGGCCCAAGACCTGGCACGAACGACACTGCC | AGACGAG | TGAAGC | MDVDGGPAMPEPW |
| | | | | CAGGCTGTGAACGACGAGCACGCTGCTAACCCGGCAGCGCACCGGCCT | CAACGCGC | TGCACA | VLRFDGACRRNPGP |
| | | | | CTGGGCTCCGCTGCACCGGTTACCGGTGCAACACGGTGGGGTCCACGTC | TGGGGCCC | AGCGCG | GGAGAALFKPCGTV |
| | | | | ACGGCGCGATCGGGGCAGCGGCTGTAGCTCGCCTGCTCACAGGCCTAC | AAGACCTG | AGGCTC | AWTCSHYMPNSSET |
| | | | | CCACGGCACCAGCACCAGCTACGCGCCGGCCTGCTTCGGCTCGGCGCTG | GCACGAAC | GACAGC | NNTAEYTALLLGVQS |
| | | | | CCCAGACCCGCCCGCTCCCCCCGCGGCCACGACGACAGCCCCCGATGCG | 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 |
| | | | | GAGGACATCGACGCAGAGATTGCCGCGCGCGACGGCGGCGAGGCGTT | GTCACGGC | ACAACG | SGDGYISAIPSRIRQLL |
| | | | | CCCCACGCTGCCTATCGGGCCCGGGACGGCTCCAGCGCGGCAGCCGCG | GCGATCGG | GAGCGG | QPFTAAQPHPRPPLQ |
| | | | | CCTTCGACTGCGGCAGCTGACCGAAGACGAGGAAGCGGCGGCGGCGTC | GGCAGCG | CAACGC | QQRQRPPRVTRTQR |
| | | | | GGCCCTGCAGGCCATGGCGGAGGAGCTAGCGTGCAAGATCGAAGACG | GCTGTAGC | CTCGCG | EHRLDEALDEMAAV |
| | | | | CAGACAGCTGGACCTCGGGCGATGGATACATCAGCGCCATCCCCAGTC | TCGCCTGC | AGCGCC | QQERPTSRSAVRRAR |
| | | | | GCATCCGGCAGTTGCTGCAGCCCTTCACAGCAGCCCAGCCCCACCCTCG | TCACAGGC | GCGTGG | RRVGRIRASMRQQQ |
| | | | | GCCTCCACTGCAACAGCAGCGACAGCGACCGCCCCGGGTGACGCGCAC | CTACCCAC | GTCTGG | LRHDFARNESKCVED |
| | | | | TCAGCGCGAGCATCGACTCGACGAGGCCCTCGACGAGATGGCGGCGGGT | GGCACCAG | GGGGGG | ILRAASAETAAEEHPE |
| | | | | GCAGCAGGAGCGACCCACCAGCAGATCGGCCGTCCGCCGTGCGCGGCGG | CACCAGCT | CCCGCG | TCPIDSGTLHEHFTAV |

| | | | | | |
|---|---|---|---|---|---|
| | CCGCGTCGGCCGGATCCGCGCCTCGATGCGCCAACAGCAGCTCCGCCAC | ACGCGCCG | GCGCAC | | NSPRINFLPDEACGA |
| | GACTTCGCTCGCAACGAGAGCAAGTGCGTCGAGGACATTCTGCGTGCT | GCCTGCTT | ATCGCA | | LFREAMADVGTPQE |
| | GCGTCGGCGGAGACCGCAGCGGAGGAGCACCCGGAGACGTGCCCCAT | CGGCTCGG | GGCTCG | | RRSALTDELTMDEVE |
| | CGACTCGGGCACGCTCCACGAAGCACTTCACGGCGGTGAACTCGCCCCGC | CGCTGCCC | GGCACG | | DQLMQAATNSSPGH |
| | ATCAACTTCCTACCAGATGAAGCTTGCGGGGGCGTTGTTCCGCGAGGCGA | 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 |
| | CTCCAGCTAGAGCCCAAGCTACAGCAGATCAAGCGGGAGAGGCGGTGGCG | | 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 |
| | | | | TCCCCTCGGCGTGCGCACGTCCCCCCCGTGCCGCGCGCCACCACGACAC | CTCCAGCA | AAACCC | | RFDGACRRNPGPGG |
| | | | | CATCCGCGCTGCGGATTGGTGGCAAACGCCGCCGCCTGAACGACGACG | GCATCGGC | AGTTCA | | AGAALFEPGGAVVW |
| | | | | GCGACAACGAAAACAGCGACGGCCGCGACGCCGACATCGAGATGCGC | GCTCCGCG | GTTTATT | | TVSHYLPGSETNNTA |
| | | | | GCCGACGACACCGACCACGCGCCGACAACCCCCGCGACCAGTGCG | 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 |
| | CTGGCGCCAACAGCCCGCCTGTCCAGCTCCAGCTCCAGCTCCA | 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 |
| | CGCCTGCGTCGAGAGCATCCTCCTCCGCGCCGCCAAACCCGACGGCACCGA | 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 |
| | | TCATCTCAGCATTGATTCCTCTCCTGCGGGCTCTGCAGCGGCTACCAGGC | 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<br>CAGATTGGTGCGGCGATCAACGCTGCCATCAAGGGCCGCCGCACCGAC<br>ACCGAGGTCCGACTCAACCAGACAGTTCCAGAGTTCAACGGACCGGCG<br>TGGCGGCCCGACATCCAGGTGCGCGACGCGCGCTCCAAGACCATGGTG<br>ATCGCGGACCTCGCGATCACCTTCGAGGACCAGCCAAACGACCAGTCCG<br>CCAGCTCGTCGCTCCAACACAGCCGCGAACACAAGATCGCGAAATACCA<br>GCCAATCGCGGCGGCACTGGAGCGGCAAGGGTGGCGGGTACACACAA<br>GTGCCATTGTGTACGGGTCGCTACGCTCGGTCCACCCCAGCAACTTCAC<br>CGTGTACACGGAGTTGCTGGGACTGCTCAAGCGCGACGCCCGTCGCCT<br>GAACACAACGCTGTCGTGCCACTGCATCCGCAGCAGCCGACGCGTGTG<br>GAACTGGCACTGCGGCCAGCACAGAGCACGACAGCATCAGCGATGCCA<br>AGAAGGCCGCGCGCACGGGTCCGGGGGGGAACCAGCGCGCGGAAGGG<br>GGCACGGCAACAACATGATGCGGGTCATATTGACCGAAAGGGCACCAT<br>CCACGTAGGACACCGCCCTCAAAAACCCAGTTCAGTTTATTGACACCCTC<br>CACTTAGTGACATGCATATTCAAACCGATACATATTCGTTAGGAGAGGT<br>TACTAACTGGTAATATATCACCATTTC (SEQ ID NO: 1069) | GGAGTGTT<br>CTTTTCGG<br>AGACGCCG<br>CCTACTTT<br>AGAGGAG<br>AGAATCCC<br>CACGGGCA<br>TCCTCATTT<br>GATCTGAT<br>CCATCGAG<br>TATCTGCG<br>AATAGTCG<br>GCGCACTC<br>CTTTTGCC<br>ATGATCCC<br>GGGGGTC<br>TCATGGTA<br>AAAAGGTT<br>TGTGGCAC<br>GGCTTAGT<br>TGACGCCC<br>CTCTTCCA | TGATGG<br>GAGAGT<br>GAGGAA<br>TCTTCTC<br>CACTGT<br>GCAAAA<br>CCATAC<br>AGTCAG<br>AAGATG<br>CTAACTA<br>CTAGTTT<br>GATACC<br>CTGTGC<br>CCCCTGC<br>AATGTC<br>CCGCGT<br>GTCGTA<br>CCCAAG<br>CCCGGC<br>TGGCAT<br>TGAGAC<br>ACATTA | TSAIVYGSLRSVHPSN<br>FTVYTELLGLLKRDAR<br>RLNTTLSCHCIRSSRR<br>VWNWHCGQHRAR<br>QHQRCQEGRAHGS<br>GGNQRAEGGTATT<br>(SEQ ID NO: 1437) |
| NeSL | Utopia-1_SM | . | Strigamia maritima | GGAGTGTTCTTTTCGGAGACGCCGCCTACTTTAGAGGAGAGAATCCCCA<br>CGGGCATCCTCATTTGATCTGATCCATCGAGTATCTGCGAATAGTCGGC<br>GCACTCCTTTTGCCATGATCCCGGGGGGTCTCATGGTAAAAAGGTTTGTG<br>GCACGGCTTAGTTGACGCCCCTCTTCCACGTCACTCGGCCTGCATCGATC<br>GACTGCTCTCCTCTTCCTCCCTCCTCTTCACGCTCTCTTCACGTGA<br>CCCCTTCCCATCCCGCCCCTCGGCTTTTGGCAAAGATCTGTGTCCTCC<br>AAAGCACCCATCTACCATTTGCTCGAGTTGCGATTGGTCGAAGCTGCCA<br>CGCCACTCGTCTACTCTGCCTCTCTGACCCCTTCCCCCCTCTCCCTCTGC<br>CGCCCTCGCCTCTCGGGTCCCTTCCCATCTCCCTCCTCCCATCACGTGCT<br>CCCTCCTCTCGTCACGTGATCGTTGCGGCTCGACATGGCCACGGTCAGA<br>CTCAAGTACCCCTATCCTCCCGAGGGTATCTTGTGTGGACCGTGTGCCGC<br>CAACACGAACGCCCCTCAGACGCGGCCATACTCCGATAAGTCGGGGTTG<br>GCCAAGCATCTAAAGCTCTATCACACAAAGCCACTCTAGTGGTGGAGTGTC<br>GTCACTGTGGTCACGAGGAGTCTGATCTGAGGAAGATGAAGAAGCACA<br>TCTCGACGAACCACACCCCGTCGCTGCGGCCGCCGCCTACCGTACCCCC<br>TCGCCTGGGTCCCCACCGCCCCTCCTCCGCCCCGGGTTATCCTCCGCCCTC<br>GCTTCATCCCACGGCCTCGGACTCCCTCTCCTTCCTCCTCGTCATCGG<br>CGTCGTCTCCGGCCTCTTCCAGAAGGTCTGTATCCCTTCCCCCTGCGAGT<br>CCCCCCGTCTCCTCCGCTTCCTCCCCCGGCCGCTCGCTCGGGTCGGAATTC<br>GCCGGACTCGCAGGGGACGGCTCCGGTCACTCCGATCGGTACCGTACG<br>GAACTCGCCGGCCGGTAGCCCTGCCCTCAGCTACTCGACGGCCTCCCCG<br>ATCGCCTCTACCATCACGACGCCACGCCATCTCTCGCCTGCGTCCCCCGC | | | |

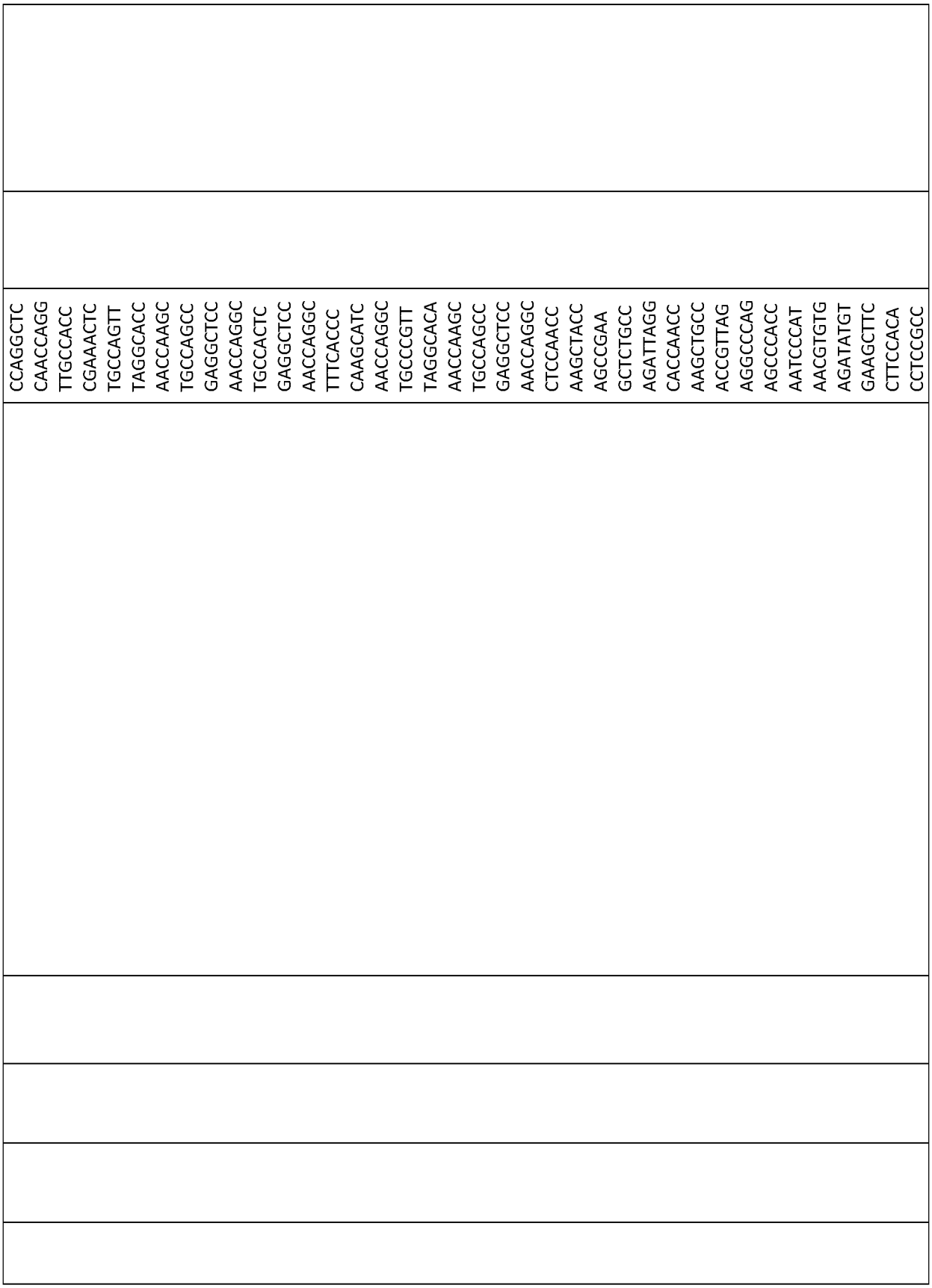
| | | | | | |
|---|---|---|---|---|---|
| | | TCTCTCTGCCGGGCCTGGCTCTCTCGGCGCGTCTCCGCCCGTCTCTCCCA | CGTCACTC | GGCTCT | SLAALTSSCIREDHRG |
| | | CGGCCGCTACAGTACCCCCCCGCTCCTCCAGCTACCGTACCCGCAGTGAT | GGCCTGCA | CGCTCCC | SLPLLCTSGWDSPWP |
| | | GGCGGCTACCGTACCCTTCGTCGCCGCGACTACCGTACCTCCGTCGGC | TCGATCGA | 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 |
| | | GACACTGGAGTGGTCCAAGGAGCTCCCCTCTCCCCCATCCTCTTCAATCT | 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 |
| | | CCTCGAACGAAATACTGTACATGTCCCACTTGCACGGAGGGCAATCCCT | ACGTGATC | (SEQ ID | GSPVPSISDGDAYKH |
| | | ACTCCCCCTCTCGGTGCTAGCGGACATCGGACAAGTGACTCACGCCGTC | GTTGCGGC | | LGVPTGTTTFPSGTD |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GCTCTTCTTCAATCAAGGGATCCAGCCGTGGCTGACCTCGCTCTCCGGAC<br>TTGCCGAGAAGTCGCATCGAAACGGGCAAAGAAGACGGTCAACGGCCC<br>CGAGTTGGCCCAGTACCTCTCCGGGTCCACTGACGGAATCTACTGCACC<br>CCGACTTCTGACATCCCGAGTCTCTGGACGACTGCCCGAGCTGCAACAC<br>GCCGACTCTCCTCCACTCTGCCACTCACGTGGACATCCCCCCCTTCCCTCTG<br>GCGTCCCCTTCCTTTCTATCAATGGGTCCCCCCTCTCCCCATTCCGAGTAC<br>AGTCCACCCTCACCAATGCCATTAGACATAACCACCTCTCCACTCTCATT<br>GCCAAACGGGACCAGGGAAATTCCTACCGTACCTCACACGACCCCGACC<br>CCTCTAACTACTGGGTCAAGGGAGGAGACTTCCTCCGATTCTGTGACTG<br>GCGATTCATACACAGAGCCCGACTGAACCTCCTCCCAGTGAACGGGGCG<br>CGCCGATGGGACGCAAACAGCATCAAGACGTGCCGGCGCTGTGGTGCG<br>CCCAATGAGACACTCGCCCACGTGCTCAACGTCTGCCCTGTCGGACTGC<br>CGGAGATGAAGAAGAGGCACGATGCAATTCATGCCCGCATCAAGAAAG<br>CCCTCCGACCCTCTCCCCATACGGTAGTACACCACGACCGGACAGTTCCC<br>GGATGCGGCCCCCTGAGACCGGACATTCTGAGAATCAGCGAGAGAGAC<br>AAATCTGTTGCCATTGTAGATATCCACGTTCCCTTCGACAATGGCACTGA<br>TGCGGTTGAGCGAGCTCACGAGACAAAGCGGGCGAAGTACGAACTGAT<br>CCGGCGGCACTACGAACACCAAGGGTATCGAGTCACGTTTGACTCACTA<br>GTAGTCACAGCGCTTGGTAGACTGTGGCGGGGGAGTGAGGCAGCGCTA<br>CAAGCGCTCCAGATCTCGTCACAGTATAGTAAGCTCCTCCGTAAGTTACT<br>CGTGGCTGATGCCATCCACGGCTCGAGAAACGTGTACGCCCATCATATG<br>ACCGGTATGGTTATGTGATGGGAGAGTGAGGAATCTTCTCCACTGTGCA<br>AAACCATACAGTCAGAAGATGCTAACTACTAGTTTGATACCCTGTGCCCC<br>CTGCAATGTCCCGCGTGTCGTACCCAAGCCCGGCTGGCATTGAGACACA<br>TTAGGCTCTCGCTCCCCCGTATACTCTCATAATTTCGTGTACGCTAATCCT<br>ACCCTACCCCTCCCTTTGACCACTCACCCAACCATGTGTATAGCTGTGCT<br>GGTGATCCCGGGGCGGTTATTCACTGGTTATCATATCATTCTAAAATGAT<br>CTTTGATCTCTCAATTAACAACTAACTTATTTCTGTTTGTTTCATTGTTTCA<br>CCTCGTAAGAGGAAGTTCATTGTGCGATAAATCAA (SEQ ID NO: 1070) | TCGAC (SEQ ID NO: 1193) | NO: 1316) | TGACGA | AIKKMTTDLQAIDHS |
| | | | | | | AATGTTC | DLAPAQKLDALRTFI |
| | | | | | | AATATAT | MPQLSFHLSHGSVPK |
| | | | | | | GACATT | APLTQLDKKIKRAAK |
| | | | | | | CTAATGT | HWLFLPQRASNEILY |
| | | | | | | TCATGTA | MSHLHGGQSLLPLSV |
| | | | | | | TTTTTTG | LADIGQVTHAVALLQ |
| | | | | | | | SRDPAVADLALRTCR |
| | | | | | | | EVASKRAKKTVNGPE |
| | | | | | | | LAQYLSGSTDGIYCTP |
| | | | | | | | TSDIPSLWTTARAAT |
| | | | | | | | RRLSSTLPLTWTSPLP |
| | | | | | | | SGVPFLSINGSPLSPF |
| | | | | | | | RVQSTLTNAIRHNHL |
| | | | | | | | STLIAKRDQGNSYRTS |
| | | | | | | | HDPDPSNYWVKGG |
| | | | | | | | DFLRFCDWRFIHRAR |
| | | | | | | | LNLLPVNGARRWDA |
| | | | | | | | NSIKTCRRCGAPNETL |
| | | | | | | | AHVLNVCPVGLPEM |
| | | | | | | | KKRHDAIHARIKKALR |
| | | | | | | | PSPHTVVHHDRTVP |
| | | | | | | | GCGPLRPDILRISERD |
| | | | | | | | KSVAIVDIHVPFDNG |
| | | | | | | | TDAVERAHETKRAKY |
| | | | | | | | ELIRRHYEHQGYRVT |
| | | | | | | | FDSLVVTALGRLWRG |
| | | | | | | | SEAALQALQISSQYSK |
| | | | | | | | LLRKLLVADAIHGSRN |
| | | | | | | | VYAHHMTGMVM |
| | | | | | | | (SEQ ID NO: 1438) |
| NeSL | Utopia-1_SP | AAGJ021405 37 | Strongylocentrotus purpuratus | AAGGCTCAAACCAGGCTGCCAACCAAGCTGCCAGCTTAGGCACCAACCA<br>AGCTGCCAGTCCAGGCTCCAACCAGGCTGCCCACCAATCTGCCAGCTTA<br>GGCTCCAACCAAGCTGCCCACCTAGCTGCCAGCTTAGGCACCAACCAAG<br>CTGCCAGCCGAGGATCCAACCAGGCTGCCCACCTAGCTGCCAGCTTAGG<br>CTCCAACCAAGCCGCCAGCCCAGGCTCCAACCAGGTTGCCACCCGAAAG<br>TCTGCCAGTTTAGGCACCAACCAAGCTGCCAGCCGAGGCTCCAACCAGG<br>CTGCCACTCGAGGCTCCAACCAGGCTTTCACCCGAAGCATTACCCAGGC | AAGGCTCA<br>AACCAGGC<br>TGCCAACC<br>AAGCTGCC<br>AGCTTAGG<br>CACCAACC<br>AAGCTGCC | | | |

| | | | | | |
|---|---|---|---|---|---|
| | | 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 |
| | | TCCATTTCCTTCTCACCCATGCATCTTAGCCCCCACTCAGCCCGCCAGCCCG | 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 |
| | CCTTTCTCGAATGAGCAATTCTGCCCCTGGAAAGGACCGCCTCGAGTAT | 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: 1439) |
| | | TGAAGAGGACATCGACAACTACTTGTCTGGCAGCAACGAGAGGGCCTGCTT | ACCCGAAT | | |
| | | GCCAATTCGGGTAATAGCGGCCAAGCTTCTTCTCTGTGGTCTCGAACCA | CTCTGCCA | | |
| | | GGTCAGCCGCAAGAAGACTTCATCTTACCCTCAGAGCCACCACCAGTGG | GTTTAGGC | | |
| | | TACAGTTGTTGTCAACCAGCAGGCAGACATCGACCACACACAAGGGACATT | ACCAACCA | | |
| | | CTACCAGCATCAATAACAAGAGGACTGAGGCTCATCCAACGAACCACAA | AGCTGCCA | | |
| | | ATGCTGAGAAACTCAAGTCCCTACCAGACCAAGGGAAAGTTGCCAGATC | GCCTAGGC | | |
| | | CCTCAGCAACGATCCATTCGCTAACGGCTCTTCTTGGCATGCAACTGGAA | TCCAACCA | | |
| | | AGTTCATCCGCTTCTGCGATTGGCGGTTCATTCACCGAGCCAGACTGAA | GGCTGCCA | | |
| | | CTGCCTCCCCACGAATGTAGCCACGAAGAGGTGGAAGGCAAATGCGAA | CTCGAGGC | | |
| | | TGGCAAGAATGGCCATCAACAAGAGACACTGCCTCACGTACTGAATCAT | TCCAACCA | | |
| | | TGCCTACCCAACATGGTCCCCATCAGACGCCGCCATGACAACATCCAAC | GGCTTTCA | | |
| | | AACGACTGGTAACAGCCATCAGACATGTCTTTGTCAATCAACA | 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 NO: 1071) | CAAGCTGT | | |
| | | | 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

| | | | | GGTCTCCG TCGCCACA CGGCCAG GCTTCATC GCACCACT GCTGAACA CACTGACG ACAGC (SEQ ID NO: 1194) | | |
|---|---|---|---|---|---|---|
| NeSL | Utopia-1_TSP | . | Trichinella spiralis | TTTCTGGTATGAATCCCAAGCGGATTCGTTACGAAATTTGCATAAGTTTT | TTTCTGGT | TGATCC |
| | | | | TGAAAAAATAGGCATTTGGTCGAGTGCTCGCACCACCATTGTCGCGGG | ATGAATCC | GTCCCG |
| | | | | TCGTCCTGATATTGCACTACATTCAGGAACGGCCTATTCCCTTCGGGGAA | CAAGCGG | AACCAA |
| | | | | TTGTGTTTTAGGAATTGGAATCGGTTTGGTTACGATCGGTCGAGTGGTT | ATTCGTTA | CGGAAC |
| | | | | CGTGAGATCGAGTGACAGCCGGGTGGCAGCGACAATGTGTAGTGCAAA | CGAAATTT | CACATT |
| | | | | GACACCAGCCTTGAAGTCGGGTCGCCGTCGCGGAAAAGAGGTCAACTA | GCATAAGT | GCCGCA |
| | | | | TGAAGGCCAAATTGTACGGGTAGAAAGACGGCGTGGATCTCGCAGTAG | TTTTGAAA | TGACTTC |
| | | | | TACGAGCGCAACGGACCTTGGTACGAGGATGGTTACGCGTGGGCGAAA | AAATAGGC | GATTTC |
| | | | | GAAGCTGATGGAGGCGTCGGTTCGTGAAGCGGGTCACCACGGAGGTG | ATTTGGTC | GCTTTTG |
| | | | | AATCGGCGAGTACTGTGGACGTCGACGTGGTGGAATCGAAGAAAATTA | GAGTGCTC | CTCTTTT |
| | | | | CGGGTAAAACGGCGCGCCGTAATAGAAGAGCGCCAAGCGGTGATGGA | GCACCACC | TGTATTT |
| | | | | AAAAGACGAGAGAGCTGTGGTGCGGAATGTGGCCAAGCCGTATGCGG | ATTTGTCG | AATTTTG |
| | | | | TAACGCTGTTGCGGACCGTAGTGAGGCCAGTTCGCCCCGCACTCCCAAC | CGGGTCGT | CTATTAA |
| | | | | GTTAGTAAATCTGGTCGTGACAAGTGTGGTCAACCCACTATAAAAGCCT | CCTGATAT | CAATTCA |
| | | | | CCACACCATCACCACCGAAAAGAAAACCGACCACTTCGAGTAGCCCACG | TGCACTAC | GTTTGTT |
| | | | | TACGCCATGCCTCTCCAAGCGCGGAGCGAGATCCAAAATTCCGTCCACG | ATTCAGGA | AACTGTT |
| | | | | CCAGATACACCATCGACTTCTGGCGGTTCAGGTAAGCAACGTGTTTTGG | ACGGCCTA | TTGTATT |
| | | | | TGTCGCCGTTGCTTAGGACGGAAAAATTGCCCGACCTCGAAGTTCTACA | TTCCCTTC | CATTTGA |
| | | | | AAGAACCGAGGAACAGGTGACTGTACGGGCAACTTTCCCAATCGCCCA | GGGGAATT | AGATCC |
| | | | | GGCGGTGGTCTGCCCACTGGGATGCGAGAAACCGTACACGGCCGTGAG | GTGTTTTA | AAATAA |
| | | | | ACCAGACGGTCAATTTGCGCACCAAACGCTGACTCGTCATTTCATGCGA | GGAATTGG | AAC |
| | | | | GTCCACAATTGTCACAGCGTCCAGTGGCATTACCGGTGTCGCAATTGTA | AATCGGTT | (SEQ ID |
| | | | | ATACGGATTTCTTGCCGGCAGATCACCGGTATCCCCTCCGAGTTGTGAA | TGGTTACG | NO: |
| | | | | CACCCACGTTCGCAGCTGTGTTAGCCGGTGGGAGATAACCCGGAAGTT | ATCGGTCG | 1318) |
| | | | | GGGAGAGAGCGAAGATTTGCATGGAGTTCGTTGCGATCTTTGCGATTAC | AGTGGTTC | |
| | | | | GTAGGCGTGTCGAAAAGAGCCGTCGGATTGCATCGCCGCCGGCACGCC | GTGAGATC | |
| | | | | AACGAGAACATTATGCAAAATACGGGTACGGCCGCTCAAATCGAGGCC | GAGTGACA | |
| | | | | TTGAGCAAACAAGTCGGCGAAATACGTGTAGCAGGCGACTATAGTCAG | GCCGGGT | |

| | | | | | |
|---|---|---|---|---|---|
| | | TTTAAGTTTGGTAAACGGGGTTCGGCAATACGTGGCACCAACGCAGAGG | GGCAGCG | | GEPSTATAIGAEEISA |
| | | CGAGACGGGTTGGACGAAGAGGAGGTTCATGAAGCCGAAGAAGAGGA | ACA (SEQ | | TGPVRADTAAQQMI |
| | | GGTGCCAGCCGAATCACGTACCATCCTAGGTGAACCCAGCACCGCAACT | ID NO: | | CRIGQWCVWPQDY |
| | | GCTATAGGAGCTGAGGAGATCTCTGCAACGGGTCCAGTTCGAGCAGAC | 1195) | | HSIPAPQCWTDTLM |
| | | ACAGCTGCTCAGCAGATGATCTGCCGGATCGGGCAATGGTGTGTCTGG | | | DLMIEQIVLQRYPDG |
| | | CCGCAGGACTACCATTCGATCCCGGCACCTCAGTGCTGGACGGACACGT | | | AGVSVMSCSAVSAAI |
| | | TGATGGACCTCATGATTGAACAAATCGTTCTGCAGCGCTATCCCGATGG | | | HHEISAEFAAQVMSS |
| | | TGCGGGTGTGAGCGTGATGAGCTGCTCAGCAGTAAGCGCAGCCATACA | | | HDASLYCIIPVNVRN |
| | | CCACGAGATATCTGCAGAGTTTGCCGCTCAGGTCATGTCTTCACATGAT | | | HWQMIVLDVAERVV |
| | | GCCAGTTTGTATTGCATCATTCCTGTGAATGTACGTAATCATTGGCAGAT | | | HYYCSLREHNTVVLSS |
| | | GATTGTGCTGGATGTGGCTGAACGAGTAGTCCATTACTACTGTTCTCTCC | | | LLSLVELSGKHTGCTS |
| | | GCGAGCATAACACTGTTGTTTTATCGTCTCTGTTATCGCTAGTGGAGCTG | | | WKIETHDGAPVQTN |
| | | AGTGGTAAGCACACACTGGCTGCACCTCGTGGAAGATCGAGACTCACGAT | | | 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 |
| | | GCCCTGGAGCCCCGTGTTACAGCGGGTGCTCAGTGCGGTAGTCGCCAGC | | | RLANWRPIALEPVLQ |
| | | AGAGTGACGAACTGGGCGAGGGCAAACGGGTTGATTTCCTTGGAAGCC | | | RVLSAVVASRVTNW |
| | | CAGAAAGGCTTCCAGCCAGCTGATGGTACCTCCGAGCACAATTTCGTCA | | | ARANGLISLEAQKGF |
| | | TGGAGGTTGCAATTCAAGAGGCAAGACGCACTAATGCACACAGCTTGCCAT | | | 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 | | | | GALGTWDPKNDAIL |
| | | | ACCATCAGCGCTAGGGCATCAGGTGGGAAATGCCTGGAACGGCGGTTC | | | | KRIGVVSQRYLRLMK |
| | | | ACGTCACCCGTCGGGACGCCCTCGCAGCGTGGTGAGCTGCAGTGCCAA | | | | VLVVSEMLEHSSRIYR |
| | | | CCGTGCCCTGGCCCTAGAAGACCCGCCCTCGCTGGAATAGCGCCCAACC | | | | KHLGLRDLLLPDTGTK |
| | | | CGCCTAGTCTCCAGCCCAGAAAACCACCCCCGCGGCAACACCAGAAGCC | | | | RRPVGTTETDPPGGD |
| | | | GGTTACGAAGTCTACAGCGCATTGATCCGTCCCGAACCAACGGAACCAC | | | | LRQKKRNTISARASG |
| | | | ATTGCCGCATGACTTCGATTTCGCTTTTGCTCTTTTTGTATTTAATTTGCT | | | | GKCLERRFTSPVGTPS |
| | | | ATTAACAATTCAGTTTGTTAACTGTTTTGTATTCATTTGAAGATCCAAATA | | | | QRGELQCQPCPGPR |
| | | | AAAC (SEQ ID NO: 1072) | | | | RPALAGIAPNPPSLQ |
| | | | | | | | PRKPPPRQHQKPVTK |
| | | | | | | | STAH (SEQ ID NO: 1440) |
| NeSL | Utopia_2_CMy | . | Chelonia mydas | GGCAGAAACTGCACSTTCTAGAAGACTCACTGCCTATCCTGAGGAAGAC | GGCAGAA | TGAGCT | MLQLRLPTPQTLRLL |
| | | | | TACCGCTTTGGAGATGGATTCTACTGCCGCTGCTTCAAAAGAAAATCTTC | ACTGCACS | GGAGTG | HPSQLPQSHSTKRW |
| | | | | ATGCTGCTTCGGAGGCTCCAGGACAGASTGAGAAGATGATCGCTCCCTC | TTCTAGAA | CCGATG | SNIYEERARAPKDTTI |
| | | | | GCCGATTCCACAGAAACCTTCAACTGCCGCTCGGACCGCTGCTGCTCCA | GACTCACT | AGAAGC | ERVSAASKIPKLDPAK |
| | | | | CGGAGCGCCCATCGGGGAACAGCCTCTAAGAGACCCTCAGCGCTTCCA | GCCTATCC | GCAGTC | RRIGAPLQLMQGNSI |
| | | | | GGACATCGCAGATGAGCAGCATCGATTGGAAAAGCAGCGCCTCCCTGA | TGAGGAA | GGGAAA | SRQLSASSQYVQHN |
| | | | | AGAAAACCCGTGGAGATGCTGCTGCAGGGAWATCTTGCACCTCGAGGA | GACTACCG | GTAACT | AVVRRVSAPPHTGNF |
| | | | | CGGCCTCCCAGGACATCGCAGCCAGGACAAACATCATCTCGCCTGCTCT | CTTTGGAG | GAAATA | TGSCRRKLALPHRPP |
| | | | | TCAGGASAAGGATGCCMSAAGAACTTMTCCCACCTCCTCMACTGCCCA | ATGGATTC | CTTTCCT | SETQPAESYLHCWTT |
| | | | | GGATCCTSATGCTGGTCGTCGTCCTGCTGTGCTGGAAGMKACCCCAACT | TACTGCCG | CATGGA | QRDPTLADQHGLQD |
| | | | | GGCGAGACCTCAGAAACCACCCAGMWGGACCKCAACATCTCACTAGAT | CTGCTTCA | TTGTATT | HPTGLPPEGSHCIKD |
| | | | | GCCTGCCCAGCCGAACCTCTCCATGCWACTCTCCCAGAGCAACSAGAAC | AAAGAAA | TTCTAAA | PRQNSSTEGQRRSG |
| | | | | CATCCAGMGAATCCGCTGATATGACTGAAGCCMATCCAACAGAGGGA | ATCTTCAT | TGGACA | NSQARRIPKRASTAA |
| | | | | GAAGGAAAGGAGAATGACTGCATCTATCTCCAGTATCCCCTCCCTACGG | GCTGCTTC | ACCTACC | SKTVLPKRTSAALKSV |
| | | | | ACACGCTCCTCTKCCCCTTCCGCTATCCGAGGGTTCCAGTACATTGGCAG | GGAGGCTC | TAATTCT | REDTALVLEDPAKWS |
| | | | | TCTCAGCAMACACCTCAAGAGAATCCATASCAAGCGGATCACCTTCCGG | CAGGACA | CAATTAC | SQHREGXRQQANPT |
| | | | | TGTGCCCTCTSCGACCTGCCTTTCGAGACGCAGATGAAATGTAAGTCTCA | GASTGAGA | TGAGGG | AVFQPEPAEIEQQPX |
| | | | | TCAAGTCACCTGCAAAGGACATCTCGAACTGGAAGAGTCCAACTTTACC | AGATGATC | ACAATCT | VRAATPWQAAWME |
| | | | | AGTCTATGTTGCCGCCACACCCCATCTCTGCTCCGAAAGCAGAAACACCACA | GCTCCCTC | CCACTCA | ELARTASFXDFDLLVD |
| | | | | TGCTCCAGCTCCGGCTTCCGACTCCCCAAACACTCCGGCTACTGCATCCC | GCCGATTC | TTGATAT | RLTKDLSAEIVSGRKG |
| | | | | AGCCAACTCCCCCAGAGCCACAGCACGAAGAGGTGGTCCAACATCTATG | CACAGAAA | ATTTTGC | TQENTPTAHRQNQN |
| | | | | AGGAAAGGGCCCGGGCGCCCAAGGATACTACCATTGAAAGAGTAAGTG | CCTTCAAC | TTTCCAC | NMREARRRNISRCYD |
| | | | | CTGCCTCAAAAATCCCTAAGCTGGACCCTGCCAAGAGGAGGATCGGCG | TGCCGCTC | AACCAA | PAAASRIQKLYRSNR |
| | | | | CTCCCTTGCAGCTTATGCAAGGCAACTCCATCTCCAGACAGCTTAGTGCT | GGACCGCT | ATCTCTG | PKAMREILDGPSSYC |
| | | | | TCCTCGCAGTATGTCCAGCACAACGCCGTCGTCAGAAGGGTCAGCGCCC | GCTGCTCC | TACAACT | AIPSERLFLYFKGVFD |
| | | | | CTCCACACACGGGTAACTTCACGGGATCCTGTCGTCGGAAGCTAGCACT | ACGGAGC | TTTCATG | 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 |
| | | TGCGGAGATTGAGCAGCCCWCGGTGAGGGCCGCCACACCATGGC | 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 |
| | | ATCGATGCCCCACCAGCACATTTTTGACATGCTGCGAGAGTTCGGGATG | 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 |
| | | GCGAATTTGGAAGAGACGGGGGGAGACTTTGCTTCGCTCTGGACTCGTG | 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 | |
| | | GCTGAGAGAATGCGGAATCGGTCAACGCTACGCTCGGCTGATGCGGCA | CGCTATCC | |
| | | ACTCATGGTGTCAGATGCCATCAGGTGGTCGAGGGACATCTACATAGAA | GAGGGTTC | |
| | | CACATCACCGGACATCGGCAATACCAGGAGGGATGAGCTGGAGTGCCG | CAGTACAT | |
| | | ATGAGAAGCGCAGTCGGGAAAGTAACTGAAATACTTTCCTCATGGATTG | TGGCAGTC | |
| | | TATTTCTAAATGGACAACCTACCTAATTCTCAATTACTGAGGGACAATC | TCAGCAM | |
| | | TCCACTCATTGATATATATTTTGCTTTCCACACAACCAAATCTCTGTACAACTTT | ACACCTCA | |
| | | TCATGAGTGATGTACCCGAGTACTTGGATTCTAATATCTAAACTGTATTG | AGAGAATC | |
| | | TTAAATCTATTCACCTAAATTTGGGTTATTGCTGATTATGTACTCTATGTA | CATASCAA | |
| | | TCATATGACTTTTAAAAACAAACTTTGTATTTGTGGATAAATCTAAGCACT | GCGGATCA | |
| | | ATACCCAGATGTACAGACACTCTTTTCCCAACCTATGTATTATATTTTTTT | CCTTCCGG | |
| | | AACATTAGCTTTAATAAAATTTTTAAA (SEQ ID NO: 1073) | TGTGCCCT | |
| | | | CTSCGACC | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | TGCCTTTC<br>GAGACGC<br>AGATGAAA<br>TGTAAGTC<br>TCATCAAG<br>TCACCTGC<br>AAAGGAC<br>ATCTCGAA<br>CTGGAAGA<br>GTCCAACT<br>TTACCAGT<br>CTATGTTG<br>CCGCCACC<br>CCATCTCT<br>GCTCCGAA<br>AGCAGAA<br>ACACCAC<br>(SEQ ID NO: 1196) | TAAGCT<br>GGTCAT<br>TTGACC<br>GACAGG<br>GCACTA<br>CCCAGG<br>TAGGGA<br>ACCGCC<br>CTTCAG<br>AAACCC<br>AGGAAG<br>ACACAA<br>ACACCCT<br>CCCTTTA<br>GTGACA<br>TACATAT<br>TTTAGG<br>CTACATT<br>TCAGTTA | MQDMEEELLLDVEM<br>ETETTEPQTSTAXDA<br>TTTTDRPTRWGPHP<br>RAVAAAAIAQLVTGE<br>XAXPALPSRQDRRPA<br>PRSHAPTRSRWGPR<br>HQAVGAAAIASLATG<br>LPASAAPVSRATKHG<br>EGRRRLQTRWGPRV<br>SIPRAARRPGSRWGP<br>PRAAGASGQLPASAS<br>GATGQLPEHVEAITT<br>TPRVASDADEGPTPP<br>DPWILRFDGACRRN<br>PGPGGAGAALFKPS<br>GAVVWTCSHYMPSS<br>NETNNTAEYTALLLG<br>VQSAVHHGATRLDIE<br>GDSSLVIAQVKGTFA |
| NeSL | Utopia-2_PCa | . | Phytophthora capsici | ACCGCCCAAGTCTCCACCGCAGCTGCGACTGCTGCTCATCGAGCCG<br>CAGTAGTCGCAGCTGTACAAGCACCACCTCCTACCGGACGCCGCCGTCGT<br>GGAGCACCACGCGCGCTGAGCCGCNTCAAGACCAAGAATTCCAGGC<br>TCGCGCGCGCGTGGCGCTTCAAGATGGCGACGCGACCAGCAGGAGTGG<br>CCGCTCTACTGACGGCGGAGACGAGGCCAAGACCGAGGACAGCGACGC<br>GGACGCAGCCAAGCGCCAAGCCGCTCAAGCAGCACCGAGACCCCATAC<br>CCMGGACACCGCGCCGTACGACGGCCACAGGGACTCGGTGCTGGCTGT<br>GTACGCACACAACGCACCTGCATTCACCTGCGCGTTGTGTGTGTACACA<br>GCACGCAACTTCGCCGAGCTGACCAAGCATCGCCATGCAGCGCATCGTC<br>ACACCCGCTTCGTGGATCACTTCCACAGCGGGTGCACGTGTGGCATCGG<br>CTTCCAGTCGCGCGCGGCAGCTACGCGACACGCTCAAGCCTGTGCAGAC<br>AGCACACACGCCACCGTAGCTGCCTCGCGCGACCCGGCMNCCASTGCSS<br>CCSGNGCCGGMAGCGASGAGGAGGACCNCGCACCCCCCGGTCCTCTSS<br>TCGCGGCAGCATCTGCAGCTGCAGCAGCCGCATCAAGCGCCACTCCAGC<br>TGCAGATACTGCCACCACACGCAGAGCGCCGTGCCCATCGCTACTCCTGGG<br>CCCCAGTACGCCCCCCCACGTGCTGGAGCCACCTCCAGAGCTCCGAGTTT<br>CCGGCAAACGCCGMCGCCTCAACACGCCGATCGACCTGCAGCCGCTGG<br>ACGTGGACGCGCTGATGCAGGACATGGAGGAGGAGCTGCTCCTGGAC<br>GTGGAGATGGAGACGGAGACGACGGAACCGCAGACCTCGACGGCCAM | ACCGCCCA<br>AGTCTCCA<br>CCGCAGCT<br>GCGACTGC<br>TGCTGCTC<br>ATCGAGCC<br>GCAGTAGT<br>CGCAGCTG<br>TACAAGCA<br>CCACCTCC<br>TACCGGAC<br>GCGCCGTC<br>GTGGAGC<br>ACCACGCG<br>CGCGCTGA<br>GCCGCNTC<br>AAGACCAA<br>GAATTCCA<br>GGCTCGCG | | |

| | | | | | |
|---|---|---|---|---|---|
| | | GGACGCCACCACCACCACGGACAGGCCAACACGATGGGGCCCCCACCC | CGCGCGTG | CGGAGA | CRNAKLRQLRNRVR |
| | | CAGAGCAGTGGCGGCTGCAGCCATMGCGCAGCTGGTGACCGGCGAGS | GCGCTTCA | GGTTAC | HALRSVEKYTLRHIDR |
| | | CAGCAGMTCCAGCTCTGCCTTCGAGACAGGACCGCCGCCCAGCACCAC | AGATGGC | TAACTG | KANAHADRLANRAL |
| | | GATCCCACGCACCCACCCGGTCTCGATGGGGACCGCGCCATCAAGCTGT | GACGCGAC | GTAAAT | DRRSSSSECEPHGSC |
| | | CGGTGCCGCAGCCATCGCGTCGCTAGCTACAGGACTACCGGCCTCTGCG | CAGCAGG | AAAAAG | MERCCGTDTTPAVQ |
| | | GCACCCGTGTCTCGCGCAACCAAGCACGGGGAAGGACGACGCCGTCTC | AGTGGCCG | CACTTT | GPTPQAAAAVPVQV |
| | | CAGACACGATGGGGACCTCGGGTGAGCATCCCACGAGCCGCACGCCGT | 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 |
| | | AGCGGTGCAAGGACCGACACCGCAAGCAGCAGCAGCAGTGCCAGTGCA | 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 |
| | | GCGTTCGCCAAGGACGAGTCCAAGTGTGTGTCCAAGATATTGGCCGGT | 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 |
| | | CGACAAACGCCCCGCGGCAACCCGGTCCGCGACGACGGCGTGCACCTAG | SCCSGNGC | | QVARKRYRLQEEHW |
| | | AGCTGCAAGGCGAAGTCATCGCCCCCCTCAGTCTCCACGATAGCTATCG | CGGMAGC | | RGRGDELVRLFLNSE |
| | | CTACCTGGGTGTGGGCGATGGCTTCGACCATGTGCGCCACCGCCTCCAG | GASGAGG | | LAASPFADCLRRNGD |
| | | CTTGAACCCAAGCTCCAACAGATCAAGCGAGAGGCGGTGGCGCTGATG | AGGACCN | | IGSLWTDVQRTLRLH |
| | | CAGTCCGGACTGGCGGGGGTGGCAGGTGGTGAAGGCGCTCAAGACCTTC | CGCACCCC | | HLSLTAQDDRDGQD |
| | | GTGTACCCCAAGGTCGAGTACGCGCTGCGCCACCTACGACCACTCCAGT | CCGGTCCT | | PLALRVPHHTKWLD |
| | | CCCAGCTCCAAGGCTTCGACCGCGCCGTCGTCCGTGGCCTGCGCCACCT | CTSSTCGC | | HKTVLRHVKLHMKIR |
| | | ACTGCGCCTCCCCCAGTCCGCCACCACCGAGTTCTTCTACACACCAACGT | GGCAGCAT | | HQTRWKGLVDQGKT |
| | | CTGGAGGTGGGCTGGGGCTGCAGTCCCTCGTGGAGATGCATCAGGCGC | CTGCAGCT | | VRVHGGLGAKFVST |
| | | TGCAGGTGGCGCATGCCTGGCAGATGCTCCATTCCAAGGACGCGGCCG | GCAGCAGC | | GAGLSDDAYRFGVK |
| | | TCGTGGCCGTTGCGAAGGAGCAGGTCTGCCAGGTAGCACGCAAGCGCT | CGCATCAA | | ARLNQVDTNAVLKR |
| | | ACCGCCTGCAGGAAGAACATTGGCGCGGGGCGTGGCGACGAGCTCGTGA | 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 | . | Phytophthora infestans | GCTCGGCCTCGCGGCTGCCTTCCCAGGCGCCGCCGACTTCGCGCTCTGG | GCTCGGCC | TAGGCG | MLADPAALAAGLAR |
| | | | | CGCGGCCCACACGCCGCCGCCGAGCCTCCAAGCGCGCCCGTTGGCTTCC | TCGCGGCT | GAAACC | APPPPSAPQDPSPAF |
| | | | | GCAGACGCAGGGCTCGCGGCGACGGCCCTCGCCAGCCCCCAAGACCCC | GCCTTCCC | AGGCCC | PAGPAGQNPRAAAP |
| | | | | CCCTACGATGTGGCACCACCCGGCAGGGCGGCCGGCAGGCTGCCCGAC | AGGCGCC | AAGACG | ARVEVHTVVAPPGR |
| | | | | TCGGTATCGCCGGGTGCTACACTCTCAGCCGCCACAGCTCGGGCCTTGG | GCCGACTT | GCCGAC | AGGMLPDPGLVDSS |
| | | | | CGGTCCGCCATTGGCCCTTGGAGCTCGACAGCGACAGCAGCGACGACG | CGCGCTCT | AGGGCC | PAAATAATPAPVAAT |
| | | | | AGGACGCTCAAGACCCCCACGCCGCCGCCCCAGGACCCCCCCGCAGACGT | GGCGCGG | CCACCC | ATTARAAARVAVEH |
| | | | | CGCGAGTGTGCTTGCCCCACCCGGCAAGGCAGGCAGCATGCTCGCCGA | 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 |
| | ACCGACAACCGCTGGGGCGCCCCGCTTCCGCGCGTGCTGGTCGCGTCTC | GGCCTTGG | TTCTAAT | ACEEEAKDSDTMDW |
| | GAATCGCCGGGGCGCCTCGCTCAGGTGCCACCCCCACGCTGGGGCCCCCC | CGGTCCGC | CAGTGT | APTWTNEETKESEPH |
| | GCTGCCGCGCACCACCATCGCAGCCCGCATTGCGACGAGGCTCGCAGCC | CATTGGCC | GAAAAC | DEAPGQVDEETIDDA |
| | ACGCCTGCCCCCCGCTGGGACCCACCGCTGCCACGAAGCCTGGTTGTGT | CTTGGAGC | TGGTTTT | DGEWLLRFDGACRA |
| | CACGCATCGCCGCCCCGCCTGCTACCGCGCGCTACCAGACGCGCCTGCATG | TCGACAGC | CGCCTTT | NPGPGGAGAALFKP |
| | CGAGGAAGAGGCCAAGGACAGCGACACGATGGACTGGGCGCCAACTT | 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 |
| | CAGCAGCAGCAGCGGCAACGCCCACGGCAGCAGCAGCAAACACG | | TTTCTAA | VTRHHREHRIDEALD |
| | GCAGCGTCGACAGCGGAGATGCAAGCGCGGCGGTGGCAGCCAACACA | | TTCGCCT | DLHALESRRPQDRTA |
| | GGCAGCGGAAGACGCGCCGCCGACGGCCCCCACGGGTGACACGCCACC | | CCGTCTT | ISKARRRVGRIRSALD |
| | 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 |
| | | GCAACGTGTACGGGGGTACGTCCAGCCCCGGCTTGGACGGAGTGGGGTA | 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 |
| | | TGCCGCACCACGACACAAGTGGCTGACGCACACAAGGACGTGCTGCGGCACG | | 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 |
| | | GCAGCCGGGCGACGGAGACCGGGGGGACTCCGTCCCAGACAAGCCGC | | AAGCGG | LKRDAKRLDRQLSVE |
| | | CGCTAGGCGGAAACCAGGCCCAAGACGGCCGACAGGGCCCCACCCAGG | | CATCGC | CIQSSRRIWNLHCSQ |
| | | TAGGGAACCGCCCTAGAAACCCATTTCGGTGGTCGACTCGAAGGCCTTA | | GGGACG | HRTRQHQARPSQGP |
| | | CCTATTTTTCCTTAGACATTCAATTAGGTAGCGACCAAATTACAAATTTG | | ACGCCT | RGSRATETGGTPSQT |
| | | GTAACGAGTAAGCCAAATGGTAATACACAAAACTTTTCTGTTCTAATCAG | | CGGCCG | SRR (SEQ ID NO: |
| | | TGTGAAAACTGGTTTTCGCCTTTTGGCGGACTTTTTCACTCGCATTTTTG | | CTCAAG | 1443) |
| | | GGCAATCGTCTGCGGCTAGCTTGCTAGCGGCGGACGAGCGGTCTCCGG | | CGCAGC | |
| | | GGGCGTTCACCTTTCCCCCGCGAGGCCAACTACACCGATCTTCTCTACAC | | CCCGCC | |
| | | TTTTCTAATTCGCCTCCGTCTTCGGCTGTCGGATTTTTTTCTTTT | | CTTCCAG | |
| | | TGACCAATCAGAGCGCGCCATGCGACTCTTCTGGCCAATCAGAGACCGG | | CACGAC | |
| | | GCCCTGTCCTCGGACAGCGAGGCCTCCACGGCCAGCCAATCAAGTCTCG | | CTCGCG | |
| | | GCAGCGACGCGTCTTCTATAGCGCAGCTGACGAGGCCGATCTGGCGG | | CCGTTT | |
| | | CCCCCGATTGGTCCGACTTTCGGCCAATCAGCGACGACGAGGGGGGCAG | | GGCGGA | |
| | | GGGTTTACACTTTTGCCCCCGTTTCGACTTCAACTTCAGGCCAAAATGGC | | TCGCCAT | |
| | | GATTTCGACCCTCCACGCGCCGTGCCACTGCTCGGCACCGGCGGCGATT | | CAAGAC | |
| | | CAGCGGGTGCAACTTCGGGCACGTGTGCAACACATGCAGCGCCCCATTG | | GTGCGA | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NeSL | Utopia-2_PR | . | Phytophthora ramorum | TCAAGCCCCGCCGCCAAGCCCAGCTGCGGCTGTTGCCGCCCCTCCAGCA<br>GCAGCAGTCGCAGCTGAACCTACAGCCCCTCCTGCTGATCGCGCCGCCG<br>TGGAGCCCCGCGCGCCGAGCCGCCCCAAGAACAAGCACCCCCAG<br>CTAGCGCGCGCGTGGAGCCCATGTCGGGCGACGTGGTCAGCAGCGACG<br>GCAGCAGCAGGACGACGGACGCGAGCGGCGACGGCGACGACGGCGCA<br>GGCAGCAGCGATGCAGCAGGCGACGTCGGCGTCGTCGCCATGGACGTT<br>GACCAAGGTGCCCGCCGACAGCAGCCGCCATGGCAGCGCGTCGGCGGC | CACGCCAAGCGGCATCGCGGGACGACGCCTCGGCCGCTCAAGCGCAGC<br>CCCGCCCTTCCAGCACGACCTCGCGCCGTTGGCGGATCGCCATCAAGA<br>CGTGCGAGAGCCAGGCGGGTCGGGCAAAATATACTTACTCTAAGTAT<br>GCCCGAATCCCTGCCCTCTCAGGCTGAACGCGGCCCCATACTTGATCTAA<br>GTATGGGAGGGATCCCTGGCCTCTCAGGCTGTACGCGAGACCCGTACG<br>GCCGAATCCCCTGGCCTCTCAGCCTGTACGCGGGGC (SEQ ID NO: 1075) | TCAAGCCC<br>CGCCGCCA<br>AGCCCAGC<br>TGCGGCTG<br>TTGCCGCC<br>CCTCCAGC<br>AGCAGCA | GAGCCA<br>GGCGGG<br>GTCGGG<br>CAAAAT<br>ATACTTA<br>CTCTAA<br>GTATGC<br>CCGAAT<br>CCCTGCC<br>CTCTCAG<br>GCTGAA<br>CGCGGC<br>CCCATAC<br>TTGATCT<br>AAGTAT<br>GGGAGG<br>GATCCCT<br>GGCCTC<br>TCAGGC<br>TGTACG<br>CGAGAC<br>CCGTAC<br>GGCCGA<br>ATCCCCT<br>GGCCTC<br>TCAGCCT<br>GTACGC<br>GGGGC (SEQ ID NO: 1321) | TGAGCA<br>CCTTGG<br>GTTGCT<br>CAAGCG<br>TGATGC<br>GAAGCG<br>GCTGGA | MSGDVVSSDGSSRT<br>TDASGDGDDGAGSS<br>DAAGDVGVVAMDV<br>DQGARRQQPPWQR<br>VGGKRRRLNDVDDE<br>DTRELAELLLEEEDEA<br>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 |
| | GCCCCGACTCGTTGGGACCCGCCCCTCCCCCGCACGCTGGTGGCGACAC | 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 |
| | CGGCCAGCCGAGCACCAGCCGCAGCCGTCGCCGTCGCCGCCGCCAA | | 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 |
| | | CCCCGCCGCGACATCGCAGCCTGAGTTGCTCTCGGACGCGCCGGACCACG | TACCAG | MAPVGAPFRSAVAH |
| | | GACGACATCGAGGACCAACTGCAGCGTGCCCGTGGCAGCTCCAGCCCT | GTATTTA | LPAATSQPELLSDAPT |
| | | GGCCTGGACGGCGTGGGGTACGACATTTACAAGGCGTTCGCGGCCCAG | CTACCTG | TDDIEDQLQRARGSS |
| | | CTCCTGCCGGCCCTGCACGCCAGCATTCGCGTGCTGCTGGCGCCACAAGC | GAAGCC | SPGLDGVGYDIYKAF |
| | | AAGTCCCGCAGAGCTGGAAGGTGGGGGTGGTGCCGCCTCCTGTTCAAGA | AAGGAT | AAQLLPALHAAFACC |
| | | AGGGGGAGCGAACCGAGCCGGCGAACTGGCGCCCGATCTGCCTGCAG | TAACCG | WRHKQVPQSWKVG |
| | | CAGGCAATCTACAAGCTGTACGCCGGGGGTGCTAGCTCGCCGACTCACCC | 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 |
| | | CGGGCTGGGCCTGCTGCCCCTGACGGAGCTGCACGCGGCGCTACAAGT | | AVKVYLYPRVEYALR |
| | | CGCGCACGGATGGCAGATGCTGAACTCGAAGGACCCTGCCATCCGACG | | HLRPFHQQLEGFDR |
| | | CATCGCGCGCGTGCAGCTCCGGCAAATCGCCGACGCGCGCCACCGGAT | | HLVRGLRHLLRLPAN |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CGACGCTCAAGCGTGGCAAGACCGGGAGGAAGAGCTGGCGCAGCTCCT | | | | ATTAFFYAPVSRGGL |
| | | | CCTGAACTCGCAGCTCGGGGCGTCCACCGGCGCACCCCCGAAGCGCCG | | | | GLLPLTELHAALQVA |
| | | | CAACGGCGACATCGGCTCGCTCTGGGTGGACGTCCAGCGCCACCTCCGC | | | | HGWQMLNSKDPAIR |
| | | | CATCTCAGCCTCAAGCTGGAGACCGCACCGGCGTGTGCGGAGACCGGC | | | | RIARVQLRQIADARH |
| | | | ACCGCCGCGGCCATGCTGCAGCTTCGCGTGCCGCACCACGACACAAGTGG | | | | RIDAQAWQDREEEL |
| | | | CTAGACCACAAGACTGTGCTGCGGCACGTGAAGCTGCACTACAAGAAC | | | | AQLLLNSQLGASTGA |
| | | | AAGCACTGGGCCAGGTGGGCGGCCATGMGSGACCAAGGCAAGACGGC | | | | PPKRRNGDIGSLWV |
| | | | ACGCACGCATGGGGGCGCGGGGAGCGGCTTCCTCACCCGGCCACGAG | | | | DVQRHLRHLSLKLET |
| | | | GCATGTGGGAAGCCGACTACCGCTTCGCGGTGGCCCGCCCGCCTGAACC | | | | APACAETGTAAAML |
| | | | AGCTCGACACGCACAGCGGTSCTGAAGCGCCGGCGCCTCCGGSMSCACG | | | | QLRVPHHDKWLDHK |
| | | | ACCGCTGTCGACAGCCGGGCTGCACCCAGGGSGGAGACGCTGGCGCAC | | | | TVLRHVKLHYKNKH |
| | | | GTGCTCAACCACTGCGCCGGCACCATGGACGCGGTCCGCGGCCGCCAC | | | | WARWAAAMXDQGK |
| | | | GACGACGCCCTCAAGCACATCGAGCGCGCGCTGCACGCGTCGTCGCCG | | | | TARTHGGAGSGFLTR |
| | | | GGCGGCCAGGACCGCGTCGAGCTCCGGGTCAACCAGACGACGGTGCCGTCG | | | | PRGMWEADYRFAV |
| | | | CTCGCCGGCCCCGCGCTACGGCCCGACCTCCAGCTGTACAACCACACCA | | | | AARLNQLDTHSVLKR |
| | | | AGAAGACGGTGGCGGTGGTCGACCTGGCCGTGGCGTTCGAGGAGCAG | | | | RRLRXHDRCRQPGCT |
| | | | GCGAGCGACGACCCGGAGAGCTCGGCGCTGGCACGCATCGCCGCACAC | | | | QGGDAGARAQPLRR |
| | | | AAGCGCGCGAAGTATGCCGGCGTCAAGCGACACCTCGAGCGCCAAGGG | | | | HHGRGPRPPRRRPQ |
| | | | TGGAAGGTCCACCTCTCGGCGCTCGTGTACGGCTCGCTCGGCGCGGTGC | | | | AHRARAARVVAGRP |
| | | | CGGCCGGCAACCACACAAGGTGCTTACTGAGCACCTTGGGTTGCTCAAGCG | | | | GPRRAPGQPDGAVA |
| | | | TGATGCGAAGCGGCTGGACCGGCAGCTCTCGGTGGCGTGCATCCAGTC | | | | RRPRATARPPAVQP |
| | | | CAGCCGCCGCATCTGGAACCTGCACTGCAGCCAGCACCGCGCGCGCCA | | | | HQEDGGGGRPGRG |
| | | | GCACCAAGCACCAGGGGGAAGTCGGGCGGCGGAGACCGGGGGGGACTC | | | | VRGAGERRPGELGA |
| | | | CGCCGCGCACCGGCCGCCGCTAGACGGCACACAGGCCCACAGCGGCCG | | | | GTHRRTQAREVCRR |
| | | | ACAGGGCCACACCCAGGTAGGGAACCGCCCTCAAACCCCGCCGGTACA | | | | QATPRAPRVEGPPLG |
| | | | TTATGGTCCGACACCTATGAGGTGCAACCTGTACACAAGTTACACACCA | | | | ARVRLARRGAGRQP |
| | | | CATAGCGACTACCAGGTATTTACTACCTGGAAGCCAAGGATTAACCGGT | | | | QGAY (SEQ ID NO: 1444) |
| | | | CGGTAATACACATAACTTT (SEQ ID NO: 1076) | | | | |
| NeSL | Utopia-2_PU | . | Pythium ultimum | TGACTGGTGTTTGATCACGATCAATGAGGTGATTAACATGAGCCGGAGC | TGACTGGT | TAAGCG | MDYDDSEFFDAICIP |
| | | | | AGGCCCCTTACACGCTGGTGCTGTAATGGTTCAGGAATGCTCTTATGAG | GTTTGATC | GGGGGT | DEDADVLDDGDEGD |
| | | | | TAACTCCACAGTATAATTTTTGTTAGCGGGGGTGAGCGCGTGCGCCGCC | ACGATCAA | CCAGAC | EGGNDDESSEPLPLAI |
| | | | | CCCACCTTTCTCTTTTCGTTTGTATTTGGTTCCTCTGGAGCGACTTCCGTG | TGAGGTGA | CCCACA | TNAPSAPLHATMLC |
| | | | | GCTTTTGTCGCGCGCTATCGCGCCCGGCATGAGGATCGTCGCTCCTGAC | TTAACATG | AGAGAG | GTVTQPWLLRFDGA |
| | | | | GCCATGTGCGCCAGTATATCGGAGCCCCCGTGCTCCGCTTGGGTGTGTGT | 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 |
| | CGGCAACTTGGACGACCCACGTGCACGCGTGCCGCCAGTGCCGA | 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 |
| | | GACGGCACCGGGACTCGACGGCATCACGAACGCTGTGTACGTGCGCTT | 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 |
| | | CTCTCGCGTTTCGATTATGTACTCCGACACTTGCGCCCGTTCCTTAGTCTC | 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 |
| | | CTCGTCAGTCACACTGACAAGCAGCTTGCACCTCAACAATGGGCTAGTG | 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 |
| | | | ATCAAGCAATTTCTGGAATACAAAGGCTACACTGTGCATGTAGTAGCAGCGC | CAACCTCC | | VPVSTPTLTPTPTPTR |
| | | | TAGTCTATGGTTCACTAGGCTCCGTTGACACCGGCAACTTTGCCGTGTAT | AGACCCGC | | RPKPAPSSTQPQQQG |
| | | | ACGGAACGTCTCGGTCTCCGCAAGGGCGCAGTGCGGCGACTGGAGTGC | GTGCCAAC | | APAQRRQQREQKKQ |
| | | | TCGCTCTCGGCGCGCGCCACACATCAACTTCGCGCACCGTATGTGGCGACGCC | 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) |
| | | | CGGCAGTGAGAGAGACACCGCCAGCAGCTCCGACACAGCAGCTCGGACAG | (SEQ ID | | |
| | | | CAACTCCGACAGCGAGACCGACAGCGACAACGAGATCAACATCGACAA | NO: 1200) | | |
| | | | CGAGATCAACAGCGACAGCGACATCGACAACGAGATCATCAGCACCGA | | | |
| | | | CATCGAGACCATCAGCACCGAGACCGAGATCATCAGCACCGAGACCGA | | | |
| | | | GATCATCAGCACCGACAACGAGATCAGCAGCTCCGACAACGAGATCAG | | | |
| | | | CTGCACCGACAGCGATAGCGACAACCAGCGTTTACAAAAGTCGACGCAC | | | |
| | | | GACCAACGCCATCTCTGGCGCAACGACGCGACGCAGTGCGTCATCCAA | | | |
| | | | GCGTGCGCCAATGCAGCCCCGAACTGCATTGACGCCAACACAACAACAA | | | |
| | | | CAACGACAGTAAGCGGGGGGTCCAGACCCCACAAGAGAGAAGCAGGA | | | |
| | | | ATCATGGTCCGCATGGACCAGTAGGGCACGCTCCACAAAGGTTATCGCC | | | |
| | | | CTCAAACCCATCACACGAAGGATCTAAAAAGAAAGCAACAATCGAAATA | | | |
| | | | GTAAATAGTAAATAGCTTAGAAAGGTCAACATGCGAAATGCATGAGGA | | | |
| | | | ACGTAAGATGGTAATACATTTTTATA (SEQ ID NO: 1077) | | | |
| NeS | Utopia | . | GTACGGCCGAATCCCTGGCCTCTCAGCCTGTACGCGGGGCTATACTTGG | GTACGGCC | TAGACG | MPRSLASEPVHSSAS |
| L | -3_PI | Phytop | TCTAAGTATGCCCCGGTCGCTGGCCTCTGAGCCCGTACACAGCTCGGCC | GAATCCCT | GCAAAG | RLPSQAPPTSRSGAA |
| | | hthora | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 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 |
| | | | AAGACCCCCACGCCGCCCCCAGGACCCCCCCGCAGACGTCGCGAGTG | NO: 1201) | TAGGGT | GPPADVASVLAPPGR |
| | | | TGCTTGCCCCACCCGGCAGGGCAGGCATGCTCGCCGACCCAGCAG | | AACGCC | AGSMLADPAALAAG |
| | | | CGCTGGCAGCCGGTCTGGCTCGCGCTCCGCCGCCCCCATCGGCACCCCA | | CTCGGG | LARAPPPPSAPQDPS |
| | | | GGACCCAAGCCCCCGCCCTCCCTGCTGGCCCCCGCCGGCCAGAACCCTCGC | | AAAACC | PALPAGPAGQNPRA |
| | | | GCCGCCGCCCCCAGCCCGTGTTGAAGTCCACACAGTCGTGGCACCACCCG | | ATTCTG | AAPARVEVHTVVAPP |
| | | | GCAGGGCGGGCGGCATGCTGCCCGACCCGGGGTTGGTTGACTCAAGCC | | GTGTTT | GRAGGMLPDPGLVD |
| | | | CAGCAGCGGCTACCGCAGCTACTCCAGCTCCGGTAGCTGCTACTGCTAC | | GGCTGT | SSPAAATAATPAPVA |
| | | | AACTGCTCGCGTCGCCGTGGAGCACCACGCGCACGCTGAGCCGAACCA | | TTTTCAA | ATATTARVAVEHHA |
| | | | AGAACATCTACCGATGGCTCGCGTGCTCTGTGGAGCCCATGCAGGTCGAC | | 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 |
| | | | TCAGGTGCCACCCCCACGCTGGGGCCCCCCGCTACCGCGCACACCATC | | ACAAAA | QATETEATENRWGT |
| | | | GCGGGCCGCATTGCGACGAGGCTCGCCGCCACACCTGCCCCCCGCTGG | | CCTTCTC | PLPRVLVASRIAGRLA |
| | | | AGCCCACCGCTTCCACGAAGCCTGGTTGCGTCACGCATCGCCGGCCGCC | | TGTTCTA | QVPPPRWGPPLPRT |
| | | | TGCTACCCGCTCTGCCAGACGCGCCTGCATGCGAGGACGAGGCCAAGG | | ATCAGT | TIAGRIATRLAATPAP |
| | | | ACAGCGACGAAATGGACTGGGAGGCGAGCCGACCGTCGAAGCG | | GTGAAA | RWSPPLPRSLVASRIA |
| | | | CCAGGGCCGGTCGACGAGGAGACCATCGACGCCGACGGCGAGTG | | ACTGGT | GRLLPALPDAPACED |
| | | | GCTCCTGCGCTTTGACGGTGCCTGCCGTGCCAACCCAGGCCCAGGCGGC | | TTTCGCC | EAKDSDEMDWEASE |

| | | | | |
|---|---|---|---|---|
| | | GCCGGCGCAGCTCTCTTCAAGCCAAGCGGTCCCGTAGTGTGGACTTGTT | TTTTGGC | PHVEAPGPVDEETID |
| | | CCCACTACGACCCGAGCACCACGGCAACGAACAACACAGCAGAGTACA | GGACTT | DADGEWLLRFDGAC |
| | | CTGCGTTGCTTCTCGGGGCCAGGGCCGCCGCCGACCACGGCGTCACGA | TTTCACT | RANPGPGGAGAALF |
| | | AGCTGCGCGTCGAGGGCGACAGCACGCTGGTCATCCAGCAAGTGCGGG | CGCATTT | KPSGPVVWTCSHYD |
| | | GCATCTTCGCGACGAGAAGCACGCGGCTCCGAGCACTGCGCAACAAGG | TTGGGC | PSTTATNNTAEYTALL |
| | | TCAAGCTCGAGCTCGCGCGCGTGGGGAGCTTCTCGCTCCACCACATCGA | AATCGT | LGARAAADHGVTKL |
| | | CCGGCAGGCGAACGGTCACGCCGACCGCCTTGCCAACGCCGGCCTCGA | CTGCGG | RVEGDSTLVIQQVRG |
| | | CCGACGCCGCACCCAGCTCGAGTGCAGTGTGCACCCCGATGGTAGGGG | CTAGCTT | IFATRSTRLRALRNKV |
| | | CTGCACCAACACCTCCGTAGCGACCGCAGCGCCTACAGCTTCTGCAGCA | GCTAGC | KLELARVGSFSLHHID |
| | | CCCTCGACCCCAACCCGGCCTCCTGCAACTACGGCCGCTCCGTTTCACAG | GGCGGA | RQANGHADRLANAG |
| | | CGACCAAGGCCACATCGACGAGGACGACGAACGTCGTGCCGACATCGA | CGAGCG | LDRRRTQLECSVHPD |
| | | CGACGGTGAAATCTACGCCCCAATGACGCTCGGGCCAGACGAAGTGCC | GTCTCC | GRGCTNTSVATAAPT |
| | | TGCTCGGCGACCGCGCCTCCGACTCCGCCAACTCAGCGACGAAGAACTC | GGGGGC | ASAAPSTPTRPPATT |
| | | GAGGCCGCTGGTGCCATCGTGGAAAGGCTCAGCGCATCGTTGTCAGCC | GTTCACC | AAPFHSDQGHIDED |
| | | AAGATCACCGACGCAGAAGACTGGGGGACGGCAGAAGGCTACATTACG | TTTCCCC | DERRADIDDGEIYAP |
| | | GCCCTGCCTCACCTGCTGTACGACAAGCTCCTACCGTACTCGCGGACAG | CGCGAG | MTLGPDEVPARRPRL |
| | | CGCCCCGTCATCAGCGGCCACCCCGGCCATCGCGCAACCAGCAAGACCA | GCCAAC | RLRQLSDEELEAAGAI |
| | | CCCGCAGCCGCGACGCGACCAACCGCAGCGCAACGTGGACGAACAGCA | TACACC | VERLSASLSAKITDAE |
| | | GCATGCGGAGTCACAGCAGGGTGAAGACCAGCGACAGCAGCCGCGC | GATCTTC | DWGTAEGYITALPHL |
| | | CGACGCGCCGCCGACGTCGGCGAGGGAAGCGTAGAGGCCGCCGACAG | TCTACAC | LYDKLLPYSRTAPRH |
| | | CGACGACACCCTCGCCAGCCAGGACAGAGTACCAGAGCCTCCCAGCAA | TTTTCTA | QRPPPRPSRNQQDHP |
| | | AGCCGGCAGCGCCGACCCCGACCCCGCGCGTTACCCGCCACCACCGCG | ATTCGCC | QPRRDQPQRNVDEQ |
| | | AGCACCGCATCGACGAGGCGTTGGATGAGCTCCATACACTGGAGCGGG | TCCGTCT | QHAESQQGEDQRQ |
| | | CGCGGCCACAGGACCGATCGGCGATTGACAAGGCGAGACGGCGCGTC | TCGGTC | QQPPTRRRRRRGKR |
| | | GGACGTGTCCGCGGGGGCCATCAACCAACACCTGCTGCGCCACCGGTTTG | TTCGGTT | RGRRQRRHPRQPGQ |
| | | ACACCGACGAGAAGGCGTGCGTGGCGGACATCTTGGAGAAAGCGCAC | GTCGGG | STRASQQSRQRRPRP |
| | | GCGGCGCGAGCGGCACGTACGGCACAGGCGGCGGGCGCAGCAACGTC | CTTTTTT | PRVTRHHREHRIDEA |
| | | TACAGGAGGAGCAGCGACAGCACCAACACAGCAAGCAGCCACGTCAGC | CTTTTTG | LDELHTLERARPQDR |
| | | CCTTGGTGATGCCGACGACGGAACCTGCCCCATCCTGGCCGACGAGCTC | ACCAAT | SAIDKARRRVGRVRG |
| | | TGGCAGTACTTCACCGGCACCAACACGCCGCGCTGGGAGTTCAACCCCG | CAGAGC | AINQHLLRHRFDTDE |
| | | CGACACCCGTCGGCGAGGCGTTCCGTACGGCGATGGCAAGGCTTCCTCC | GCGCCA | KACVADILEKAHAAR |
| | | GGCGACGAGGCTCCGCGAACTGCTGACAGAAGCCCCGACTGCGGACGA | TGCGAC | AARTAQAAGAATST |
| | | AATCGAGACCCAGCTACAGCACGTGCGGGGATCCTCGAGCCCAGGACT | TCTTCTG | GGAATAPTQQAATS |
| | | CGATGGCATCGGCTACGACGTCTACCAGCGCTTCGCACAGCAACTGCTC | GCCAAT | ALGDADDGTCPILAD |
| | | CCGGTGCTCACCGCTTCGTTCAAGCGATGCTGGACAGCCAAGATGGTGC | CAGAGA | ELWQYFTGTNTPRW |
| | | CACAGAGCTGGCAGGTGGGGGTCGTGCGACTGCTGTACAAGAAGGGC | CCGGGC | EFNPATPVGEAFRTA |
| | | GCGCATGACGACCCGGCAAACTGGCGGCCCATTTGCCTCCAGCAAGCCA | CCTGTCC | MARLPPATRLRELLT |
| | | TCTACAAGCTTTACACAGGTGTACTTGCTCGGCGACTCGTGCGGTGGCT | TCGGAC | EAPTADEIETQLQHV |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | AGCGAG | RGSSSPGLDGIGYDV |
| | | | | | GCCTCC | YQRFAQQLLPVLTAS |
| | | | | | ACGGCC | FKRCWTAKMVPQS |
| | | | | | AGCCAA | WQVGVVRLLYKKGA |
| | | | | | TCAAGT | HDDPANWRPICLQQ |
| | | | | | CTCGGC | AIYKLYTGVLARRLVR |
| | | | | | AGCGAC | WLDVNDRHAPGQK |
| | | | | | GCGTCT | GFRAVNGCGEHNFL |
| | | | | | TTCTATA | AATLIDQARRKRRSLY |
| | | | | | GCGCAG | EVWYDFRNAFGSVP |
| | | | | | CTGACG | FQLLWDSLQRLGAP |
| | | | | | AGGCCG | ADFIDMCKGLYHQA |
| | | | | | ATCTGG | AFVIGNAADGPTAAI |
| | | | | | CGGCCC | RQQVGVFQGCPLSP |
| | | | | | CCGATT | QLFNVAISPLLFALRR |
| | | | | | GGTCCG | LPETGVQLSGDDRVG |
| | | | | | ACTTTCG | VSAYADDLKTFSSTKA |
| | | | | | GCCAAT | GATKQHELVAAFLA |
| | | | | | CAGCGA | WTGMKANAAKCSS |
| | | | | | CGACGA | MGVRRNSNGATEAD |
| | | | | | GGGGGC | NLDLALDGTPIPSMT |
| | | | | | AGGGGT | HMQSYTYLGIGDGF |
| | | | | | TTACACT | DHVHRRIELAPKLKTL |
| | | | | | TTTGCCC | KQDTTALLESGLAPW |
| | | | | | CCGTTTC | QVVKAVKVYLYPRVE |
| | | | | | GGCTTC | YALRHLRPEDQLLESF |
| | | | | | AACTTCA | DLHLRAGLRHLLRLPK |
| | | | | | GGCCAA | NANNDFFYSPVSRG |
| | | | | | AATGGC | GLGLLPLVELHAALQI |
| | | | | | GATTTG | AHGWQMLNSTDPA |
| | | | | | GACCCT | TRRIAREQLHQIADA |
| | | | | | CCACGC | RHRLDKAHWKERGD |
| | | | | | GCCGTG | ELCQLFLNLDLGTSA |
| | | | | | CCACTG | HAPPKRRNCDIGSLW |
| | | | | | CTCGGC | VDVRKNLQAFGLKLE |
| | | | | | ACCGGC | TAPADAESGTPALPL |
| | | | | | GGCGAT | QLRVPHHEKWLTHR |
| | | | | | 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 |
| | | CCCGCGAGGCCAACTACACCGATCTTCCTCTACACTTTTCTAATTCGCCTCC | CCCGCC | RDAKRLDRTLSVACI |
| | | GTCTTCGGTCTTCGGTTGTCGGGCTTTTTTTCTTTTTGACCAATCAGAGCG | 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 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | AAGTAT GGGAGG ATCCCTG GCCTCTC AGGCTG TACGCG RGGACC AAGTAC AGCCGA ATCCCTG GCCTCTC AGCCTG TACGCG GGGCTA TACTTG GTCTAA GTATGC CCCGGT CGCTGG CCTCTGA GCCCGT ACGC (SEQ ID NO: 1324) | | MQDQVDAEQQARN RWGPPLPRPLVASRV AARLGEVPPPRWGP PLPRGVVVSRIAARLE AVPVPRGGPPLPRSF VATRIADRLAPPSPDL SLLDEEMKESEPPDP THHSADEDSTDAETA DAVMEPAFVSDPPT ATPREWRLQFDGAC RGGPNPGGAGALLY NPEGAVVWTGSHY MPGAKETNNSAEYT |
| | | | | | TGCGCGGC AGACCAAG ACGCGCAG CCAACAAC AGACCGTG CAGCGTGC GAAAGG GCCCCA CCCACG TAGGGA ACCGCC CTCGGG AAACCC AGTCTG | TAGACG GCAACA CTCTGG CCCGCG TCGGCC GAAAGG GCCCCA CCCACG TAGGGA ACCGCC CTCGGG AAACCC AGTCTG | |
| | | | TGCGCGGCAGACCAAGACGCGCAGCCAACAACAGACCGTGCAGCGTGC GGGTGGAAAGCGCCGCCGCCTGAACGCTGGTGACGATGAAGACCAGC GAGAGCTGGCCGAGCTCCTGCTCGTGGACGAGGACAAGGCTGGCGCCG AACACCCCGCGCTCAGGCTGCCCACGGCCAGCGCTCATCCGGCCTCCGT CCTCTCCGTGTACGCGCACGCTGCAACTCGCTTCGACTGCACGCTGTGCA CGTACACGGCTGCCAGCCTCGCTTCGCTCAAGCGCCACCGCTCGTCTCG GCACCGACACGGCCTTCCTCGACAAGTTCTTGGCGGGCTGCGCGTGC GGCACGCCCTTCGCATCGAGGTTGGCCGCAGCCAGACACGCGCAAGCCG TGCGCCAACCTCTGCACCACCTCGGCGACGACTTCGACGGCAGCAAAGG CATCAAGCCCCACTGCTGCCGGAGGCAGACCCACCGTCCGTGCAGTGGT CACCGCCGCGCCCGACCTGCCCCGCCAGTATCCCTCGGAGCTCGTTGCG TCCCCCCCGCAGCCGAGCTCCACCAACGTTGCAATGCAGGACCAGGTCG ACGCAGAGCAGCAGGCGCGAAATCGCTGGGGCCCACCGCTCCCTCGAC | TGCGCGGC AGACCAAG ACGCGCAG CCAACAAC AGACCGTG CAGCGTGC GAAAGG GCCCCA CCCACG TAGGGA ACCGCC CTCGGG AAACCC AGTCTG | |
| | | Phytophthora ramorum | | | | | |
| | | . | | | | | |
| | Utopia-3_PR | | | | | | |
| NeSL | | | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | 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 |
| | GCCAACCGGGCGCTCGACCTGAAGAGCGACAAAGTCGAGTGCAAGGAA | 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 |
| | CCAGCCAGGTGAGCCGTCACCGACGCGCCGCCGCCGCCGAGGAAAGCG | 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 |
| | TGGCAGCAGCTCGCACGGCACGCGACGCACAACAGTGAGGGCG | 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 |
| | | TCATCGACCACGCGCGCCGCAAGCACCGGCCACTGTACGAGGGTGTGGT | 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 |
| | | GACCGGCATGGCGCGAACCCGGCCAAGTGCAGGTCGATGGGCGTGC | CAGCCGAG | | VKGYLYPRVEYALRH |
| | | GCCGCAACGGCAACGGCGCCATCGAGGCGGACCACCTGGAGCTGGCGC | CTCCACCA | | LRPDDQLLESFDLHLR |
| | | TGGACGACACGCCGATCCCTACGTTGACGCACCTGCAGTCGTACACGTA | ACGTTGCA | | RGLRHLLRLPKSANN |
| | | CCTGGGGATCGGCGACGGCTTCGACCACGTGCGCCGCCGCGTCGCGCT | (SEQ ID | | DFVYAPVSRGGLGFL |
| | | CGCGCCCAAGCTCAAGTTGCTCAAGCAGGACGCGACGGCGCTGATGGA | NO: 1202) | | PLVELHAALQIAHGW |
| | | GTCCGGCCTCGCGCCGTGGCAGGTGGTGAAGGCCGTGAAGGGGTACCT | | | QMINSPDPAIRRIAR |
| | | CTACCCTCGGGTGGAGTACGCCCTCCGGCACCTGCGACCCGACGACCAG | | | EQLHQVADARHRLD |
| | | CTCCTGGAGAGCTTCGACCTCCACCTCCGCCGCGGCCTCCGCCACCTGCT | | | KDHWKQRGDELCEL |
| | | CCGCCTCCCCAAGAGCGCCAACAACGACTTCGTGTACGCGCCCGGTGTCG | | | LLNGELGTSAHAPPK |
| | | CGCGGGGGCCTGGGCTTCCTGCCGCTGGTGGAGCTGCACGCGGCGCTG | | | RRNGDIGSLWVDVR |
| | | CAGATTGCGCACGGGTGGCAGATGATCAACTCGCCCGACCCAGCCATCC | | | KNLKAFGLKLATAPA |
| | | GGCGCATCGCGCGCGAGCAGCTTCACCAGGTTGCCGACGCGCGGCACC | | | DPESGAPAKPLQLCV |
| | | GGCTCGACAAGGACCACTGGAAGCAGCGCGGTGACGAGCTGTGCGAG | | | PHHAEWLDHRNVLR |
| | | CTCCTCCTCAACGGGGAGCTCGGGACGTCGGCCCACGCGCCACCCAAAC | | | HVKQHMKNKRWRA |
| | | GGCGCAACGGCGACATTGGCTCGCTGTGGGTCGACGTCCGCAAGAATC | | | WCSHVDQGRTARA |
| | | TCAAGGCCTTCGGGCTGAAGCTCGCAACCGCGCCCGCCGACCCGGAGT | | | HGGVGSGFLTRPRG |
| | | CCGGCGCACCTGCCAAGCCGCTGCAGCTTTGCGTGCCGCACCACGCTGA | | | MWESDYRFAVAARL |
| | | GTGGCTGGACCACCGCAACGTGCTGCGGCACGTGAAGCAGCACATGAA | | | NMLDTVNVLARRRL |
| | | GAACAAGCGCTGGCGGGCCTGGTGCAGCCACGTGGACCAGGGGCGGA | | | RAHDRCRHPGCRWK |
| | | CGGCGCGCGCACGGAGGCGTCGGCAGCGGTTTCCTGACGAGACCGC | | | 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 |
| | | | | GTCGCGAGTGTGCTTGCCCCACCCGGCAGGGCAGGCAGCATGCTCGCC | CCCACACG | ACCGCC | | HRRSAHRGTRFKDIF |
| | | | | GACCCAGCAGCGCTGGCAGCCGGTCTGGCTCGCGCTCCGCCGCCCCCAT | CCGCCGCC | CTTTAAA | | TSGCQCSLVFYARIVA |
| | | | | CGGCACCCCAGGACCCAAGCCCCGCCTTCCCTGCTGGTCCCGCCGGCCA | 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 |
| | | CACAAGAACCTCCTTTGCAGCCAGAGCCACAAGCAGCAGCAACAGT | TATTGGCC | TCACTCG | APSGAVVWTCSHFM |
| | | CGCAGCAGACATCGCAGCAACAGTAGCAGCAGACATTGCGGCAGCAGC | CTTGGAGC | CATTTTT | PSRSETNNTAEYTALL |
| | | AGCGAACGCAGCAATGGACGTGGATGGTGGACCGGCAGCCGACGAGA | TCGACGGC | GGGCAA | LGAQSAVHHGAKRL |
| | | CATGGCTGTTGCGCTTTGACGGCGCATGCCGACGGAACCCAGGACCTG | GACAGCA | TCGTCT | NIEGDSHLILSQVRGA |
| | | GTGGAGCAGGTGCCGCACTCTTTGCGCCCAGCGGTGCCGTAGTATGGA | GCGACGAC | GCGGCT | FACNNKRLRSLRNRV |
| | | CGTGCTCGCACTTCATGCCGAGCCGCTCCGAAACGAACACACCGCCGA | GAGGACG | AGCTTG | QASLRQLDWYRLQHI |
| | | GTACACGGCGCTACTGCTCGGCGCACAGAGCGCAGTCCACCACGGCGC | CTCAAGAC | CTAGCG | DRKANQHADRLANR |
| | | CAAACGCTTGAATATCGAGGGCGATAGCCACCTGATCCTCTCGCAAGTG | CCCCACGC | GCGGAC | ALDLRRTVTECGPHA |
| | | CGCGGTGCCTTCGCGTGCAACAACAAGCGGTTGCGGTCTCTCCGAAACC | CGCCGCCC | GAGCGG | ETRNRCFQTPQPLVE |
| | | GAGTCCAGGCCTCTCTTCGGCAGCTCGACTGGTACCGCTTGCAGCACAT | CAGAACCC | TCTCCG | PGETHCVPGSDEVLA |
| | | CGACCGCAAGGCCAACCAGCATGCCGACCGCCTCGCGAACCGCGCCCT | CCAGAAGA | GGGGCG | ANTAMEDATAVPTE |
| | | GGACCTGCGCCGCACGGTGACCGAGTGCGGCCCACACGCGGAGACTCG | CGTCGCGA | TACACCT | DDEAEVAARDGGEV |
| | | CAATCGGTGCTTTCAGACACCGCAGCCCCTCGTCGAGCCCGGCGAGACG | GTGTGCTT | TTCCCCC | FPTIAIGPDSAPAKQP |
| | | CATTGTGTACCGGGGTCTGATGAGGTCCTCGCTGCTAATACCGCGATGG | 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 |
| | | CTTCCGACGCAGCCAGCACGACCACAGCCCCGGGAGCCTCGTATGCAGC | | 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 | |

| | | | | | |
|---|---|---|---|---|---|
| | | GACATCGGTTCCCTATGGAACGACGTCCGCGACACGCTGCAGACCCTCG | GACATC | | RNGDIGSLWNDVRD |
| | | GACTCAAGTTCGCCGCAGGCGACGAGGAGGAAGCACCTGGACTACTGC | AACTTCA | | TLQTLGLKFAAGDEE |
| | | AGCTCCGCGTGCCGCACCACACCAAGTGGCTCAGCCACTCGACGGTGTT | GGCCAA | | EAPGLLQLRVPHHTK |
| | | GAGGCACGTGAAGCTGCACATGAAACTGCCGCCGCATGGACACGTGGAA | AATGGC | | WLSHSTVLRHVKLH |
| | | GTCCAAGGTCAGTCAGGGCACGACGGTGCGAGAGCACGGCGGCGCGTGG | GATTTC | | MKLRRMDTWKSKVS |
| | | GCTCACGCTTCATCACGGCGGGGTGCAGGCCTGTCCGACGCCGAGTATCG | GACCCT | | QGTTVREHGGVGSR |
| | | CTTCGCTATTGCGCCCCCGCGCCCACCTGATTGATACCAACAGCACCCTCA | CCACGC | | FITAGAGLSDAEYRFA |
| | | AGCGCCGTCGACTCCGGGCCAACGACACCTGCCGGGCTCCGGGATGCT | GCCGTG | | IAPRAHLIDTNSTLKR |
| | | CCTACACGGAACCCCCCCGCCCACATCCTGAACAAATGCAGTCCCAACAT | CCACTG | | RRLRANDTCRAPGCS |
| | | GGACGCGATCAGGAAACGCCACGACGATGCTCTCGAGCGTATCGCTGA | CTCGGC | | YTEPPAHILNKCSPN |
| | | CGCCCTGCGGCGCAAGGTGGAGAAGTCCGGCGGCAGGCTCGAGGTGG | ACCGGC | | MDAIRKRHDDALERI |
| | | CCATCAACAAGACGGTACCGGAGTACGACGGCGCGGCCCTGCGCCCCG | GGCGAT | | ADALRRKVEKSGGRL |
| | | ACATCGTACTCCGGAACACGGAGACCAAGCGTGCCATCATCGCCGACCT | TCAGCG | | EVAINKTVPEYDGAA |
| | | GGCGATCACCCACGAGAACCAGCCGACCGACGCGACCACGTCTTCAGC | GGTGCA | | LRPDIVLRNTETKRAII |
| | | GCTACAGCAAAGCCGTGACAACAAGATTACCAAGTACCAGACGGTCGC | ACTTCG | | ADLAITHENQPTDAT |
| | | GGCGGCAATGATGCGCGCCGGTTGGCGGGTTCGCGTGACGGGCATCGT | GGCACG | | TSSALQQSRDNKITKY |
| | | CTACGGCTCGCTCGGCTCAGTCCTGCCGAGCAACTTCAAGGTCTACACC | TGTGCA | | QTVAAAMMRAGW |
| | | GAGCTGCTGGCGCTGCTCAAGCGGGACGCCCGCCGCCTTAACCGCCAG | ACACAT | | RVRVTGIVYGSLGSVL |
| | | CTCTCCAGCCACTGCATCCGGCCAGCGCGCGGATATGGAGCGCGCACT | GCAGCG | | PSNFKVYTELLALLKR |
| | | GCCGACGACACCGTGAGCGGCAACGAAGCGGCAATGCCTCGCGAGCGT | CCCATTG | | DARRLNRQLSSHCIR |
| | | CGCGTGGGTCCGGGGGGGCCCCGCGGCGCACATCGCAGGCATCGGCA | CACGCC | | ASARIWSAHCRRHRE |
| | | CGGCGGTAAACGGGTCACTTGACCGACAGGGCACCACCCAGGTAGGGA | AAGCGG | | RQRSGNASRASRGS |
| | | ACCGCCCTTTAAAACCCAGGAAGACACAAACACCCTCCACATAGTGACA | CATCGC | | GGAPRRTSQASARR |
| | | TACATATTTTAGCCTAGATTTCAGTTACGGAGAGGTTACTAACTGGTACA | GGGACG | | (SEQ ID NO: 1448) |
| | | TAAAATTACACATTCTGTTCTAATCAGTGTGAAAACTGGTTTTCGCCTTTT | ACGCCT | | |
| | | GGCGGACTTTTTCACTCGCATTTTTGGGCAATCGTCTGCGGCTAGCTTGC | CGGCCG | | |
| | | TAGCGGCGGACGAGCGGTCTCCGGGGGCGTACACCTTTCCCCCGCGAG | CTCAAG | | |
| | | GCCAACTACACCGATCTTCTCTACACTTTTCTAATTCGCCTTCGTCTTCGG | CGCAGC | | |
| | | TCTTCGGCTGTCGGATTTTTTTCTTTTTGACCAATCAGAGTGCGCCATGC | CCCGCC | | |
| | | GACTCTTCTGGCCAATCAGAGACCGGGCCCTGTCCTCGGACAGCGAGGC | CTTCCAG | | |
| | | CTCCACGGCCAGCCAATCAAGTCTCGGCAGCGACGCGTCTTTCTATAGC | CACGAC | | |
| | | GCAGCTGACGAGGCCGATCTGGCGGCCCCCGATTGGTCCGACTTTCGGC | CTCGCG | | |
| | | CAATCAGCGACGACGAGGGGGCAGGAGTTTACACTTTTGCCCCCGTTTC | CCGTTT | | |
| | | GACATCAACTTCAGGCCAAAATGGCGATTTCGACCCTCCACGCGCCGTG | GGCGGA | | |
| | | CCACTGCTCGGCACCGGCGCGATTCAGCGGGTGCAACTTCGGGCACG | TCGCCAT | | |
| | | TGTGCAACACATGCAGCGCCCATTGCACGCCAAGCGGCATCGCGGGAC | CAAGAC | | |
| | | GACGCCTCGGCCGCTCAAGCGCAGCCCCGCCCTTCCAGCACGACCTCGC | GTGCGA | | |
| | | GCCGTTTGGCGGATCGCCATCAAGACGTGCGAGAGCCAGGCGGGGGTCG | GAGCCA | | |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| | | | | | |
| | | | | | |
| GGCAAAATATACTTACTCTAAGTATGCCCGAATCCCTGCCCTCTCAGGCT GTACGCGGCCCCATACTTGATCTAAGTATGGGAGGATCCCTGGCCTCTC AGGCTGTACGCGAGACCCGTACGGCCGAATCCCTGGCTTCTCAGCCTGT ACGCGGGCTATACTTGGTCTAAGTATGCCCCGGTCGCTGGCCTCTGAG CCCGTACACA (SEQ ID NO: 1080) | | | | GGCGGG GTCGGG CAAAAT ATACTTA CTCTAA GTATGC CCGAAT CCCTGCC CTCTCAG GCTGTA CGCGGC CCCATAC TTGATCT AAGTAT GGGAGG ATCCCTG GCCTCTC AGGCTG TACGCG GTACGG GTCGGC ACTTGG GCCTGT GGCTAT ACGCGG CTTCTCA CCCTGG CCGAAT GTACGG AGACCC TACGCG AGGCTG GCCTCTC GCTGTA 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 |
| | | | | GCCCCAAGAACAAGCACCCCCAGCTAGCGCGCGCGTGGAGCCCATGTC | AAGCGCTC | ACAGCG | DQGARRQQPPWQR |
| | | | | GGGCGACGTGGTCAGCAGCGACGGCAGCAGGACGACGGACGCGA | CGCCGCCG | GCCGAC | VGGKRRRLNDVDDE |
| | | | | GCGGCGACGGCGACGACGGCAGGCAGCAGCGATGCAGCAGGCGAT | AGCCCAGC | AGGGCC | DTRELAELLLEEEDEA |
| | | | | GTCGGCGTCGTCGCCATGGATGTTGACCAAGGTGCCCGCCGACAGCAG | 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 |
| | | | | TGCGAACCACGCACACGCGTGCGACAGCCTCAACCGCACCTTCTCGGTA | 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 |
| | | | | GGCTGCTGCGCTTCGACGGCGCTGCCGCGCCAACCCTGGCCCAGGTG | (SEQ ID | CTACCTG | PDGSVVWTCSHYMP |
| | | | | GCGCCGGCGCAGCGCTGTCCCAGCCCGACGGCTCCGTTGTATGGACTTG | NO: 1204) | GAAGCC | SSSETNNTAEYTALLL |
| | | | | CTCCCACTACATGCCGAGCAGCGAACCAACAACACCGCCGAATAC | | AAGGAT | GTRAAADHGTTTLRV |
| | | | | ACGGCGCTGCTGCTCGGTACGCGAGCCGCCGCCGACCACGGAACCACG | | TAACCG | EGDSTLVIQQVRGIF |
| | | | | ACCCTGCGAGTGGAGGGCGACAGCACACCTTGGTGATCCAGCAGGTGCGT | | GTCGGT | ATRSVTLRHILRDQVK |
| | | | | GGCATCTTTGCCACGCGGAGCGTCACTCTCCGCCACCTCCGCGACCAAG | | AATACA | LELARVGRFSLHHIDR |
| | | | | TCAAGCTGGAACTCGCCCGGGTGGGGAGATTTTCGCTCCACCACATTGA | | CATAACT | QANAHADRLANRAL |
| | | | | TCGCCAAGCCAATGCCCACGCCGACCGCCTCGCCAACCGAGCCCTCGAC | | TT (SEQ | DLRRTVSECGVHPDG |
| | | | | CTCCGGCGGACAGTGTCGGAGTGCGGCGTCCACCCCGACGGGAACGGC | | ID NO: | NGCTPTAIDDRPLAP |
| | | | | TGCACGCCCACGGCCATCGACGACCGAACCCTAGCGCCGACTCAGCAGC | | 1327) | TQQPPDAPPPPPAA |
| | | | | CACCTGACGCCCCGCCACCGCGCCCGCCGCCGACATTGAGATGGAGG | | | DIEMEDPDDEDLADI |
| | | | | ACCCGGACGACGAGGACCTCGCCGACATCGACGACGGCGAGGTCTACG | | | DDGEVYAAMRVGP |
| | | | | CGGCCATGCGCGTTGGCCCGAATGCCACTCCGCAGCGACGACGCCCGTG | | | 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 |
| | | TCACGTGCGCCGCCGGTCGAGCTCGCTCCAGCACTCAAGCAGCTTAAG | | GRGVLKTDDLQLDLA |
| | | GACGACGCGACGACGCTGCTGCAGTCGGGCCTCGCCCCGTGGCAGGTG | | GTPIPALSMSASYTYL |
| | | GTGAAGGCGGTGAAGACGTACCTCTACCCCCGGGTGGAGTACGCTCTC | | GIGDGFDHVRRRVEL |
| | | CGCCACCTTCGCCCCTTCCAACAGCAGCTTGAGGGCTTCGACCGGCACC | | APALKQLKDDATTLL |
| | | TCGCGCGCGGCCTCCGCCACCTCCTGCGACTGCCGGGCAACGCCACGGC | | QSGLAPWQVVKAVK |
| | | GGAGTGCTTCTACGCTCCGGTGTCGCGCGGGCTGGGCCTGCTGCC | | TYLYPRVEYALRHLRP |
| | | ACTGACGGAGCTGCACGCGGCGCTCCAAGTCGCGCACGGATGGCAGCT | | FQQQLEGFDRHLAR |
| | | GCTGAATTCGAAGGACCCTGCCATCCGGCGCATCGCGCGCGTGCAGCTC | | GLRHLLRLPGNATAE |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CGGCAGATCGCCGACGCGCGCCACCGTATCGACAGCCGCGCATGGGAG<br>GGGCGCGACGAGGAGTTGTGCGAGCTGCTGCTCAACTCGCAGCTCGGG<br>ACGTCGCCCGACGCGCCACCGAAGCGTCGGAACGGCGACATCGGTTCT<br>CTTTGGGTCGACGTCCAGCGCCACCTCCGCACGCTCGGCCTCAAGTTCG<br>CGACCGCGCCGGCCTGCGCGGACGCCGGCTCTGCGGCGACCACGTTGC<br>AGCTCCGCGTGCCGCACCACGATAAGTGGTTGGACCATCGCACTGTCCT<br>GCGGCACGTGAAACTGCACGTGAAGCACCGGCACTGGTCCAAGTGGGC<br>GGCCATGCGCGACCAAGGCAAGACGGCCCGCGCCCATGGCGGCGCGG<br>GGAGCGGCTTCCTCACCCGGCCACGAGGCATGTGGGAAGCCGACTACC<br>GCTTCGCGGTGGCCGCCCGCCTGAACCAGCTCGACACGCACAGCGTGCT<br>GAAGCGCCGGCGTCTCCGAGCCCACGACCACTGCAGGCAGCCGGGCTG<br>CTCCAGGGCGGAGACGCTGGCGCACGTGCTCAACCACTGCGCCGGCAC<br>CATGGACGCGGTCCGCGGCCGCCACGACGCCCTCAAGCACACATCGA<br>GCGCGCGCTGCACGCGTCGTCGCCGGGCGGCCAGGACCGCGTCGAGCT<br>CCGGGTGAACCAGACGGTGCCGTCGCTCGCCGGCCCCGCGCTACGGCC<br>CGACCTCCAGCTGTACAACCACACCAAGAAGATGGTGGCGGTGGTCGA<br>CCTGGCCGTGGCGTTCGAGGAGCAGGCGAGCGACGACCCGGAGAGCT<br>CGGCGCTGGCACGCATCGCCGCACACAAGCGCGCGAAGTATGCCGGCG<br>TCAAGCGACACCTCGAGCGCCAAGGGTGGAAGGTCCACCTCTCGGCGC<br>TCGTGTACGGCTCGCTCGGCGCGGTGCCGGCCGGCAACCACACAAGGTGC<br>TTACTGAGCACCTTGGGTTGCTCAAGCGTGATGCGAAGCGGCTGGACC<br>GGCAGCTCTCGGTGGCGTGCATCCAGTCCAGCCGCCGCATCTGGAACCT<br>GCACTGCAGCCAGCACCGCGCGCCAGCACCAAGCACCAGGGGGAAG<br>TCGGGCGGCGGAGACCGGGGGGGACTCCGCCGCGCACCGGCCGCCGCT<br>AGACGGCACACAGGCCCACAGCGGCCGACAGGGCCACACCCAGGTAG<br>GGAACCGCCTTCAAACCCGGCCGGTACATTATGGTCCGACACCTATGAG<br>GTGCAACCGGTACACAAGTTACACACCACATAGCGACTACCAGGTATTT<br>ACTACCTGGAAGCCAAGGATTAACCGGTCGGTAATACACATAACTTT<br>(SEQ ID NO: 1081) | | | | CFYAPVSRGGLGLLPL<br>TELHAALQVAHGWQ<br>LLNSKDPAIRRIARVQ<br>LRQIADARHRIDSRA<br>WEGRDEELCELLLNS<br>QLGTSPDAPPKRRN<br>GDIGSLWVDVQRHL<br>RTLGLKFATAPACAD<br>AGSAATTLQLRVPHH<br>DKWLDHRTVLRHVK<br>LHVKHRHWSKWAA<br>MRDQGKTARAHGG<br>AGSGFLTRPRGMWE<br>ADYRFAVAARLNQL<br>DTHSVLKRRRLRAHD<br>HCRQPGCSRAETLAH<br>VLNHCAGTMDAVRG<br>RHDDALKHIERALHA<br>SSPGGQDRVELRVN<br>QTVPSLAGPALRPDL<br>QLYNHTKKMVAVVD<br>LAVAFEEQASDDPES<br>SALARIAAHKRAKYA<br>GVKRHLERQGWKVH<br>LSALVYGSLGAVPAG<br>NHKVLTEHLGLLKRD<br>AKRLDRQLSVACIQS<br>SRRIWNLHCSQHRA<br>RQHQAPGGSRAAET<br>GGTPPRTGRR (SEQ ID NO: 1449) |
| NeSL | Utopia-5_PI | . | Phytophthora infestans | CTCAAGCCTAGCAGCGGCTACCGCAGCTACTCCAGCTCCGGTAGCTGCT<br>ACTGCTACAACTGCTCGCGCTGCTGCTCGCCGTCGCCGTGGAGCACCACG<br>CGCACGCTGAACCGAACCAAGAACATCTACCGATGGCTCGCGTGCTCGT<br>GGAGCCCATGCAGGTCGACGAGTGCAGCAGCTGCGACCGCAGCACTTT<br>GACAGCCGACGACGGGAGCGGTGACGACGTTGCAGCCCCAAGCTCCTT<br>GAACAGCAACGACGTCGCCGCCCCCATGGACGTTGACTCCGGCACTGAT<br>TGCCCTCCAGCGTTGCAACAGCCGCCCCCAGCGACCGCGTGCGCTGCACG | CTCAAGCC<br>TAGCAGCG<br>GCTACCGC<br>AGCTACTC<br>CAGCTCCG<br>GTAGCTGC<br>TACTGCTA | TGACGC<br>ACCGTG<br>ACATAG<br>TGCGGC<br>ACGTGA<br>AGATGC<br>ACATGA | MARVLVEPMQVDEC<br>SSCDRSTLTADDGSG<br>DDVAAPSSLNSNDV<br>AAPMDVDSGTDCPP<br>ALQQPPQRPRALHV<br>GSKRRRLDADDEEEA<br>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 |
| | | CTGGTGAGCTCCCCCCGATGTTGAAGTGCAGGCACCAGAGACAGAGGCC | NO: 1205) | ATGGTG | WGTPLPRVLVASRIA |
| | | ACCGAGAACCGCTGGGGCACTCCACTTCCGCGCGTGCTGGTCGCGTCTA | | GGGTTG | GRLAQVPPPRWGPP |
| | | GGATCGCCGGGCGCCTCGCTCAGGTGCCACCCCCACGCTGGGGCCCCCC | | GTAGTG | LPRTTIAGRIATRLAA |
| | | GCTACCGCGCACTACCATCGCGGGCCGCATTGCGACGAGGCTCGCCGCC | | GCTTCCT | TPAPRWDPPLPRSLV |
| | | ACACCTGCCCCCCGCTGGGACCCACCGCTGCCACGAAGCCTGGTTGTGT | | CACACG | VSRIAARLLPALPDAP |
| | | CACGCATCGCCGCCCGCCTGCTACCCGCGCTACCAGACGCGCCTGCATG | | GCCGCG | ACEEEAKDSDTMDW |
| | | CGAGGAAGAGGCCAAGGACAGCGACACGATGGACTGGGCGCCAACTT | | 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 |
| | | CACCAACACACCTCTGTGGCGACCGCAGCGCCTACTGCTCCTGCAGCGCCC | | AGCGCC | RGCTNTSVATAAPTA |
| | | TTGCCCCCAGCCCGGCCTCCTGCAACTACAGCCGCTCCGTCTCATGATGA | | GGCGCC | PAAPLPPARPPATTA |
| | | CGACCACAGCGTTCAAGGCGACATCGACGATGGCGAAGTATACGCTGC | | TCCGAG | APSHDDDHSVQGDI |
| | | AATGTGCATCGCGCCCCGATGCGGTGCCGCACCGCCGACCGCGCCTCCGC | | CACATG | DDGEVYAAMCIGPD |
| | | CTACGGCACTTGACGGACGAGGAGTCCGAGGAGGCTGGCGACGTCGTG | | ACAGGT | AVPHRRPRLRLRHLT |
| | | GAGAGGCTGGCAGCGTCGCTGGCGGCGAAGATTGCAGACGCGCCCGA | | GCAGAC | DEESEEAGDVVERLA |
| | | CTGGGAGACGGCGGAGGGCTACATCACAGCCCTACCGTACGCTCTCTAC | | ACCCAG | ASLAAKIADAPDWET |
| | | GACAAGCTGCAGCCGTACTCCCAGGCCCAGCCTCAGCCTCCAAGCCAAC | | GATGCA | AEGYITALPYALYDKL |
| | | AGCAGCAGCAGCAGCAGCGACCACGGCAGCAGCAGCAAACACGG | | 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 |
| | | TCGGGGACGGCTTCGACCACGTGCGCCGCCGCCTCGAGCTCGCGCCAAT | CTTCAGC | | QHELVATFLRWTGM |
| | | GATGAAGCAGCTCAAGCACACGACGCGACGGCGCTGATGCAGTCGGGGCT | 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 |
| | | GGGAAGCCGACTACCGGTTCGCGGTGGCCGGCCGCTTAAACCAGGTAG | | 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 | |
| | | AGGCCCAGCCAAGGACCAAGAGGCAGCCGGGCGACGGAGACCGGGGG | | 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 | TGATGC | MSGDVVSSDGSSRT |
| | | | | | | CGACTC | CCGCCA | TDVSGDGDDGADGA |
| | | | | | | GCAAGC | ACTACA | GSSDAAGDVGVVA |
| | | | | | | CTTACCT | AGGTGC | MDVDQGARRQRPP |
| | | | | | | ATATTTT | TTACTGA | WQRVGGKRRRLND |
| | | | | | | AGACGT | GCACCTT | VDDEDTRELAELLLEE |
| | | | | | | AGCGAC | GGGCTG | EDEVGAQAPALRLFA |
| | | | | | | CAAATT | CTCAAG | ASAHPASVLSVYAHN |
| | | | | | | ACAAAT | CGTGAT | AQRFVCTLCAYTAAS |
| | | | | | | TTGGTA | GCGAAG | FASLKRHRDSRHRRV |
| | | | | | | ACGAGT | CGGCTG | SFVDKFSAGCACGTP |
| | | | | | | AAGCCA | GACCGG | FGSRLAAARHAQAC |
| | | | | | | AATGGT | CAGCTG | ASLSIPRTVTAPAAAG |
| | | | | | | AATACA | TCGGTG | DLSPTATGANATASA |
| | | | | | | CAAAAC | GCGTGC | AATSPDLPRPASPELA |
| | | | | | | TTT (SEQ ID NO: 1328) | ATCCAG | ASPPQTSPFDVAIQA |
| | | | | | | | TCCAGC | DAAEQTAWTRWDP |
| | | | | | | | CGCCGC | PLTRAAVAARVASRL |
| | | | | | | | ATCTGG | AVVPAPRWGPPLSR |
| | | | | | | | AACCTG | TLVASRIAARLDAQTS |
| NeSL | Utopia-5_PR | . | Phytophthora ramorum | CCAGGCATCACCCCCGCCGCCCCCAAGCGCTCCGCCGCCGAGCGCAGCT | CCAGGCAT | | | |
| | | | | GCGGCTGCTGCCGCCCCACTGGTGACGGCAGTCGCAGCTGCCCCTGCCA | CACCCCCG | | | |
| | | | | CCTCTCCTGTTGGCCGCGCCGCCGTGGAGCCCCCGCGCGCCGCCGAGCC | CCGCCCCC | | | |
| | | | | GCCCCAAGAACAAGCACCCCCAGCTAGCGCGCGCGCGGAGCCCATGTC | AAGCGCTC | | | |
| | | | | GGGCGACGTGGTCAGCAGCGACGGCAGCAGGACGACGGACGGACGTGA | CGCCGCCG | | | |
| | | | | GCGGCGACGGCGACGACGGCGCCGACGGCGCAGGCAGCAGCGATGCA | AGCGCAGC | | | |
| | | | | GCAGGCGATGTCGGCGTCGTCGCCATGGATGTTGACCAAGGTGCCCGC | TGCGGCTG | | | |
| | | | | CGACAGCGGCCGCCATGGCAGCGCGTCGGCGGCAAGCGTCGACGTCTC | CTGCCGCC | | | |
| | | | | AACGACGTCGACGACGAAGACACCCGCGAGCTGGCAGAGCTCCTGCTC | CCACTGGT | | | |
| | | | | GAGGAGGAGGACGAGGTTGGGGGCCCAAGCACCAGCGCTCAGGCTATT | GACGGCA | | | |
| | | | | CGCGGCCAGCGCTCACCCGGCCTCCGTCCTCTCCGTGTACGCACACAAT | GTCGCAGC | | | |
| | | | | GCGCAGCGCTTCGTGTGCACGCTGTGCGCGTACACGGCCGCCAGCTTCG | TGCCCCTG | | | |
| | | | | CGTCGCTCAAGCGACACCGGGACTCACGACACCGGCGCGTGTCCTTCGT | CCACCTCT | | | |
| | | | | GGACAAGTTCTCGGCGGGCTGCGCGTGCGGCACGCCCTTCGGCTCGAG | CCTGTTGG | | | |
| | | | | GCTGGCAGCGGCGAGACATGCGCAAGCGTGCGCCAGCCTCTCGATCCC | CCGCGCCG | | | |
| | | | | ACGCACGGTAACAGCTCCGGCGGCAGCAGGAGATTTGAGCCCCACTGC | CCGTGGAG | | | |
| | | | | CACGGGAGCCAACGCCACCGCCTCGGCGGCCGCAACTTCGCCCGATTTG | CCCCGCGC | | | |
| | | | | CCCCGCCCAGCTTCCCCGGAGCTCGCTGCGTCCCCCCCGCAGACGAGCC | GCGCGCCG | | | |
| | | | | CCTTTGACGTTGCCATACAGGCAGACGCTGCAGAGCAGACAGCGTGGA | AGCCGCCC | | | |
| | | | | CGCGCTGGGACCCTCCGCTCACCCGCGCTGCGGTGGCAGCCCGTGTCGC | CAAGAACA | | | |

| | | | | | |
|---|---|---|---|---|---|
| | CTCCCGCCTCGCCGTGGTTCCCGCGCCGCGATGGGGGCCGCCGCTGTCA | AGCACCCCC | CACTGC | RWGPPLPPAMVASR |
| | AGGACGCTGGTGGCCTCCAGGATTGCAGCACGCCTGGACGCGCAGACG | CAGCTAGC | GCGCAG | IASRLAAMPAPRWG |
| | TCACGATGGGGACCGCCCCTGCCACCGGCGATGGTGGCGTCCCGGATT | GCGCGCGC | CATCGA | PPLLRTVIASRIADRLL |
| | GCGTCGAGACTCGCCGCCATGCCGGCACCCCGCTGGGGCCCGCCGCTG | GGAGCCC | GCACGG | PPELAADEETKDDDV |
| | CTGCGGACCGTGATCGCATCCCGCATCGCCGACCGACTGCTCCCACCGG | (SEQ ID | CAGCAC | HMDNAASVDVDEES |
| | AGTTGGCGGCCGACGAGGAGACGAAGGATGACGACGTCCACATGGAC | NO: 1206) | 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 |
| | TGGGCAGCTTCTCGCTCCACCACATCGACAGGCAAGACACAACGCACACGC | | 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 |
| | CGACCTCCACGCGGTGGAGAGAGCTACGCCCAGCGACCGCCGCTCCATT | | 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 |
| | CTTCAGTGGCACCGCGGACGGCGTTAAGCGACACGCACGCTCGTGGC | | | CPLSPHLFTAAISPLLH |
| | CGACTTCCTGCGCTGGACCGGCATGGTGGCGAACCCCAACAAGTGCTGC | | | ALDRLKDTGVRLSAD |
| | ACCATGTCGGTCCAGCGCGACGGACGGCGTGCTGAAGGCCTGCGAC | | | DRPGASAYADDLKIF |
| | CTGGAGCTGCAGCTGGACGGAGCCCGGATCCCCTCGCTGATCATGAAC | | | SGTADGVKRQHALV |
| | GCGTCGTACGCGTACCTGGGGACAGGCGACGGCTTTGACCACGTGCGC | | | ADFLRWTGMVANP |
| | CGCCGCATCGAGCTCGTCCCGGCGCTCATGCAGCTCAAGGACGATGCG | | | NKCCTMSVQRDGRG |
| | ACGGCACTGCTCCAGTCGGGGCTGGCGCCGTGTCAGGTGGTGAAGGCG | | | VLKACDLELQLDGARI |
| | GTGAAGACTTACCTCTTCCCGCGGGTGGAGTACGCACACTCCGCCACCTCC | | | PSLIMNASYAYLGTG |
| | GCCCCTTTCAGCAGCAACTGGAAGGCTTTAATCGCCACCTCGTTTGCGG | | | DGFDHVRRRIELVPA |
| | CCTTCGCCACCTGCTCCGACTTCCGGTCAGCGCCACGACCAGCTTCTTCT | | | LMQLKDDATALLQS |
| | TCGTGCCGGTGTCGCGCGGCGGGCTAGGCCTGTTGCCACTGACGGAGC | | | GLAPCQVVKAVKTYL |
| | TGCACGCAGCTCCGGCTGATCGCCGACCGGCGCCACCGCCTAGACCCCG | | | FPRVEYALRHLRPFQ |
| | ACCACTGGAAGGAGCGGGAGGAGAAAATATGCGAGCTGCGGATCAAC | | | QQLEGFNRHLVCGLR |
| | TCGCAGCTCGGGACGTCGGCCCACGCCCCACCCAAGCGCCGCAACGGC | | | HLLRLPVSATTSFFFV |
| | GACATCGGCTCGTTGTGGGTCGACGTCCAGCGCCACCTCCGAGGACTCG | | | PVSRGGLGLLPLTELH |
| | GCCTCAAGCTAGAGACCGCCGGCGTGCGCGGAGACCGGCACCGCG | | | AAPADRRPAPPPRPR |
| | GAGGCGACGCTGCAGCTTCGCGTGCCGCACCACGAGAAGTGGCTGGAC | | | PLEGAGGENMRAAD |
| | CACCGCACGGTCCTGCGGCAGGTGAAGCTGCACATGAAGAACAAGCAT | | | QLAARDVGPRPTQA |
| | TGGCAGCGGTGGGCCGGATGAAGGACCAAGGCAAGACCGCCCGCAC | | | PQRRHRLVVGRRPA |
| | CCATGGTGGTGCAGGGAGCGGCTTCCTCACGCGGCCTCGCGGCCTATG | | | PPPRTRPQARDRAG |
| | GGAGACCGACTACCGCTTTGCAGTGGCGGGTCGGCTCAACCAGCTGGA | | | VRGDRHRGGDAAAS |
| | CACGCACAGCGTGCTGAAGCGCCGGCGCCTCCGCGCGCACGACCAGTG | | | RAAPREVAGPPHGP |
| | TCGACAGCCGGGCTGCTCCCGTGCCGAGACACTGGCACACGTGCTGAAT | | | AAGEAAHEEQALAA |

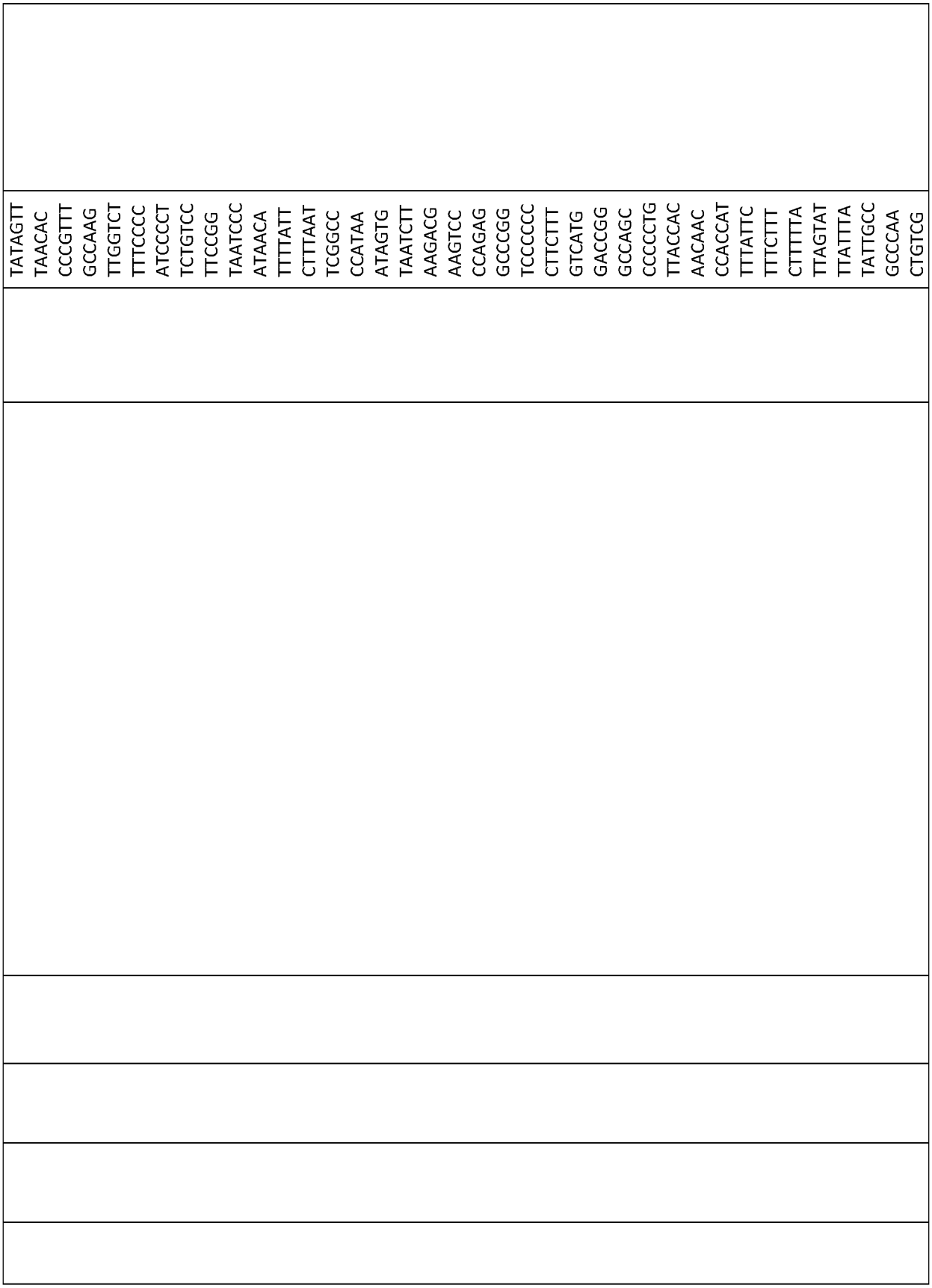
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CACTGCGCCGGCACCATGGACGCGGTCCGCGGCCGCCACGACGATGCG | | | | VGRDEGPRQDRPHP |
| | | | CTCAAGACCATCGAGCGCGCGCTCATCGCGTCGTCCGGGCGGCCAGGA | | | | WWCRERLPHAASRP |
| | | | CCGCGCCGAGCTCCGGGTCAACCAGACCGTGCCCTCGCTCGCCGGCCCC | | | | MGDRLPLCSGGSAQ |
| | | | 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 | ATCAACCCACTACTACACCCTCTACAAAGAACCCACTACAGCAAGATCTT | ATCAACCC | TAGCCA | MATSSSSVSSGNVQT |
| | | | | GGCGACCAACTACACCTGCAGCCTCCCGTCGACTCTTCACCACCTGCGCA | ACTACTAC | AAAGAT | EVRCVYHGKGDLLLE |
| | | | | CCTCCCTCCGACTCCAAGCGGCAGACAGTGGCACCACCAGCACTCCAAG | ACCCTCTA | TTGGTG | CPVAHCPSIHPTVATI |
| | | | | GGTCCACAGCCAGAAGAGTGCACCCCCCACTGCTAGGAAGAGCTAGGC | CAAAGAAC | TTCTTGG | TKHLKKHHTPQFEQI |
| | | | | CATTGCCCGCTGTGGGCCTTGTGCTGGCTGTCTGCCGCTGATTGGCACG | CCACTACA | CGAGCT | TTKNLTITYTCSQCSF |
| | | | | TTAGGGAGTGTGCGCTGAGCACGATAGGGAGGCTGCTGAAGCCGAAG | GCAAGATC | GCAGTC | STTGLTQHHISKHYKT |
| | | | | AACCCGCGCCGGGGACTTTAGGACTTTAACTAAAATCTGTAACACTAAA | TTGGCGAC | CCGGCA | CKGVGAVQEGNKGR |
| | | | | ATTGGAATACTGGATTACTATTGGATTACTACTGGACAATAACCTGATCT | CAACTACA | GAGTAG | FCCPACGTRWALLCK |
| | | | | ACAACAACCAATTGGATTACAGAACCAAATTACAATCTCAACTACAAAC | CCTGCAGC | GCGAGA | ARHHFNNVHFEYDT |
| | | | | AAAAAGTGAGTCCCCGGCCGGGTTTCTATATTACCCAATTACTGTTTTTA | CTCCCGTC | TCGTGA | PPIAAFSGTPYKLKKR |
| | | | | CAATTTTAAAATTTTATAGTATTTTACTAATTTTTCCGCCCCGCTAGCACT | GACTCTTC | GCTCTAC | KFTIINKALTYSCPLPL |
| | | | | TAAATTGGCACCCCCCCCCCCCSTCCAAAAAAAAATAGATAACCCTCCAC | ACCACCTG | ACCGCTT | NQLLCPLWSCSLTILN |
| | | | | CACCTAAACCCCGGTCACCCCCTCAGAACTAATTGGAACCCGGATCTTTT | CGCACCTC | CACGAT | KPLSSVQQETAHGD |
| | | | | CCCTCTATCAAATTTCCGGTTCAGTTTGGCTTAATTACCTAAAACCCGTCT | CCTCCGAC | TTGACG | GSQGQSYVPTQLRQ |
| | | | | TTATTTTATCCCTTTATTGCCCCTTAAACCAAATTATTCAAATTGAAATCC | TCCAAGCG | AGTTCTT | VLRARCHCGNPPIGK |
| | | | | GGCCAAATTGCCTTGCTGAACCACCATTTTTGTTTTTCTTATAAAAGTAA | GCAGACA | CTTGAG | GHWASCQGKRPLSS |
| | | | | ATTTTTCAACAACCCAAAATCATAAAATTAAAACTGTTATCTGATTAGTCT | GTGGCACC | TCGATCC | PKGGRSSPTPPANLT |
| | | | | AAATTCTGTTATTGACAAAAAACCTCCATACCAACCTTAATCTTATCATTAT | ACCAGCAC | CCAACG | LHFLNYLPFQLPSQSS |
| | | | | CTTCACCCAGACTGCTCAATCTAGTTCTTTTTTCAATCAATAAGGAGGTG | TCCAAGGG | GCAACG | SPQSSTLDPTACKAR |
| | | | | TTTTCTTTGGTCTGTTGTTTTATAGTTTACCGTTATAATAGTCTGTTGTCCT | TCCACAGC | CACTCCA | VPIPSFLRGDCEVTFFI |
| | | | | ATAGTTTTCCGTGGTGTATTTCTGGTCTACTATAGTTGGTGCCTCTAAAA | CAGAAGA | AATTTAA | IPSVNFYRPYLSYPLR |

| | | | | | |
|---|---|---|---|---|---|
| | CATTATTATACCGAAGAGTGCTAGCGATTTCTATTATTAACTCACCCACC | GTGCACCC | CGACCG | MFWRNRTSSGHCSL | |
| | AGTACTGCTGTGCATCCACGCAGCACCACGCCTGCCATCACGTACAGTG | CCCACTGC | CATGGC | HRVVRGFVRERLVPH | |
| | CGCTTGCTGCCTTGTGTTTTGTTTTGCAGACACCACTGGACGATAGATGG | TAGGAAG | GACGCC | PRSKSPARTPLEFLCE | |
| | CGACTTCGTCTTCCTCCGTTTCATCGGGGAATGTTCAGACCGAGGTGCGT | 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 | RKWLVSILNIIHEREGS | |
| | 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 | |
| | CGACCCGTACCAGCTCCCCTTACGCTTTGCCTTCTCCACCAGTCGCTGG | ACTACTGG | GTGTCC | ERRCGIPLDIIERKFRQ | |
| | TCCAATGATTTCCTGTGAGGACCACAGCGCCCCCCCTTCCGTTCGCTCGA | ACAATAAC | TTGTGA | VSMPGWIDHRRYAA | |
| | GCTCCCCCATTCCTAACTCCCCAGCCTCCGTAAGCTCCGTGGAGGCGCAC | CTGATCTA | GTAATT | GASPSLVTQAETDVA | |
| | TTGTCTGACTTGTTAGACAAAGTGTCTTCCGGAGAACTGCGGCCGCTGT | CAACAACC | ATTTTCA | ITSEEVEAVLSGLNVQ | |
| | CCCCCACCCTGCCTTCTTCAGGCTTTTTTGGCCCTCTGTTGCCACCCACTC | AATTGGAT | CTAATCC | SAPGSDGLSYRFWK | |
| | CCCCTCCCCCGACCTACCCCCCTCGGCCGAAAAGGCATCTCCTTCAGGCCTT | TACAGAAC | AGGGAC | GLDPSGRLLSCLFEIV | |
| | TCCTACCTACCCTGCAGAGAAGTAAAAATCGCGTCCATTGCACGGCCAT | CAAATTAC | GGGGCC | RRHGRIPGAWPTCS | |
| | CCCCTGCCTCACAGAGGGTCGGCTGCGACGCGGACCGAACAGGGCCTT | AATCTCAA | ATGAAC | VILLCKDAQGDVQDV | |
| | CGCTGAACCCTAATTATCAGCAAACCTCCCCCCCGTCTACTCCGAGTTTC | CTACAAAC | TGAACT | GNWRPITICRTLYKLY | |
| | TCCCCCATAGTTAGACCCCCTAAATTCCCCCGGTCGGGCGCCAAAGTAA | AAAAAGTG | ATGTCT | AAVIARRIQTWAKQ | |
| | ATTCCAAGTCCAAACCCCCCCGGCGTCCGACCACGCCGTGCCAAGCCAAT | AGTCCCCG | GCACGC | GGVLSRLQKGFMPV | |
| | TGAGCCAGGGACCGAGTCGGCATCACCCGTTGATGTCGACACTATTAGT | GCCGGGTT | TGCCCCC | 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 |
| | GGACGTTGCGGGGAGGGAGGGAGGGTCGCACTGGAGGATCCTCTGCTC | 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 | | | |
| | TGTCCTTGTGAGTAATTATTTTCACTAATCCAGGGACGGGGGCCATGAACT | CTGTTGTC | CCCTTGC | | | |
| | GAACTATGTCTGCACGCTGCCCCGCCTAGCCTCGGCCACACAAATAAATAG | CTATAGTT | GGCCGT | | | |
| | TCAAGCCGGGGCGTACTACCAATTGATTCGGATTTGGCGGGGGCCCATCA | TTCCGTGG | GAAAGC | | | |
| | CGCGCCTCCGCCCCCACACACAAAAACACTTTTAAACTCTTCGGTTCCCC | TGTATTTC | CCTTTTC | | | |
| | AACCACTACAACAAAGCGAGCGGCCCCTCTTAGATCCAATTTTAAAATTT | TGGTCTAC | ACCTTAA | | | |
| | TAAATCAGTGCACTCAACTTTTTTACGTGTTGTGTTTTTTGTATTTTTTCCC | TATAGTTG | TTCGCCC | | | |
| | ACAGATTTTGTATTTTTATATTATATTTTATATACACAACACTATTTTTATA | GTGCCTCT | CTTAGTC | | | |
| | CACTACCTTGCACTGTCCCACTTTTTGTAATTATCACCTTTTACCTTTTATG | AAAACATT | CAAATTT | | | |
| | CCGCTCGCTGAGCTCCTTGCACGGATGACCAGACAAACTTTTATAAAATT | ATTATACC | TTCCCCA | | | |
| | ATAACATTGTTTTAATTGTCGCGGGGTATCAGTGGCGCCCCCTTGCGGC | GAAGAGT | TTGCCT | | | |
| | CGTGAAAGCCCTTTTCACCTTAATTCGCCCCTTAGTCCAAATTTTTCCCCA | GCTAGCGA | GTAAAA | | | |

| | | | | | |
|---|---|---|---|---|---|
| | | | TTTCTATTA | GTGCGT | |
| | | | TTAACTCA | TGCCGT | |
| | | | CCCACCAG | CGACTC | |
| | | | TACTGCTG | GAACTC | |
| | | | TGCATCCA | GTCCTTT | |
| | | | CGCAGCAC | TTACCGC | |
| | | | CACGCCTG | CTCTCTG | |
| | | | CCATCACG | TTACTGA | |
| | | | TACAGTGC | ACTATG | |
| | | | GCTTGCTG | ACCAGC | |
| | | | CCTTGTGT | TTGGCT | |
| | | | TTTGTTTT | GTTGAA | |
| | | | GCAGACAC | GTCGGC | |
| | | | CACTGGAC | TTTAATT | |
| | | | GATAG | CGCCGG | |
| | | | (SEQ ID | CTTCACA | |
| | | | NO: 1207) | TTATATT | |
| | | | | TTTTTGT | |
| | | | | TGGGTA | |
| | | | | ATCTGTT | |
| | | | | TTTTTTC | |
| | | | | ATATATA | |
| | | | | AAACCA | |
| | | | | TTCGGC | |
| | | | | TTAAAAT | |
| | | | | CACCAA | |
| | | | | TCCCCAT | |
| | | | | CATCCAC | |
| | | | | CCTGGT | |
| | | | | CACCCAT | |
| | | | | TGAAAC | |
| | | | | ATCTCTT | |
| | | | | TTTAGTC | |
| | | | | AATTATT | |
| | | | | TTTTCAG | |
| | | | | ATTACCC | |
| | | | | CTCCCAA | |
| | | | | ATTGTCA | |
| | | | TTGCCTGTAAAAGTGCGTTGCCGTCGACTCGAACTCGTCCTTTTTACCGC | | |
| | | | CTCTCTGTTACTGAACTATGACCAGCTTGGCTGTTGAAGTCGGCTTTAAT | | |
| | | | TCGCCGGCTTCACATTATATTTTTTTGTTGGGTAATCTGTTTTTTTTTCATAT | | |
| | | | ATAAAACCATTCGGCTTAAAATCACCAATCCCCATCATCCACCCTGGTCA | | |
| | | | CCCATTGAAACATCTCTTTTTTAGTCAATTATTTTTTTCAGATTACCCCTCCCA | | |
| | | | AATTGTCATATAGTTTAACACCCCGTTTGCCAAGTTGGTCTTTTCCCCATC | | |
| | | | CCCTTCTGTCCTTCCGGTAATCCCATAACATTTTATTCTTTAATTCGGCCC | | |
| | | | CATAAATAGTGTAATCTTAAGACGAAGTCCCCAGAGGCCCGGTCCCCCCC | | |
| | | | CTTCTTTGTCATGGACCGGGCCAGCCCCCCTGTTACCACAACAACCCACC | | |
| | | | ATTTTATTCTTTCTTTTTATTAGTATTTATATTGCCGCCCCAACT | | |
| | | | GTCGTGTCGTGGCGCAGGGGGGTTCCCTCTGTCGGCCCAGTGGACGAC | | |
| | | | CGTCTAAAAAAACAGGCACAGGCAGGCACTTATCACCCGGTACCTCC | | |
| | | | GGGTACCGGAGGCTGGTTTTCAGCGCACTGTACGTGATGGCCAATTTTA | | |
| | | | TTCATTGCATTTTATCCGCGTCGTGGTGTTTGCGTGGATGCATTAATAAA | | |
| | | | AGATGAAATTCC (SEQ ID NO: 1084) | | |


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

| | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | manson i | TTTATCCACAACAACAGTTGTGATTCTATTCTCCANATATTCCTTGTGCTT | ATTTTTGA | GAATAG | | STCLICFAIFPTHNILLS |
| | | TTGTCAACATTATTCTATACCAACTGTACCACCTACTTCTTCATCTCACGT | ACTACTGC | CCCCCTT | | HATAIHHISCPPTPVQ |
| | | TTTAATTCTGGTCTTATTTTCTCATCATTAGTCACGGAGAGGGCCTATGA | AGTCTGAG | CACTCTT | | DGSQQMSCVLCAAA |
| | | ACGGTCCGTGACGCGAAATTCTATCCGCGATTTCGACCTCTCCTGCTAGT | TGCTTCTA | AGACAT | | FSSNRGLTQHIRHRHI |
| | | GGTCCCCGAAGTACGGTTCCTCTGGCCTGTCAGTTGTGTTAAAACTATAT | 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 | | PILPSDVTWALKEMIH |
| | | GAGGAAAATAGAGTACGCACACATCCAGAGACTCTTCCACCACCGTCCC | ACGGAGA | AACGAC | | GTAGGIDRLTSYDLIM |
| | | AAAGATGCCTCCAACACCGTTCTAGACGGTCGGTGGAGAAACCCCTATG | GGGCCTAT | AATTACT | | RFGKNGLAGYLNML |
| | | TCGCAAACCATTCAATGATTCCAGACTTCGACTGCTTCTGGACAACAGTC | GAACGGTC | TTATCTA | | LALAYLPTNLSTARVT |
| | | TTTACTAAAACAAATTCCCCAGACAGCCGGGAGATTACTCCAATCATCCC | 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 |
| | | TCGCTCCAAATGCCGGAACTCCTTCAACGGAAACTCGATCGGATCTCCAT | | CATTACT | DYLGIKFNFKGRCPV |
| | | ACTTCTACACGAGGCTGGATGGTCGGTTAACCCTGAAAAAAGCCGGACC | | ATTAAC | AHIDLLNNYLTEISCA |
| | | CTGGACCTAATCTCTGGTGGCCATTCCAAAATCACAGCGCTCTCTCAGAC | | ATTATTA | PLKPQQRMKILKDNL |
| | | AGAATTCACCATCGCGGGGATGCGTATACCACCGCTTTCTGCCGCCCGAC | | 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<br>ATTGCTTGCTCTGCCCGCTTACTTCCAATAGTTGTCATATTATGTCTTTGT<br>TTACTTGCCATGTCTAACGACAATTACTTTATCTACCTTAGTTTGTCCTCTT<br>GGTTTCGATTGCCTTCATATGTTCATGGCGGAATCTGATGTTTATAATGA<br>CTATTCCTATTACCACCACTACAACTACTATTATTATTTTCATTACTATTAA<br>CATTATTATAAACATTATTACTATTATTATTACTACTATTATTACTTCTACA<br>ATTAATATTATGGCTACTCCTCTCAGCACACCAATAAAATATCAATCAAA<br>CATCTCAATTATATCCACCTATTAAACTCTCTCTATTTCCCCTGAGTTATA<br>AACTTACAATTCAGTCTAACCGAATATCTCTCTTTTACAAATCTTAAGTAT<br>GTAATTTTGTGCCAAACCCATTTGGGTCTGTACAATTTGATACTTAAAAA<br>TAAATGTTATTAGCC (SEQ ID NO: 1085) | | | ACAAAT<br>CTTAAGT<br>ATGTAA<br>TTTTGTG<br>CCAAAC<br>CCATTTG<br>GGTCTG<br>TACAATT<br>TGATACT<br>TAAAAA<br>TAAATG<br>TTATTAG<br>CC (SEQ ID NO: 1331) | |
| R2 | R2-1_BTe | . | Bombus terrestris | TCAATAGTTACTCGGGGGCAGGCGGGATATTGGTCTTGCCTTGCCCAAG<br>TCACACTCCTACCTCCTCGTGGTACCGCCGGTAACACGCGCACGTCCACA<br>TCAGCGAGGGGCGTACTCCCCCGGATGTGGCGGCGCGTGGCTAAACGG<br>AGTGTGGCGACGAAGGAGCGAAAGACTAACAACTATAACGGTCTTCCG<br>TAACGGCTACTTGGAGCCGTGAATAATGGAGCCTATATTAAACCCTGGA<br>ACTCGTTCCTTCGTTCTGTTGACGACTGGAACGGCAAAGGACATGATTT<br>GGATAACAATTGGAAACTTAATATCGAATTCACAATTAAATTGCAAAATT<br>CGATAATAATACGAATAGTGCCTCGGACGCAGCGCCTCTCAGTCCCGGA<br>GGCGCTGTGGCGGATTTATCCGCTTCAGAGGGGACTACGGATAACGAC<br>CAGGCTATGTCCCCGGCTATGAGTTTGATNACGGTGCCGCTCGTGGGCA<br>ATCGGGTCGCCTGTCCAGNATGTGAGAAAAGGGAGGCTAACCTTTTCTT<br>CTTAAATCTGTCGGACCTGGACAGGCACCTAACACAACACCACCCGGAT<br>GCCCCGATCTNTTGGTCNTGTATCGACTGCGCGAAATGCTTCCCGAAGC<br>TTCATGGGGCTAGATGCCACATACCTAAGTGTGGTGGCGCTAGCAGCCA<br>GGCCAGGACAGGNGAGTTCCAGTGCGAGGCCTGTCCCATGAGCTTCGG<br>GTCGCGCAGAGGGCTATCCACCCACGAACGGCATGCGCACCCTGCCGTC<br>AGAAATATTAAAAGAAGGGGAGCGGACCCCCCAGAAGAAAATACAAAA<br>TCTTGGAAAGTAGAANNGGTAGCACGCTTGAAGGGGCTATGGGAAATA<br>TTCAAAGACCATAAATATCCCAACAAAGAAATCAGCAAATTCCTCACCAC<br>NAAAACGGTNGATCANATNAAATATCAAAGAAAAAAAGCTAAACTTAAT<br>TGGTGNGGAAAGCCCCCAAGAGGCTACCTCACTGGCAACAGAGGGAG<br>GGTGCGATCTCGTTAGTTCAGGCAATGCCAGTTTTGGCTCGCCTGTAGG<br>CCGCAACGAGAACGAAGAAGAGCTCATCCACGAATGGAAGCTCTCGCT | TCAATAGT<br>TACTCGGG<br>GGCAGGC<br>GGGATATT<br>GGTCTTGC<br>CTTGCCCA<br>AGTCACAC<br>TCCTACCT<br>CCTCGTGG<br>TACCGCCG<br>GTAACACG<br>CGCACGTC<br>CACATCAG<br>CGAGGGG<br>CGTACTCC<br>CCCGGATG<br>TGGCGGC<br>GCGTGGCT<br>AAACGGA<br>GTGTGGC<br>GACGAAG<br>GAGCGAA<br>AGACTAAC | TGAAAA<br>CACGAT<br>AACGAT<br>TATGAA<br>TCAAAA<br>TAAGAA<br>AGTAAA<br>CATCCCA<br>GAAATT<br>GTCTAC<br>GTCTTAT<br>TTGTTAT<br>CTATTTA<br>TTTGTTT<br>A (SEQ ID NO: 1332) | IAKFDNNTNSASDAA<br>PLSPGGAVADLSASE<br>GTTDNDQAMSPAM<br>SLXTVPLVGNRVACP<br>XCEKREANLFFLNLSD<br>LDRHLTQHHPDAPIX<br>WSCIDCAKCFPKLHG<br>ARCHIPKCGGASSQA<br>RTGEFQCEACPMSF<br>GSRRGLSTHERHAHP<br>AVRNIKRRGADPPEE<br>NTKSWKVEXVARLK<br>GLWEIFKDHKYPNKE<br>ISKFLTTKTVDXXKYQ<br>RKKLNLIGXESPQEAT<br>SLATEGGCDLVSSGN<br>ASFGSPVGRNENEEE<br>LIHEWKLSLKNEINKP<br>TEVPPILKEVYNRLML<br>IWEEHQDDRDSLTES<br>LDHFIRTALYELINKIN<br>KNQTDLKTKRAAKTK<br>SPKNNRNSRKRFSYA |

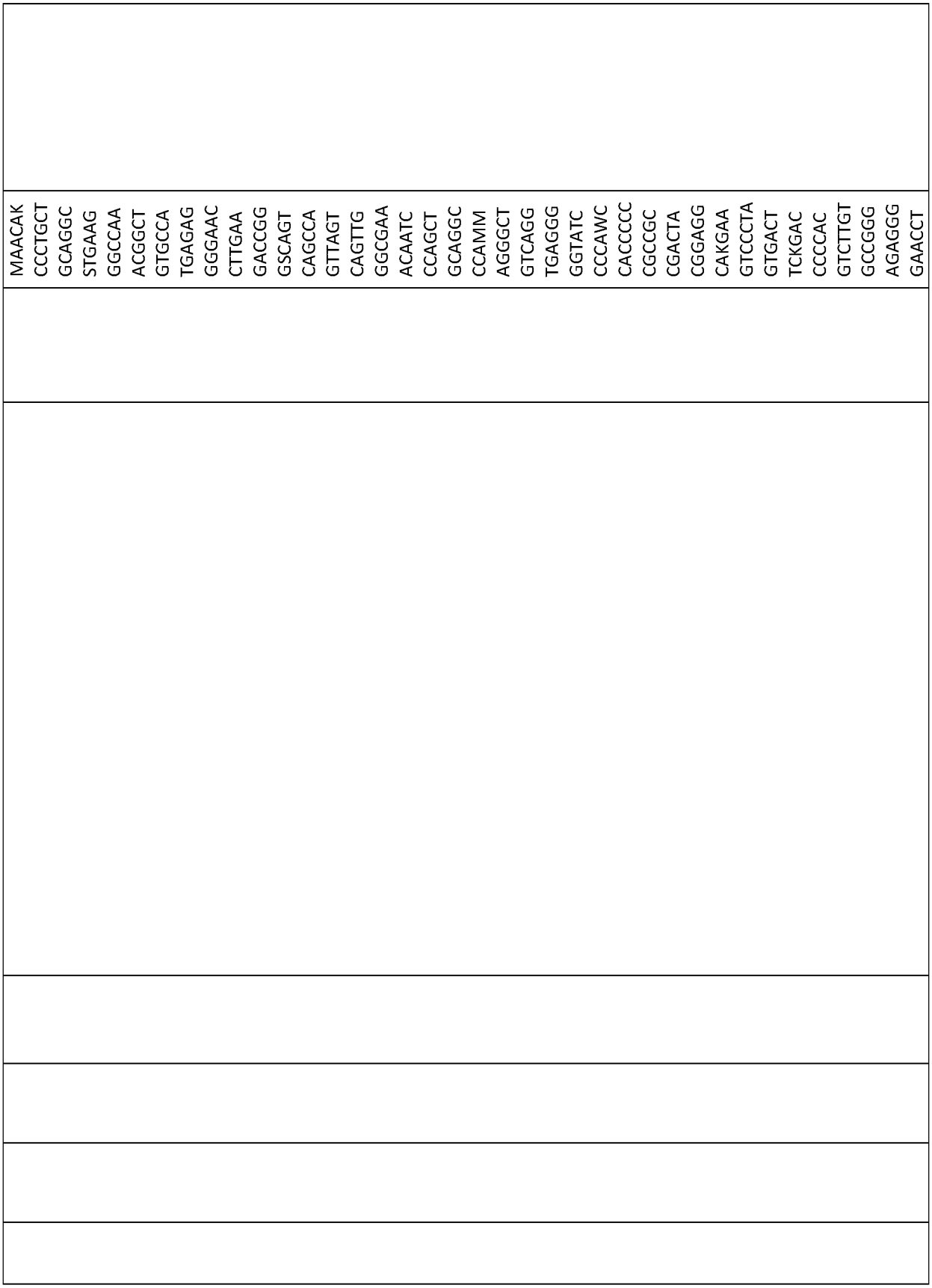
| | | | | | |
|---|---|---|---|---|---|
| | | AAAAAACGAAATAAATAAGCCAACCGAGGTGCCCCCTATTTTAAAGGAG | AACTATAA | | RCQELFHECPRRLAD |
| | | GTTTACAATCGGCTGATGTTAATCTGGGAAGAGCACCAAGATGATCGAG | CGGTCTTC | | AVVNNDQAYLEPAR |
| | | ACTCTCTAACGGAGAGTCTCGATCACTTTATACGCACAGCTCTATACGAG | CGTAACGG | | QPPGSEEVRGLYEKL |
| | | 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 |

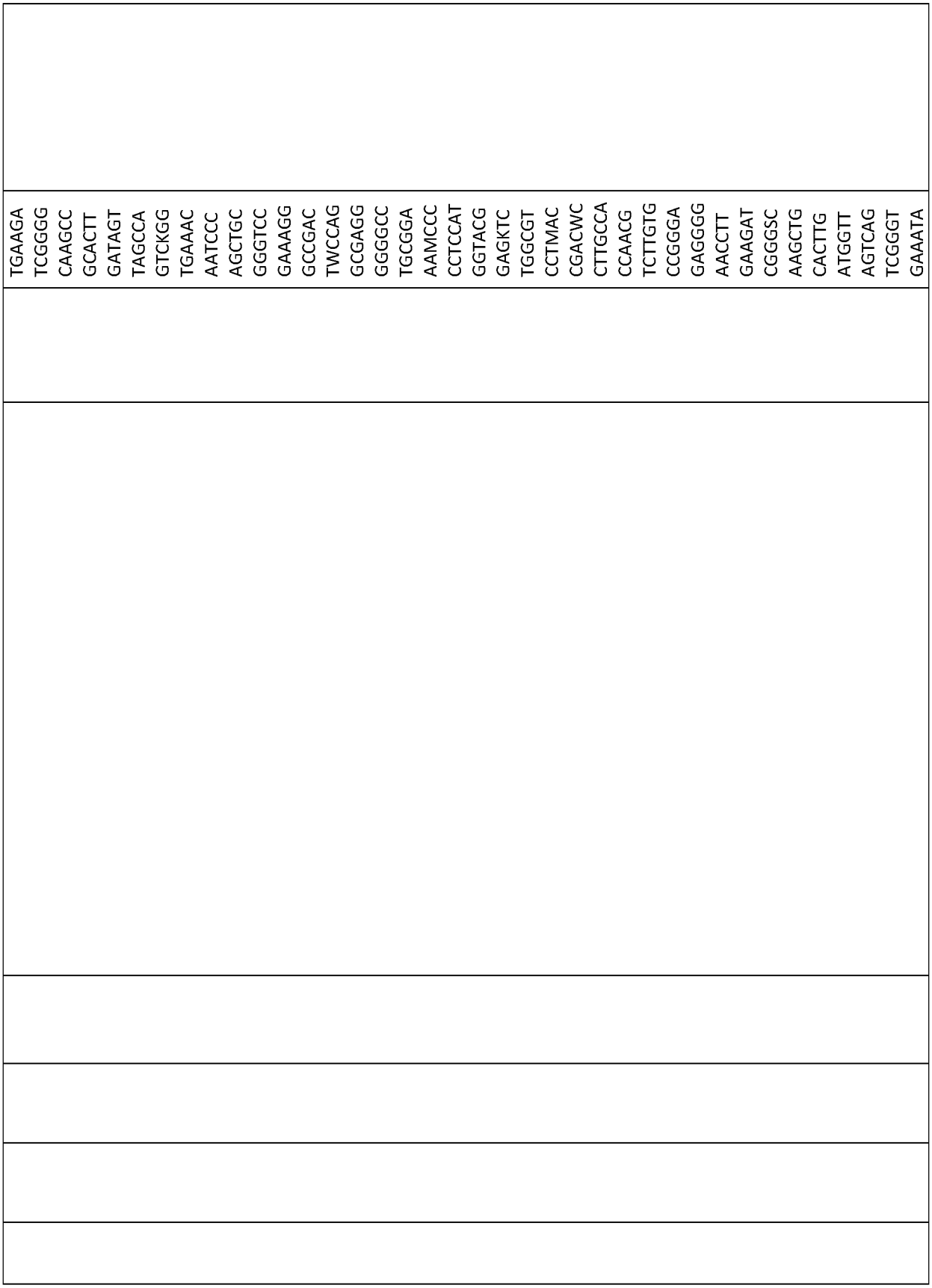
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CCTCCTTGGAGGATATTGAAAAAGCTCGTAAACGTCTAAGGAAAGAGCA | | | | RLIDALRLRTNTFGTR |
| | | | TATCAGCCAATGGGCTGATTTAAAATGCCAGGGACAAGGCGTTCCTGAT | | | | SVLARADKNIDVTCR |
| | | | TTTATCAAAAATAAAACTGGCAACCTGTGGCTTGAGGACCATAGTCTAC | | | | RCRAQPETLGHILGLC |
| | | | TCAAGCCATCGAGACTCATCGATGCCCTTCGATTAAGAACAAATACCTTT | | | | QHTKGLRIKRHDEVK |
| | | | GGTACAAGATCGGTGCTGGCACGGGCCGACAAAAATATTGATGTAACA | | | | SLLEGRLKSKKNNEVF |
| | | | TGTCGAAGATGTCGTGCCCAGCCCGAGACCCTTGGACACATACTCGGGC | | | | VEPTIKAGGSLFKPDL |
| | | | TATGTCAGCACACTAAAGGCTTAAGAATCAAAAGGCACGATGAGGTTAA | | | | VIKNGERVLVVDVTV |
| | | | ATCCCTCCTCGAAGGAAGATTAAAAAGTAAAAAGAACAACGAAGTATTT | | | | RYENKNYLALAEKEKI |
| | | | GTAGAGCCGACGATTAAGGCCGGAGGCAGTTTATTTAAACCAGACCTC | | | | EKYRPCLRALKEIFNA |
| | | | GTAATTAAAAACGGGGAAAGGGTTCTCGTGGTCGACGTAACTGTCCGCT | | | | 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 |
| | | | | CCCGGTGCWGAGGCCCGGGGCCGATACCACCACCCCCGGTGMGAGM | 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 |

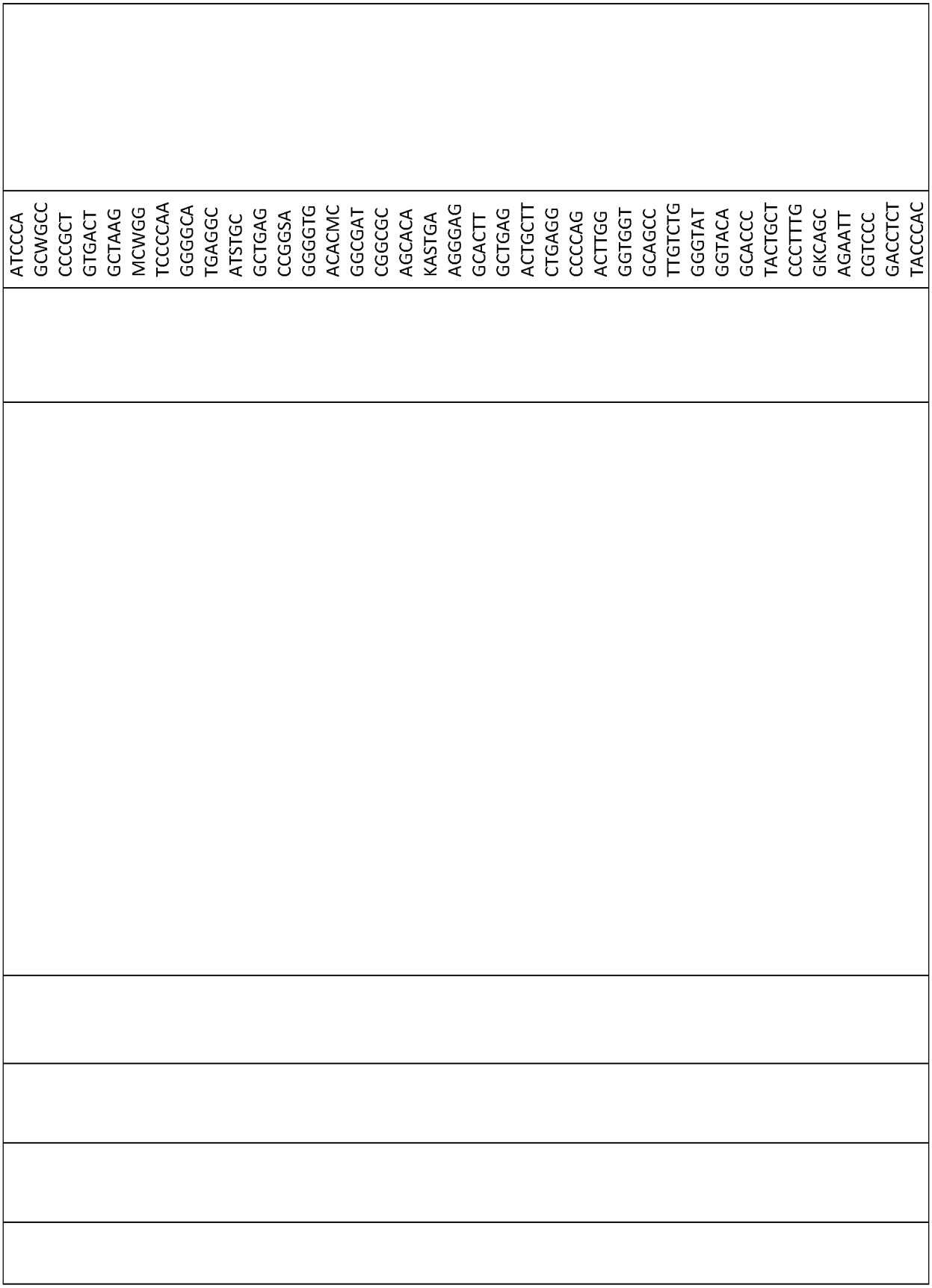
| | | | | | |
|---|---|---|---|---|---|
| | CGACAAATGGAGGTWGCTCKSTCTGWGTCCTGAGCAGACCACATWGG | GTGCCCCA | AATGAC | | LGVPPGPQRAPSKKG |
| | GTGGTGAKGCTGAGTCKGCCTCCGTGGTGSAGGWWGAGTCCATGACT | CAAGCGTK | CAGTGG | | PXEGGKPTSWMNKR |
| | CCGGAGATGGAAGCCCAAAGCCCAGCTGAMCCGCCTGGGAAAATCAG | TGGCAGS | TGTTGA | | 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 |
| | GGAGGCTGTCAAGCACACATGATGGMAATGGCCWMSAAMTCGGCTCCA | TACAGATG | GCCCAG | | LXSRVLTARLTAACPI |
| | GGTCCGGATAAGCTCACCCTGAGAGATCTGCGCCGCGCKGACCCCGAG | 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 |
| | TTGTTCAACATTGCGATGGAACCCCCTGATAGSCAAGCTCAAGACAGTCA | 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 |

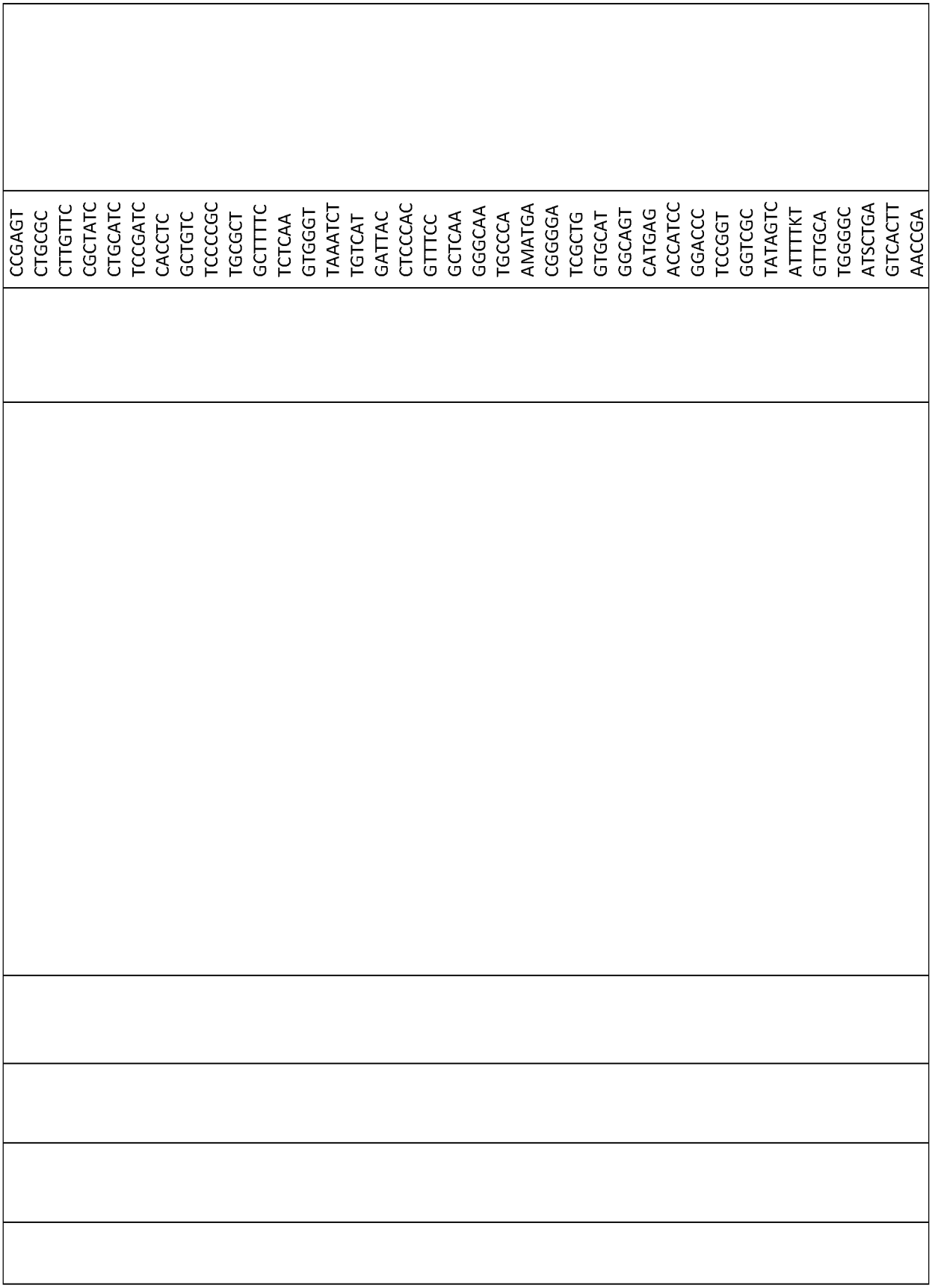
| | | | | | |
|---|---|---|---|---|---|
| | | | CKGCCTCCTCCASTCCCGCCACCGGGATGGGGGCCTKGGGTTGCTGCGC | ACCCTCC | GISDEVERLXLAXGG |
| | | | CTGGCGAGAGMCWTTCTGGCAGCGSAAGTGAGGAGGCCGATTCGTGT | GCMTCT | DXSNVPRFEDPXAPK |
| | | | AGCCACCTCATCTGATGAAGTCACGCGCGGAAAGTATCCTACGCATGTGGG | GCGCTT | SXXVQGPHEAAQET |
| | | | ATAAGTGATGAAGTGGAGCGKCTCTGKTTGGCMAGWGGGGGGAGACAT | 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 |
| | | | CTGCCATGCAGGAGGCACGGATCCTCCGGCACAACAAGCTGTGCAAGA | 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) |
| | | | CCCACTCTGGGGACCCCCAAAAATTAGAAAAAACCCAAAACAGTTGTGTT | AGGGAA | |
| | | | TAAGTGTGTTCTTGTTTGTCCCTTTGGCTTCACCTCCAAGTTGCGATCCCC | AGGGAG | |
| | | | CATCTCCCCTGCGCTGTCTTTCTGAATGACCAGTGGTGTTGAGGCTGGTG | CCGTCCC | |
| | | | TGACCTCGGTCACCTCCAAGGCCAAGTGCCCTGGCCCCGAGTAGGACCA | TGACCG | |
| | | | GGTGGCCCAGCTTGCTGGGCACCCGTCACCGCCCAGGGCAAATGGAAG | TTACCG | |
| | | | GGATCTATCCTGACCACTACCAGGCTAAGTGTGGTGCTGCCTAGCCTGC | GGCTTG | |
| | | | CGTAAGGTCAAGCGCCCTGCTGCCGCTCGGGTAKCAGTCCTCGTTCACT | ATGGTG | |
| | | | CGTCCCTCCTAGTACCCTCCGCMTCTGCGCTTCTGCTATCCACTGTGGCC | CTGGCT | |
| | | | GGWGATGCCGAGGWTGSWMCAWCCTCGACACCTCCAGGGCCAAGC | AGCCCG | |
| | | | GCCTTGGCCMCAGGTAGGACCTGGCACCTGCCCAGGGGGGCCAGACACT | CCGTAT | |
| | | | GCCTGCGGCAAGGGAAAGGGAGCCGTCCCTGACCGTTACCGGGCTTGA | GTCAAG | |
| | | | TGGTGCTGGCTAGCCCGCCGTATGTCAAGCACTCCACAGCTGCTCGAGT | 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 |
| | | | GAGGCACTTGCTGAGACTGCTTCTCTGAGGCCCCAGACTTGGGGTGGTGC | | 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 |
| | | | | AAGACGAACTAGATACGCAACGTCAGCGTGAAGATGTAAAACTCCGAT | 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 |

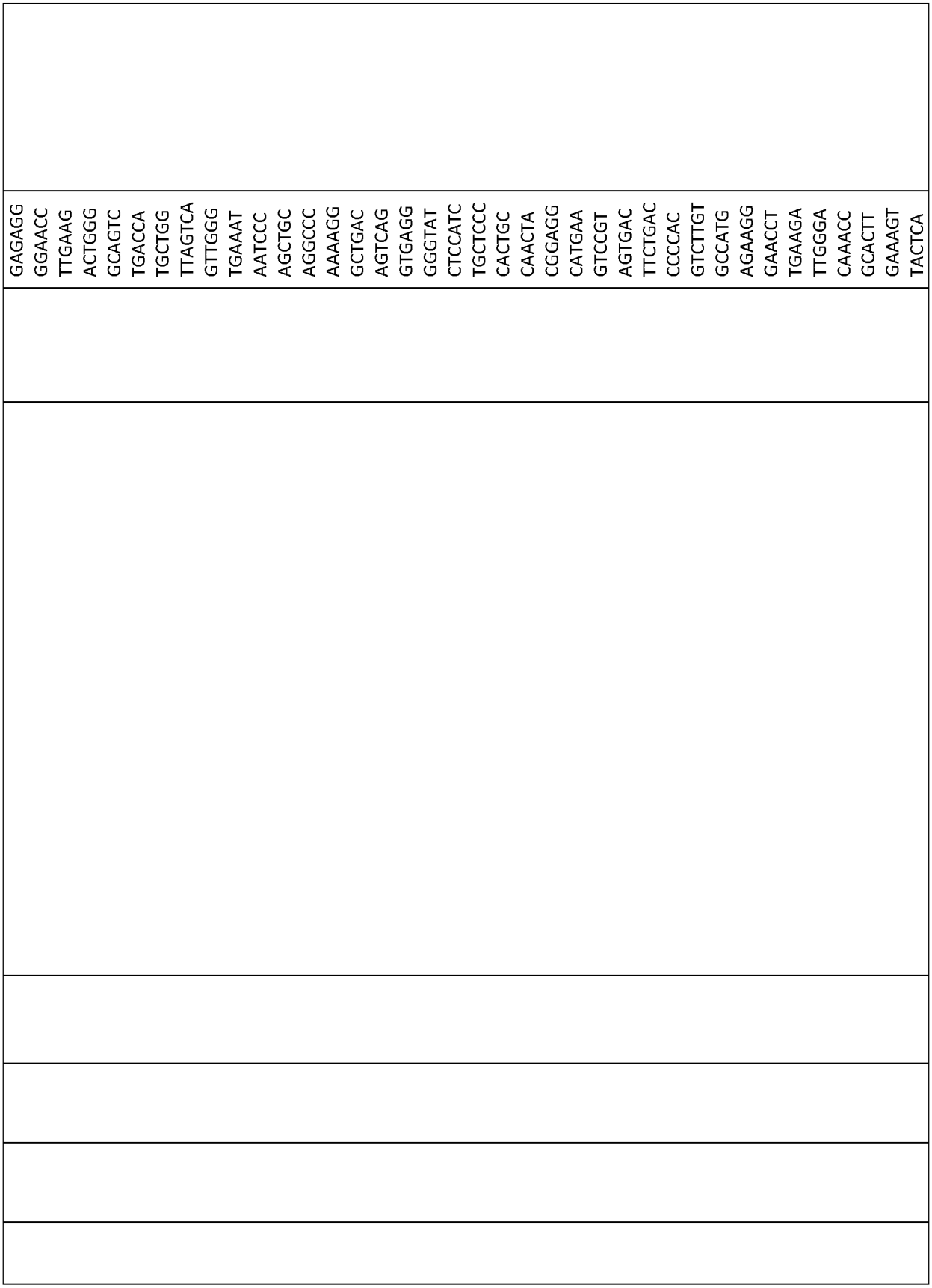
| | | | | | |
|---|---|---|---|---|---|
| | | CAAAATACCATTCTCTACCCGACTGGCCAGAACTATCTTCATTCCGAAGA | | TGGTTG | ANRPQDFRPITVPSV |
| | | CTGTGACGGCAAATCGACCGCAAGACTTTCGTCCAATAACAGTCCCCTC | | CTCGTAT | LVRQLNAVLASRLAS |
| | | GGTTTTGGTCAGGCAATTAAACGCTGTTCTGGCTTCTCGATTGGCTTCTA | | CTTTAGA | KVNWDPRQRGFLPT |
| | | 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 |
| | | AGGTCATTATGGCCCGCAACATGGATGGGTGAGTCAGCCCACGCGCTT | | | YMSVDGCGLREAGH |
| | | GCTAACAGGAAAGGAATATTTGCACGGTGTCAAATTGCGGATAAATGCC | | | YGPQHGWVSQPTRL |
| | | CTACCCTCGAAGTCTCGTACGACGAGGGGAAGGCACGAATTGGAGAGA | | | LTGKEYLHGVKLRINA |
| | | CGGTGTCGTGCAGGATGTGATGCTCCCGAGACAACAAACCACATCTTGC | | | LPSKSRTTRGRHELER |
| | | AAAAATGCTATCGTACGCATGGGAGGCGGGTAGCTAGACACACAACAGCG | | | RCRAGCDAPETTNHI |
| | | TAGTAAATGCCGTCAAGCGGGGACTTGAACGGAAAGGCTGCGTTGTCC | | | LQKCYRTHGRRVAR |
| | | ATGTCGAACCAAGTCTGCAATGCGACTCGGGCTTAAATAAACCGGACCT | | | HNSVVNAVKRGLER |
| | | GGTGGGAATCCGACAGAATCACATTTATGTGATAGACGTTCAGGTTGTG | | | KGCVVHVEPSLQCDS |
| | | ACAGACGGACATTCCTTAGACCAAGCGCACCAGCGCAAGGTCGAAAGG | | | GLNKPDLVGIRQNHI |
| | | TACGACAGAGCTGACATAAGATCACAAATGCGGCGATTTTTCGGAGCGA | | | YVIDVQVVTDGHSLD |
| | | CAGGTGAAATCGAGTTTCATTCCGTTACACTCAACTGGAGAGGAATCTG | | | QAHQRKVERYDRAD |

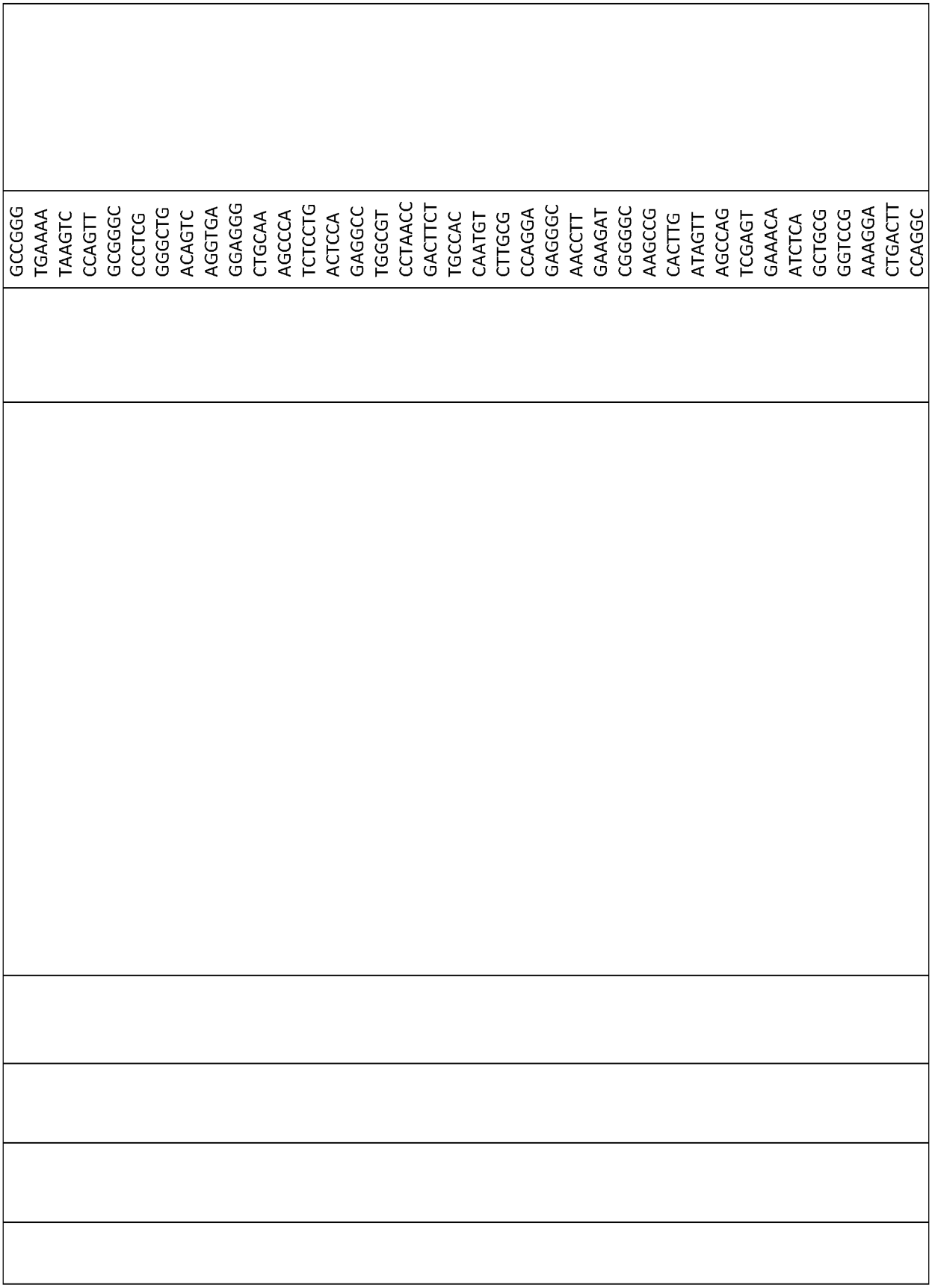
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GAGTGGTCAGTCGGTAAAACGATTGATTGCAAAGGATCTCCTCATCGCT<br>GAAGATACCAAACTCATCAGCGTCAGAGCAGTAAATGGCGGAGTGACG<br>TCCTTCAAATATTTCATGTATTGTGCTGGGTATACTCGAAGCTAGATGTA<br>CTAACCTCTAGCTTTTCTTATACTTTTGCCTGCTACCTTGGCATTACATCTA<br>AAAAGGTACAAACATCGCATTGTCATAAAGAGGTGGTTTTAGTACGTAG<br>GCGCTGTGGGACTTCATTGTCCCGGTGATGCAGTGAATCGTGCATACGA<br>GATTGTCCAGTAGTTGGTTGCTCGTATCTTTAGAAGATTTCCTTCCTCGG<br>CGATCAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1088) | | | | IRSQMRRFFGATGEI<br>EFHSVTLNWRGIWS<br>GQSVKRLIAKDLLIAE<br>DTKLISVRAVNGGVT<br>SFKYFMYCAGYTRS<br>(SEQ ID NO: 1456) |
| R2 | R2-1_Gav | . | Gavialis gangeticus | AGGCATCTCCTTKAAGGGTAATGGTCTGGTTACATGGTCATAGCAGGTT<br>TGTGTCAGGTACCTCCCAGTGGTTCCCGCCGGGTGSCAMAGCCCCAGG<br>GCTGTCGGTAGCTCGATCCTGGTACAGTACGGCCAAGGGAGTTCTTCCT<br>TGCTGTCGGGTGCCTCGCAAGCACKTGGCAGCCCCAATCGCTTCATTGC<br>GAAAAACACAAACGTCCTAAGGGGATGATCAGCTAGTTCAGTTCTGCCGC<br>TAGCCAAAACTGTTTGCCACCCAGTTACAGATAGCGTCTGTGCTGACCA<br>GCTGCCCCGCGGGCTTGGGGTGCAGTGGAGGCCGGCCCGTGGCCAGG<br>CCGGACACGGGCCGTGGAGCCTGCTCCCAGTCCAGAGAGTTCCCCCTCG<br>GAAGCTGCCAGGGCAGCACCAGCCGGGGAGGGCCACGGCCCCGGTCA<br>TGAGTCCCCCTCGGTGCAGAGGCCTGAGGCCGATACCACTGCCCCTGGT<br>GTGAGCGCGCCCACCAGGGAGGGTGAACCACCCTCCACCAGGGTTTTC<br>CTGGTGAGGCTGCCTGATTCAAACCCGCCGTGCCCAATTTGCAGGGACC<br>ATGTGGGTAAACCCTCCGCGCTGGCCCTCCACTGCGTGGAGAGCCATGC<br>GTGGGCGGATGTGCAATACCAGTGCACCCATTGCAAAAAGGTCAGTGC<br>TAACAAACACAGCATCCTCTGCCACATCCCATGTTGCCAGGGGAGGGTG<br>CCCGAGTGGACCGGGAAGGACTGGGCTTGTCCTGAGTGCCCTGCCTCCT<br>TCAATAAGAAAGTTGGCCTATCGCAGCACAAGCGGCACGTGCATCCGGT<br>AACACGTAATGTGGAACGGGTTGCAGGGAGCCTATCGAGGGCTGGTTT<br>AAGGCCCCAGACCAGACGCGGGTGCTGGTCGGTGGAAGAAGAGGAAA<br>CTCTCACCTGCCTAGACGCGATGTTCCGTGGTGCCCGGAACATCAACCA<br>GCTGATCGCTGCTGAAATGGTAACGAAGATGCCTAAGCAGATCAGTGA<br>CAAACGGAGGCAGCTCGGTCTGTGTCCTGAGCAGACCACATTGGGTGG<br>TGATGCTGAATCGACCTCCGTGGTGGAGGAAGAGTCCATGACTCCGGA<br>GATGGAAACCCAAAGCCCAATTAACCCGCCTGGGAAAATCAGGAAGAT<br>CCTGGCCCAGAGGGCACGCCGGTGGCTGAAGAAGGGGCAGGGTCTGT<br>CGGACAAGGTGCGAGAAGTCCTGGGCGCATGGGTGGAGGGTCAACCC<br>AGGATTCATGCCTGGGTCGACTCAGTCTCCCTTGATGTTTTGACTTTGTT<br>CTTGGGGGTGCCCTCAGGACCGCAGAGGGCTCCGAACAAGAAGAGGCC<br>CAAGGAGGGTGGCAAACCAACGTCCTGGATGAACAAATGTGCCGTCAA<br>ATGGGGCACATTCCTTCGGTACCAGCACCTGTTTGGTGCCAACAGGAAG | AGGCATCT<br>CCTTKAAG<br>GGTAATGG<br>TCTGGTTA<br>CATGGTCA<br>TAGCAGGT<br>TTGTGTCA<br>GGTACCTC<br>CCAGTGGT<br>TCCCGCCG<br>GGTGSCA<br>MAGCCCC<br>AGGGCTGT<br>CGGTAGCT<br>CGATCCTG<br>GTACAGTA<br>CGGCCAAG<br>GGAGTTCT<br>TCCTTGCT<br>GTCGGGT<br>GCCTCGCA<br>AGCACKTG<br>GCAGCCCC<br>AATCGCTT<br>CATTGCGA<br>AAAACACA<br>TGATCAGC<br>TAGTCAGT | TGAGAA<br>GTTGCG<br>AGTTCTT<br>ATGCAA<br>GTTGAA<br>TACCACT<br>CTKGKG<br>ACCCCA<br>AAAAAW<br>WAAACC<br>CCAAAA<br>CAGTTG<br>TGTTTAA<br>GTGTGT<br>TCTTGTT<br>CGTCCCT<br>TTGGCTT<br>CACCTCS<br>AAGTTG<br>CGATCC<br>CCCCATC<br>TCCCCTG<br>CGCTGC<br>CTTTCAG<br>AACGGC<br>CGGTGG<br>CGCGAC<br>CTCGGT | PAAPRAWGAVEAGP<br>WPGRTRAVEPAPSP<br>ESSPSEAARAAPAGE<br>GHGPGHESPSVQRP<br>EADTTAPGVSAPTRE<br>GEPPSTRVFLVRLPDS<br>NPPCPICRDHVGKPS<br>ALALHCVESHAWAD<br>VQYQCTHCKKVSAN<br>KHSILCHIPCCQGRVP<br>EWTGKDWACPECPA<br>SFNKKVGLSQHKRHV<br>HPVTRNVERVAGSLS<br>RAGLRPQTRRGCWS<br>VEEEETLTCLDAMFR<br>GARNINQLIAAEMVT<br>KMPKQISDKRRQLGL<br>CPEQTTLGGDAESTS<br>VVEEESMTPEMETQ<br>SPINPPGKIRKILAQR<br>ARRWLKKGQGLSDK<br>VREVLGAWVEGQPR<br>IHAWVDSVSLDVLTL<br>FLGVPSGPQRAPNKK<br>RPKEGGKPTSWMNK<br>CAVKWGTFLRYQHLF<br>GANRKLLVAIVLDGA<br>DRNQCTLLLEEVFQA<br>YREKWGLEEVLRAYR<br>GKWEVESSFEGLGRF |

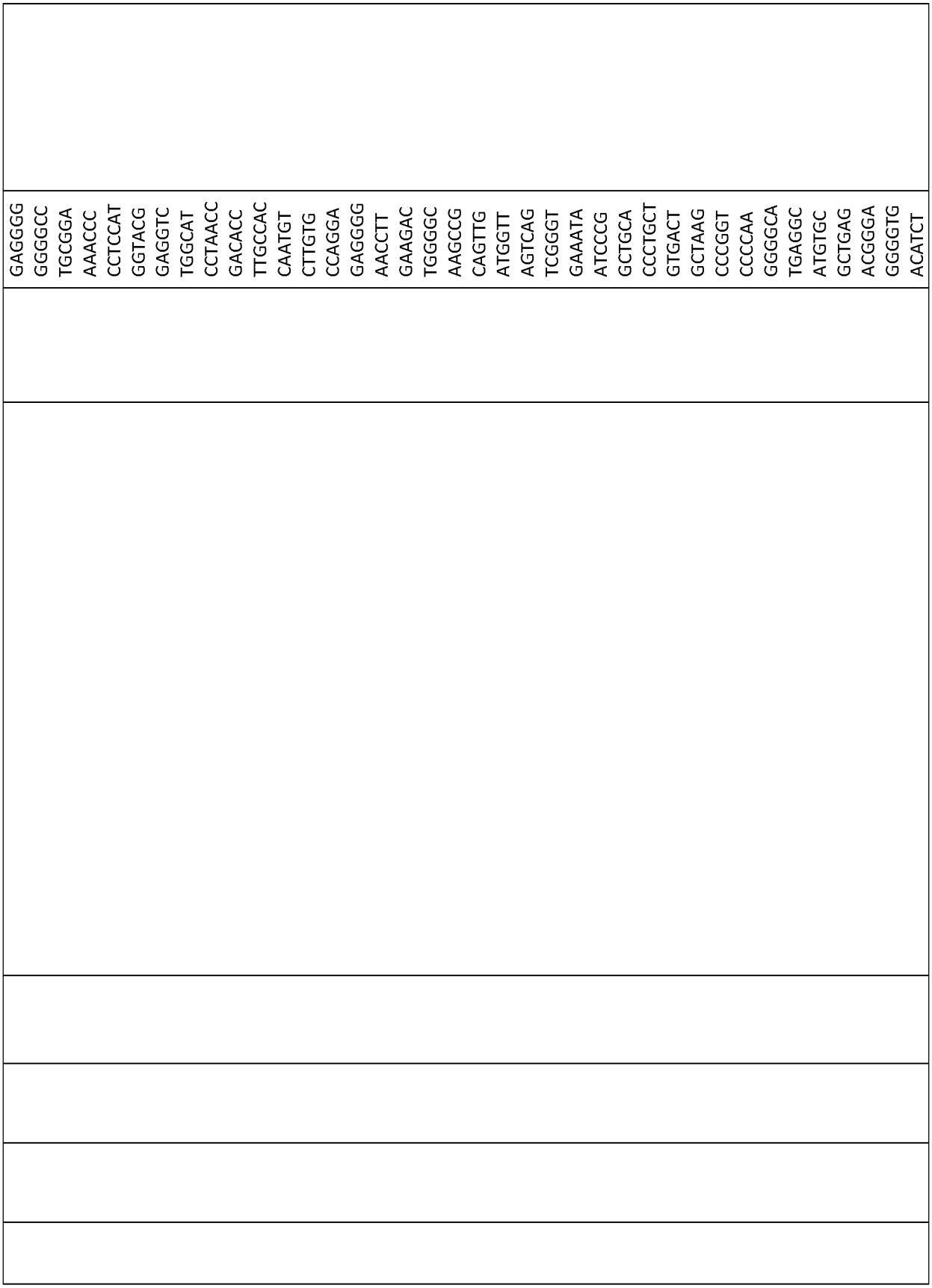
| | | | | | |
|---|---|---|---|---|---|
| | CTCCTGGTGGCSATCGTCCTGGACGGCGCTGACCGTAATCAGTGCACCC | TCTGCCGC | CACCTCC | GVRRDADNFAFKALI |
| | TCCTGCTAGAGGAGGTCTTCCAGGCCTACCGAGAAAAGTGGGGGCTAG | TAGCCAAA | AAGGCC | TPEEVVKHMMAMA |
| | AGGAAGTCCTTCGGGCCTACCGAGGAAAGTGGGAGGTAGAGTCATCTT | ACTGTTTG | AAGTGC | SKSAPGPDKLTLRDLR |
| | TTGAGGGTCTCGGACGGTTCGGGGTCCGCCGGGATGCGGATAACTTCG | CCACCCAG | CCTGGC | RADPEGDALAELFSL |
| | CATTCAAGGCCCTGATCACTCCTGAGGAGGTTGTCAAACACATGATGGC | TTACAGAT | CCCGAG | WLITGTVPDGLKECR |
| | AATGGCCTCGAAATCGGCTCCAGGTCCGGATAAGCTCACCCTGAGAGAT | AGCGTCTG | TAGGAC | SVLIPKTVDREKLGQL |
| | CTGCGCCGCGCTGACCCCGAGGGAGATGCTCTTGCCGAACTCTTCAGCC | TGCTGA | TGAGTG | GNWRPITIGSIVLRLF |
| | TGTGGCTGATTACCGGCACGGTCCCGGACGGACTCAAGGAGTGTCGGT | (SEQ ID | GCCCAG | SRVLTARLAAACPINP |
| | CTGTGTTGATACCCAAAACGGTGGACCGGGAGAAGTTGGGCCAGCTGG | NO: 1212) | 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 |
| | GAAGCTCGAGACAGACGGAGTAAAAGTCGGGAGTGCCTCCCT | | 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 |

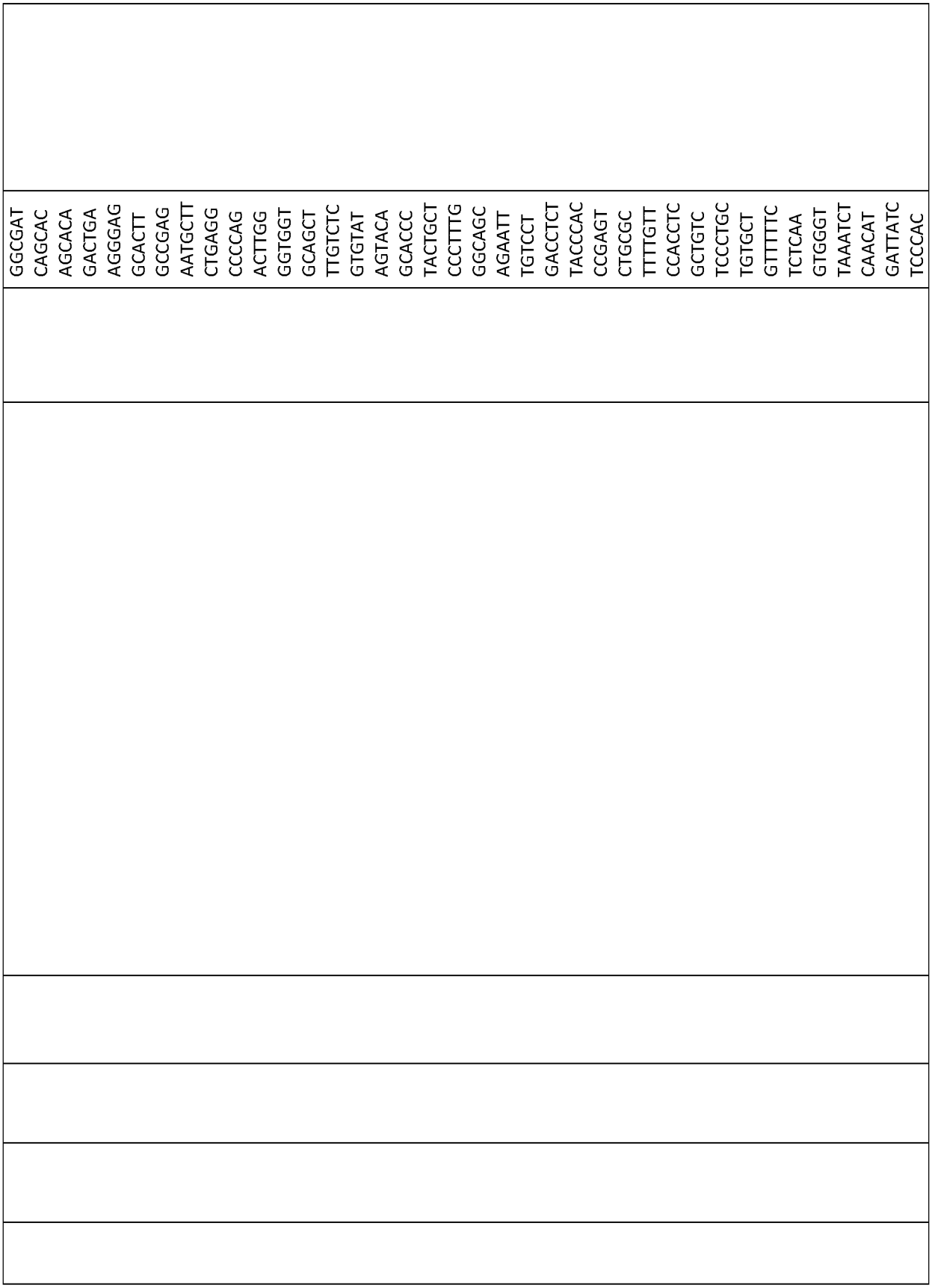
| | | | | | |
|---|---|---|---|---|---|
| | | | | TGATGTT | VALKLRSNIYPTREFL |
| | | | | GAGGTT | GRSQASTNVGCRHC |
| | | | | GGTGCA | THPHETLGHILGICPA |
| | | | | TCCTTG | VQEARIIRHNKLCKIL |
| | | | | GTCACCT | AAEGKKCEWTVYYEL |
| | | | | CCAGGG | QLLNAAGELCKPDLIF |
| | | | | CCAAGC | VRDGTXLVVNVTVG |
| | | | | GCCTTG | YEGGPAXLLSTAAEK |
| | | | | GCCACA | ATKYLDXNAQIQELT |
| | | | | GGTAGG | GAEQVTYFGFPIGAR |
| | | | | ACCTGG | GKWHADNXRVLSEL |
| | | | | CACCTG | GLSNSRKERVARLLX |
| | | | | CCCAGG | WRALLGSVDMVNIF |
| | | | | GGGCCA | ASKHRQENLSDXALS |
| | | | | GACACT | PS (SEQ ID NO: |
| | | | | GCCTGT | 1457) |
| | | | | GGCAAG | |
| | | | | GGAAAG | |
| | | | | GGAGCC | |
| | | | | GTCCCT | |
| | | | | GACCGT | |
| | | | | TACCAG | |
| | | | | GCTTGA | |
| | | | | GATGGT | |
| | | | | GCTGGC | |
| | | | | TAGCCC | |
| | | | | ACCATAT | |
| | | | | GTCAAG | |
| | | | | CACTCCA | |
| | | | | CAGCTG | |
| | | | | CTTGAG | |
| | | | | TTTGTTG | |
| | | | | GCTTCAC | |
| | | | | CTTCATC | |
| | | | | CCACCTA | |
| | | | | GTGTCTT | |
| | | | | CTGCCTC | |
| | | | | TGCACT | |

| | | | | | |
|---|---|---|---|---|---|
| | | | GGGCTGCAAAGCCCATCTCCTGACTCCAGAGGCCTGGCGTCCTAACCGA | | ATTTTCA |
| | | | CTTCTTGCCACCAATGTCTTGCGCCAGGAGAGGGCAACCTTGAAGATCG | | TCCCAAC |
| | | | GGGCAAGCCGCACTTGATAGTTAGCCAGTCGAGTGAAACAATCTCAGCT | | TCGTACC |
| | | | GCGGGTCCGAAAGGACTGACTTCCAGGCGAGGGGGGGGGCCTGCGGAA | | 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 |
| | | | CTGTAAATAACTCAAAAGAGGTACCCTCCGGGGTTCGGTAAA (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

| | | |

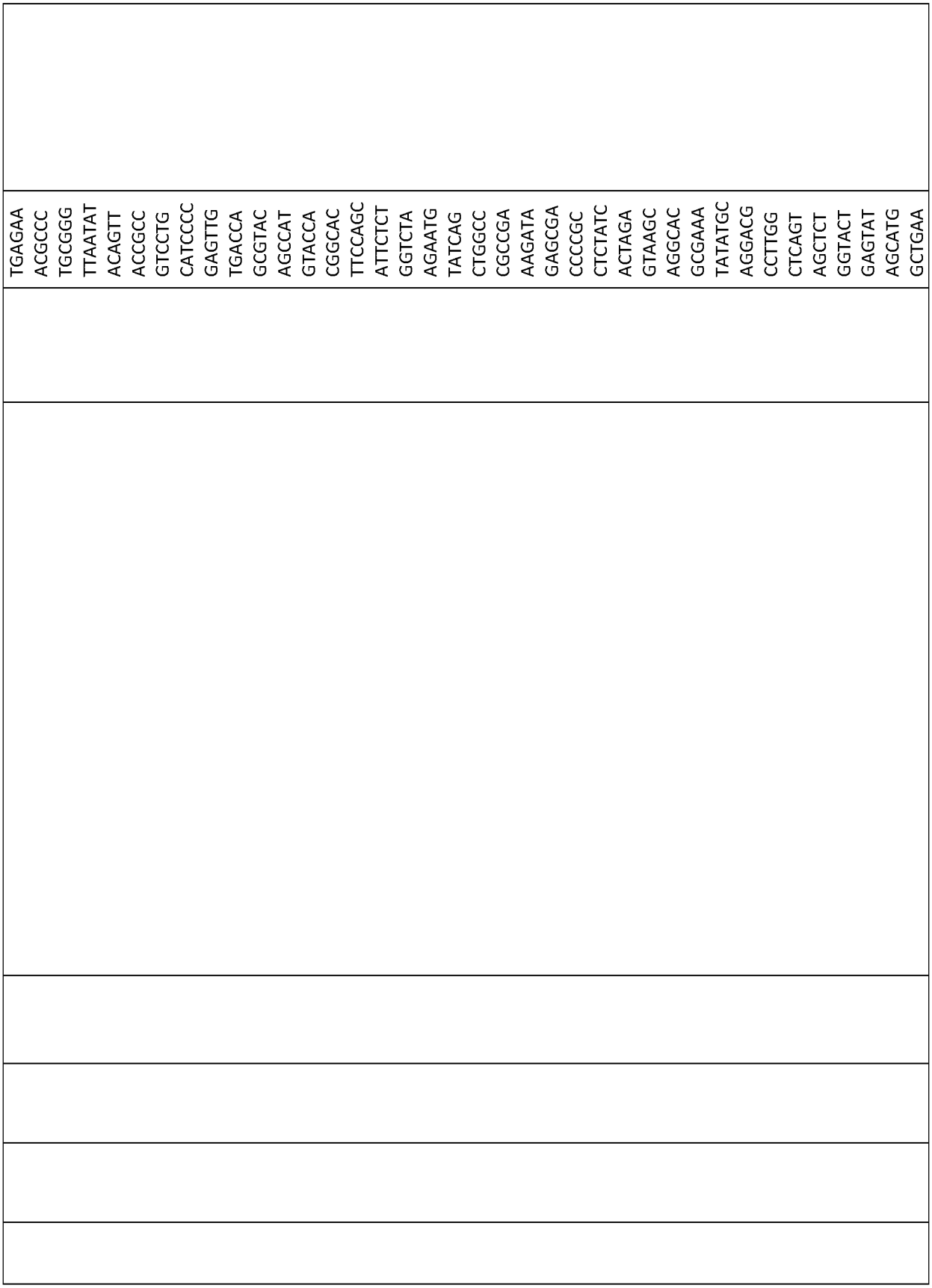
| | | | | | |
|---|---|---|---|---|---|
| | TGGGCGTCGAGGGTGAAGGTGTCGGTGCGTCTGGTAACGGCACTGATG | GCTTGCCT | CCAGGT | SDRVTTDPVPAPVPE |
| | CGCAGTTATGGTATGGCTGCACGGGCTGTGACGAAGCCTTCTCGTCCCT | AGCTGGA | GGCAGG | SPGLLPGLPGPSLPGC |
| | GCGAGGACTGAGAATTCATGCGGCCCAAAGAAACATGGAAACCAAGA | GAAATCCG | AAAGGG | 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 |
| | AACGCGGGCATTAACATCACCGTACCAAAGCCCGGTGGGACCCCGAG | 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 |
| | CATCGGCCGAGCACCACCAGGCTCCTTCCAGGGAGCGCGTCTCTGGGAC | 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 |
| | CGCCTTCCTGCAAACCTGGCGAGAGATCATGACGGGGGCCCAGCCCTGCA | | TTAAGA | SAEIAVNSAAGPDGF |
| | CTCGAGGCACCGCCTCTACCTACCCGGGGGGGAAGTCGACCTGTTCTTCC | | ACAGCC | SARLLKSVPALLLRVM |
| | CGGCGACGGCGCAAGAGATCCAGAGCGCTGAGATAGCCGTCAACTCGG | | CCACAA | VNLLLLVRRVPAALR |
| | CTGCTGGACCCGACGGGTTCTCAGCCCGTCTCCTCAAGTCCGTCCCGGC | | AGTGTT | DARTTFIPKVPDAVD |
| | CCTCCTCCTAAGGGTCATGGTTAACCTTCTGCTCCTCGTCCGACGTGTCC | | ACCTATA | PSQFRPITVASVLQRL |
| | CGGCGGCCCTCCGGGACGCGAGAACGACCTTCATCCCGAAGGTCCCCG | | TTAACA | LHRILAKRALEAIPLNF |
| | ATGCAGTGGACCCCTCCCAATTTCGCCCAATAACGGTGGCCTCCGTTCTC | | ATAAAG | RQRAFQPVDGCAEN |
| | CAGCGCCTACTACATCGCATCCTGGCCAAGAGGGCGCTGGAGGCCATTC | | TTGAAG | IWLLSTALNEARTRR |
| | CCCTCAACTTTCGACAAAGAGCCTTTCAGCCGGTGGATGGCTGTGCCGA | | CCTCAAC | RPLHMASVDLTKAFD |
| | GAATATATGGCTGCTGTCCACCGCGCTCAACGAGGCAAGAACCAGACG | | CACGCA | RVTTDAILRGARRAG |

| | | | | | |
|---|---|---|---|---|---|
| | | GCGCCCGCTACACATGGCGAGCGTCGACCTAACCAAGGCATTCGACCG | | TTGCGG | LSGEFIGYLKELYTTSR |
| | | GGTCACCACGGATGCCATCCTGAGGGGCGCAAGGCGCCGCCGGGCTGTC | | GTTAGA | TLLQFQGESLLVEPTT |
| | | CGGGGAGTTCATCGGATACCTGAAGGAGCTCTACACAACATCCAGGAC | | TGGCGT | GVRQGDPLSPILFNL |
| | | CCTCCTGCAGTTCCAGGGAGAGAGCCTGCTTGTCGAACCCACGACCGGC | | GGCTTG | VLDEYLSSLDPDISFVS |
| | | GTGCGACAGGGCGACCCACTGTCGCCCATCCTCTTCAACCTGGTCCTGG | | GCCCGC | GDLRLDAMAFADDLI |
| | | ACGAGTACCTCTCCTCCCTGGACCCGGACATCTCCTTCGTCTCGGGCGAC | | CGCCAT | VFASTPAGLQDRLDA |
| | | TTGCGCCTCGATGCGATGGCATTCGCTGACGACTTGATCGTCTTTGCCTC | | GATGAG | LVEFFDPRGLRVNVK |
| | | AACCCCAGCCGGCCTGCAGGATCGGCTCGATGCCCTCGTCGAGTTCTTC | | CTGGAA | KSFTLSLQPGRDKKV |
| | | GACCCAAGGGGGCTCAGGGTGAACGTGAAGAAGAGCTTCACGCTATCG | | CCCTCCA | KVVCDQIFTIGGTPLP |
| | | CTGCAGCCGGGACGAGACAAGAAGGTCAAGGTGGTGTGTGACCAGATC | | CCTGGT | ASKVATPWRYLGMT |
| | | TTCACCATCGGAGGAACCCCACTCCCAGCCTCCAAGGTCGCAACCCCTT | | GGGCCG | FTPQGSINKGTSEQL |
| | | GGCGCTACCTGGGGATGACCTTCACCCCCCAGGGCTCAATCAACAAGG | | CACGAG | DLLLTRTSKAPLKPQQ |
| | | GCACCAGCGAGCAGTTGGACCTACTGCTCACGAGAACCAGTAAGGCCC | | ACCACC | RLVVLRNYLLPRLYHR |
| | | CCCTCAAGCCACACAACAGAGGCTGGTGGTCTTAAGAAACTACCTGCTCCC | | GGCTCT | LVLGPWSAALLLKM |
| | | GAGGCTATACCATCGCCTCGTGCTTGGACCTTGGTCGGCCGCCCTCCTAC | | TTCTACT | DTTIRGAIRRWMDLP |
| | | TGAAGATGGACACCACCATTCGAGGAGCCATTAGACGCTGGATGGATC | | AAGGCC | HDTPLGFFHAPVTEG |
| | | TCCCGCACGACACACCGCTGGGTTTCTTCCACGCCCCAGTAACGGAGGG | | GGTCTC | GLGINSLRASIPAMVL |
| | | AGGCCTAGGAATCAACTCCCTGCGAGCATCAATTCCAGCCATGGTGCTC | | CGTGAC | QRLDGLHFSTHPGAE |
| | | CAACGGCTGGATGGACTTCACTTCAGCACGCATCCCGGAGCTGAGGTCG | | TGCGGT | VAIQLPFLTGLHRRAE |
| | | CCATCCAGCTGCCGTTCCTGACAGGACTCCATCGAAGAGCGGAAGCGG | | TGGGAT | AAAQYQGQRLLSKA |
| | | CGGCCCAATACCAGGGACAGAGACTACTGTCCAAAGCGGACGTCCACC | | AAACTC | DVHRMWSARLHGS |
| | | GGATGTGGAGCGCAAGACTCCACGGGAGCTGCGACGGAAGACCCCTTC | | CAAGCA | CDGRPLRESKRVPAA |
| | | GGGAGTCCAAGAGAGTGCCGGCTGCCCATCGTTGGGCCGCGGAAGGC | | CTGAGC | HRWAAEGTRLLSGR |
| | | ACCAGACTACTCTCGGGAAGGGACTTCATCTCGATCACGAAACTCAAGA | | GGTAAA | DFISITKLKINALPTLE |
| | | TAAACGCGCTACCTACACTCGAGCGCACCAGCCGGGGCCAGCACAAGG | | AAAAAA | RTSRGQHKDIQCRA |
| | | ACATCCAGTGCAGAGCTGGCTGCCAGGCTGTTGAATCCCTGGGCCACGT | | AAAAAA | GCQAVESLGHVLQA |
| | | CCTACAAGCTTGCCATCGAGGACACCGTGGCCGAATCCGGCGGCATGAT | | AAAAAA | CHRGHRGRIRRHDNI |
| | | AACATTGCCCGCTACGTCTGCGGCCGACTGACCCAGATTGGCTGGGCGG | | AAAAAA | ARYVCGRLTQIGWA |
| | | TGAAGTGGGAGCCCCCACTACTCTGTCGCTGGAAGGACCCTCAAACCTGA | | (SEQ ID | VKWEPHYSVAGRTL |
| | | CATCGTTGCCCATCGTGGAGCCGAGACTGTCGTGCTCGACGCCCAGGTC | | NO: | KPDIVAHRGAETVVL |
| | | GTCGGCACCAGCATGCGACTGGGCTTCCACCACGCTCAAAAGAAAGAA | | 1336) | DAQVVGTSMRLGFH |
| | | AAGTACTCTCTCCCAGACCTCCTCCACCAAGTCTGCGAGGGACGGAGAG | | | HAQKKEKYSLPDLLH |
| | | ACGCAGCCCGGGTGTCAACAATCACCCTCAATTTTCGAGGTGTTTGGGC | | | QVCEGRRDAARVSTI |
| | | ACCTGAGAGCGCCCAGGACCTGAAGTCCCTGGGCCTGACGGACAACGA | | | TLNFRGVWAPESAQ |
| | | CCTAAAGCTTCTCACCGTCCGCTGCCTCCAGGGCGGCGCGCAGTGTTTC | | | DLKSLGLTDNDLKLLT |
| | | CGGCTGCACCGCCGAATGACCACCGTGGTGAAGGCCACGGGCGATGAA | | | VRCLQGGAQCFRLH |
| | | GCCAACGCCCTCCCCGCCCATTCGGGCTTGCCGCCAACACAGCTTGGTG | | | RRMTTVVKATGDEA |
| | | GCCGAACCCTGGGTCCCTCTGCCCACAATCAGAGTGCAAGAACTACTTA | | | NALPAHSGLPPTQLG |

| | | | GTGTGACGGAGTCCTCAAGCCCCCACAAGTGCCTGCCAGGTGGCAGGA<br>AAGGGCAACTACTGGTGAGCGACCCAAGCAAGGCGGAGCCAAGACCA<br>AGCTGGAGCCAAGAGCAACTCCAGGAGGCAGGGGTGGATATCAAGAG<br>CAACCCCAAGGGACACAGACCACACGGGCAACTACTGGTGAGCGCCCAAG<br>ACAGGGGTGGATATTAAGAACAGCCCCCACAAAGTGTTACCTATATTAAC<br>AATAAAGTTGAAGCCTCAACCACGCATTGCGGGTTAGATGGCGTGGCTT<br>GGCCCGCCGCCATGATGAGCTGGAACCCTCCACCTGGTGGGCCGCACG<br>AGACCACCGGCTCTTTCTACTAAGGCCGGTCTCCGTGACTGCGGTTGGG<br>ATAAACTCCAAGCACTGAGCGGTAAAAAAAAAAAAAAAAAAAAAAAAA<br>AA (SEQ ID NO: 1090) | | | GRTLGPSAHNQSART<br>T (SEQ ID NO: 1458) |
|---|---|---|---|---|---|---|---|
| R2 | R2-1_MLe | Mnemiopsis leidyi | TGGGGGCCCCTTGGACTTGCTCCCTGGGGCAGGACACCAGTGAAAGGA<br>GATCCTCAAGACAGGACAGAGAGACAGGCACACCAACCCTTCGAACCT<br>GAGGAGACACCCAGATCTGGTTACCCCATTCCTGTAACCATGGTTGCTCT<br>CCGTGCGCTGTTAAGGAAACCCAGCCTAGTACCCTCGGGGAAAGGTTG<br>CGTAGTACTTAGCAGTGTGCGGATCAAACCTCTACCGGTCTCTCTAGCG<br>ATGAAAGTTTCTCCGACTGGAACTTGAGGGACTGGCTAGCCCAGCTAGC<br>CTGTAGAGCAATGCGTATACGATGCCTTGGCGACAATGGCGACCGCTGC<br>TTAGCAGACGGAGGTTAGTGAAAGGGCGACTTGCTGTTCATAGTCACGT<br>GAGTCGTCTAGAAACTGCACCGATGCTGCCCTCTGTTCCAGGAGAAGGA<br>CAGTGGAGGACTAAATCGTAGCGCGAGCGGTGTTATGTCGAACACTAG<br>CCACTCGAAACTTAACCTCAAAATGGATAATAAACTCAAGACTAGTCTCG<br>AGACTCCCTCCGGAGTCCGAGCAGATTCGATAATAACGCGGGTGAGGA<br>CCAGCAGTAACCGTGGTGAACACAGCAATGGTGTCACGTATCCCCGGTG<br>CGAACAGGGGGTTGCGCCTCTGGATACGCATGGGGGGATATGTGACGC<br>TCCTCCACAGGTGACGGTTCCCGCCACTGAGACGGATAAACAGAAAAAA<br>TGTGAATACTGCGAGTTCACATACCTGAAACCGAGACAGATAGGAACCC<br>ACATGAGAAAACGGCATCCCCAGGAATGGAACGACATCAAACGGACCA<br>AATTCCTGAGCGAAAAGAGACAAAAACGTTGGCTTGACGAAGACTTCG<br>AGCTTCTCTGCATAGGCCAAGAGGAATACCTTGTGCTATCCTCGATTGG<br>CAAACAGGGCAAAGGGATCAACCAGTATATTCAAACCAAGTACTTCCCA<br>ACCCTTAGTACAGATGCTATTAAATCTCAAAGGAAATCCAGACGGTTTTC<br>AGAGTACAGCGAAAAGAGATCACGTGAACTCCAACCTTGCAATACCAGC<br>AGCGACCCCGAAGAACTTCCGAATGAAGCGGTCACCGAAAACAGTCCG<br>CTCTCTTTTGATCCACTTGACCGAGATGTTGTAAAGAAAATAAGCTCCAA<br>AGATCACGGCGATCAGATCCTCTTAGTACAGGAGCACCTCATCAACGGA<br>CGGTATCAAGAAGCGAACACGTTGGCCAAAGCTATATTCGAGAAGCTCT<br>CCGGTAAATTCCCGAACCTGAAGACCGGCGACCATCGGCCTGGTAAACA<br>ACAAACAGCAAGGAAAGTCGGTAAGAAAAGAGTTAGAGGAAGCGGTA | TGGGGGC<br>CCCTTGGA<br>CTTGCTCC<br>CTGGGGCA<br>GGACACCA<br>GTGAAAG<br>GAGATCCT<br>CAAGACAG<br>GACAGAG<br>AGACAGG<br>CACACCAA<br>CCCTTCGA<br>ACCTGAGG<br>AGACACCC<br>AGATCTGG<br>TTACCCCA<br>TTCCTGTA<br>ACCATGGT<br>TGCTCTCC<br>GTGCGCTG<br>TTAAGGAA<br>ACCCAGCC<br>TAGTACCC<br>TCGGGGA<br>AAGGTTGC<br>GTAGTACT<br>TAGCAGTG<br>TGCGGATC | TAGTGA<br>CGAGAA<br>AAGTCG<br>CTTTATC<br>CTTACAT<br>AACAGT<br>GTGATA<br>GTCATCC<br>TTATCAC<br>AACTGT<br>CCTGGC<br>GAACAA<br>AATGGC<br>AGAGGG<br>ATAGTA<br>CTCGGT<br>CCAACC<br>AGAAGG<br>AAGCCC<br>ACACGA<br>TGCCAA<br>ACTTGCT<br>GTAACC<br>CACGTG<br>AGCGGA<br>AAACAT<br>CCTCCG<br>AGTATA | MSNTSHSKLNLKMD<br>NKLKTSLETPSGVRA<br>DSITRVRTSSNRGEH<br>SNGVTYPRCEQGVA<br>PLDTHGGICDAPPQV<br>TVPATETDKQKKCEY<br>CEFTYLKPRQIGTHM<br>RKRHPQEWNDIKRT<br>KFLSEKRQKRWLDED<br>FELLCIGQEEYLVLSSI<br>GKQGKGINQYIQTKY<br>FPTLSTDAIKSQRKSR<br>RFSEYSEKRSRELQPC<br>NTSSDPEELPNEAVT<br>ENSPLSFDPLDRDVV<br>KKISSKDHGDQILLVQ<br>EHLINGRYQEANTLA<br>KAIFEKLSGKFPNLKT<br>GDHRPGKQQTARKV<br>GKKRVRGSGKKLSPS<br>KQNRRELYAIVQKQ<br>WRTKKRSKVINQILT<br>GNLNKEQSYTHTPD<br>QLAQFWSTLFGRVSP<br>RDDRPINHRRSVIPEL<br>DKPLSVEEVEAALKG<br>AKDAATGIDGVPISH<br>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 |

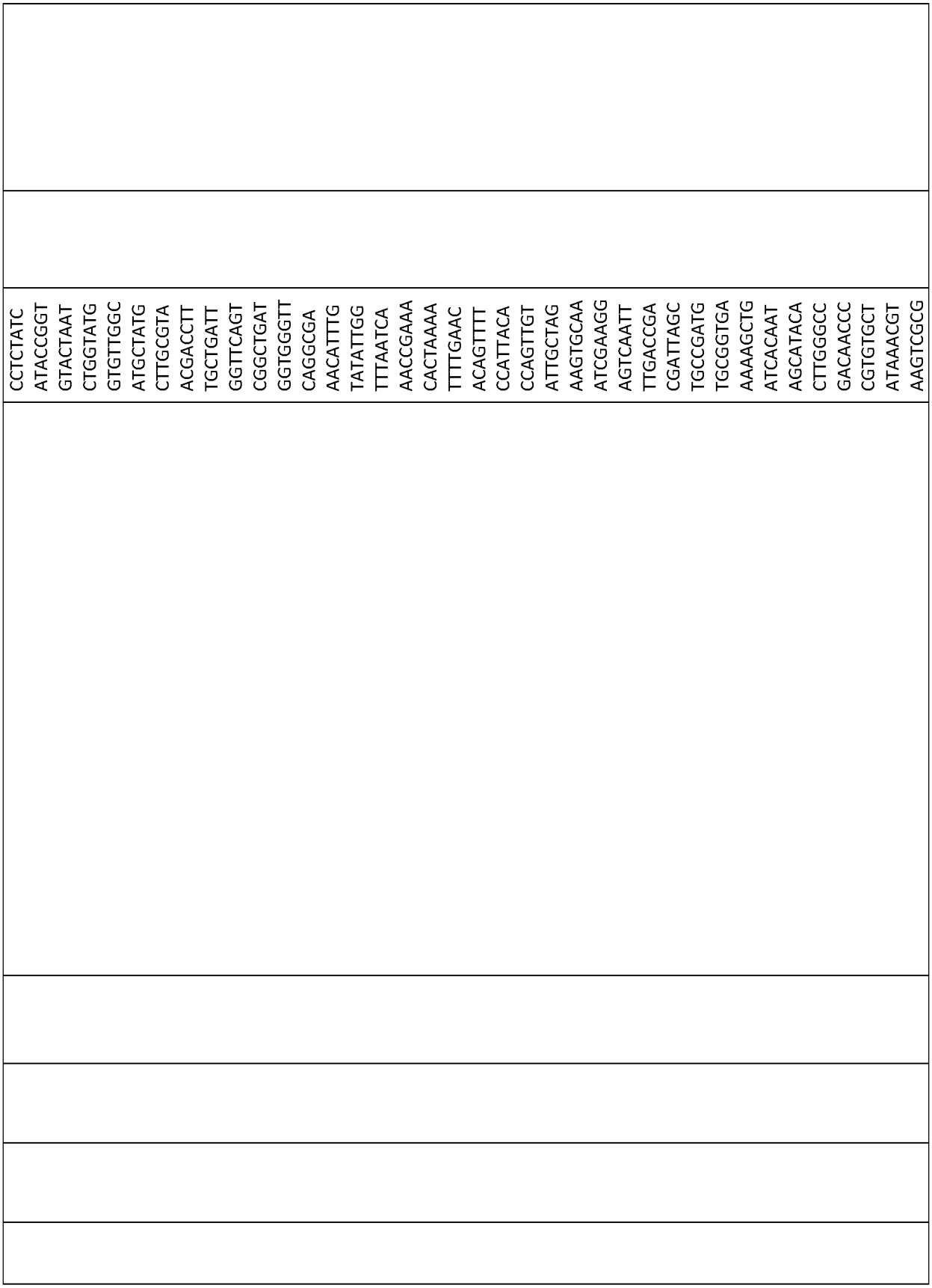
| | | | | | |
|---|---|---|---|---|---|
| | | CTGAGGCTAATATTTGTGACAAATGTCCAGGCAACAGGCAGGCCACTCT | | TTGACG | VADQGISRHEDEDQ |
| | | TGGCCACATCTCGCAAACTTGTCCGGCGACTCACGGGAGGAGAGTCAA | | GCGCTA | QKKIVKYDVDGYKRA |
| | | ACGACACGACAAGATTGTGAACAGAATCGCAAAAGCTCTAAAAGAACG | | CACCGG | AYKMLGIDYGSIPCN |
| | | 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 | |
| | | AGGATACAGGATGTGGACCGCTCTAGGCGCGGCGGACGGTGTAGACGGC | | GTAAGC | |
| | | GACCTTACCATAGCTGCGACTGCTGTGCGATCCGGAACGGGCCTTTCTC | | TGGTAG | |
| | | ACTTACCTCAAGCACCGTGTCGGATCGCGCGGGGTGGGCGCGGGATTCCT | | 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

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | CTATTAAT | TAATTTA | MNERLTDELTTEFILS |
| | | | CTAATTGGGATGAAGAAGGGGGACACGAGTTTGTGTGTGCATCCAG | GGGATGA | AGGTAA | DMFLWDYPCTDQN |
| | | | TTTCCATGGTGCATGCAGGAGTGGTGGTTTAAATGGCGAGACTCTACAG | AGAAGGG | AATCGT | KCYPCNLVFLDHRTW |
| | | | GGCTTCCATGGCTACACGGGATGCAAGGCATCAGACATTTTGGCACAG | GGACACG | GGGATT | SSHMARVHPHANKT |
| | | | GCAATCCTTTTGGTCTCTACCGCAATCATGTCTTAGACCTCAGTAGCGAC | AGTTTGTG | GTTTTGA | YKCRICNRTADSIHKI |
| | | | CACTACAACCACAGTGGTGACTGCTGTTGAGTGAAGGACGACTGAGCG | TGTGCATC | TGGCAA | ASHYGRTCKSLIGKTN |
| | | | CTGGATAACAACTTTCTTGCGTGGCCCAACATCGAAGCAACCACTTCGG | CAGTTTCC | TCTGCCT | AITTTIDETLFSCLHCS |
| | | | AGCTGGCACAAGGCAAGAGGGCAGCCCAAGGTGTGAATCATCTCAACT | ATGGTGCA | AGTCGC | RGFTTKTGLGVHTRR |
| | | | TCACTGCAGGAAGAAATGCTGTGCAAGGATGAGTGTGAACGACACCAA | TGCAGGA | GGCCTT | THPTEHEAILQQNTP |
| | | | CGGGATTGTTGCTGACCAGGAGGTGCCAACCAAATTTGAATGGATTGAC | GTGGTGGT | CCATTTT | GRKVRWGEEEVEIM |
| | | | TTTGGGCCTGGTTTCTCCTGCGTGTATTGCACGGAAAAACAAGTGGCTA | TTAAATGG | GGGTAG | AHKEAQQKDEDINM |
| | | | CACGTGTGGCCGTCGTGTCCTGGGGTTTCGCAACACAACTCCACAAGAT | CGAGACTC | GCAGCA | NQLIQNSVMPHRTLE |
| | | | CGACAACTATGAGGATGACAATGTACTTAAAGAACACAAAGAGACTGACG | TACAGGGC | GACCCA | AIKGKRRNIKYKELVR |
| | | | CCAAAGGGGATTTAAACCGCCAAATCGTACATTGGGTCTCACTACAATT | TTCCATGG | TCTATAT | TLKETTYKVENQCLV |
| | | | TTTTTACGTGTATTTATTTTCCTAAGTGTCTGTACTTGCCATTCTTCGCTGC | CTACACGG | AACAAA | NLVLPTTSEITTTPSE |
| | | | TTTTTCTGCATTAATTGCATATCGTATGCAAATAAGCGAATTAACCACCA | GATGCAAG | CTACTTT | GDQPAIRAEKEQSPT |
| | | | CCGTGCAACTATATGCAGATGTTACAGCTGAGCCCTCTATCATACCGGT | GCATCAGA | GCCTTTC | AAEDLQVIINDLKSQ |
| | | | GTACTAATCTGGTATGGTGTTGGCATGCTATGCTTGCGTAACGACCTTTG | CATTTTGG | ATAGGG | NFSHNQALLLLNSHV |
| | | | CTGATTGGTTCAGTCGGCTGATGGTGGGTTCAGGCGAAACATTTGTATA | CACAGGCA | GTACCC | EKFLNRSKPIKRKDHV |
| | | | TTGGTTTAATCAAACCGAAACACTAAAATTTTGAACACAGTTTTCCATTA | ATCCTTTT | GACCCT | NQQEIDENRHRRQS |
| | | | CACCAGTTGTATTGCTAGAAGTGCAAATCGAAGGAGTCAATTTTGACCG | GGTCTCTA | ACCAAC | KQTKYRRYQYLYHTN |
| | | | ACGATTAGCTGCCGATGTGCGGTGAAAAAGCTGATCACAATAGCATACA | | | |
| R2 | R2-1_PM | . | Petromyzon marinus | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | CTTGGGCCGACAACCCCGTGTGCTATAAACGTAAGTCGCGAATTATAAA | CCGCAATC | TTTCGG | KKALLDEITSDRSGPSI |
| | GAAAACAAACCGGACGGACTACTCGGTGACGAACTAACATCGCTCATG | ATGTCTTA | GGAAGT | YPTEESIRGTFVTLFES |
| | AATGAGCGATTAACAGACGAGCTGACTACGGAATTTATCCTTTCGGACA | GACCTCAG | AAAAGA | NSPPDNIPSKLKNDQ |
| | TGTTTTTATGGGACTACCCATGCACAGATCAGAACAAATGTTATCCATGC | TAGCGACC | AA (SEQ | SCIDIVKAITLDELIKTL |
| | AATCTTGTTTTTCCTAGACCACACAGAACTTGGTCATCACATATGGCACGGGT | 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 |
| | TCTTCAGGTCATAATTAACGATCTAAAGAGCCAGAATTTTAGCCACAATC | 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 |
| | CTAGATATCTTAAATATCATCCTTTACATACAGGATATACCACACAAATATG | 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<br>ATTCTATGATGTTTGCAGACGACCTTATCCTATTTGCAGAAACGGAATGT<br>GGCATGAATAAACTCTTAGATATAACTACTAAATTCCTCGATGACAGACA<br>CTTGAAAATAAATATAAACAAATGCAATTCGTTAAGATTTATCAAGTACG<br>GCAAACAGAAGACATTTAGTGTTGCAACGACATCATCGTACTTTATAAAT<br>AACGAACCCATTAATCCGGTATCATATGTAAAGGGATTCAAATATCTAG<br>GCATTGAATTTGACCCAAGGGGAAAACGATCTATAAGCTGTAACCTGCT<br>CGCAGCAATGTTAAACAAACTGACCAGAGCACCGTTAAAGCCAGAAAA<br>GAAAGTATATTTAATCAATAACAATTTAATACCTCGTATTATTCATCAATT<br>GGTCCTCGGAAAAGTTACCAAGGGTTTATTGATGTCACTTGATTCTGAA<br>ATTAGGAAAACAGTAAAGCTTCTGCTCAGGTTGCCACACGATACGCCCG<br>ACAGTTTCTTTTATACATCAGTATCCAACGGAGGAATGGGTATAAGAAA<br>TTTATGCGACTCAGTTGCACTATCTATAATAAACAGACACAACAAATTGA<br>TAACTTCAGATGATCTAGTAATAAGAGCATTATCACAACAATCATACACT<br>ATTGCAACGTTAAAACAGGCCCATATCATTGCAGGCTCCAAATTTCCTTC<br>GAAATCTTTAAATCAGAACAAATGGTCAAATAAACTATATCAAACACA<br>GATGGTCGGGGGTTGGTATACTGCCAATCTCAAACAGAAAACAATTCAT<br>GGATAACAGGGAATCATAGAACAATAAAATCGTATAATTACATAGACAT<br>GGTTAAACTAAGGATTAATGCACTACCGACTAAATCGAGATGCAATCGA<br>GGGACGTTAGAGACCAAGCAATGTAGATTTAAATGTCGAAGTATTAACA<br>ACCAAATTTCAGAGGAAACATTGGCACATATCTTGCAAAAGTGTGATCG<br>CAGTCATTATTCAAGAATCGCAAGGCATGATTCTTTGGTGCAATTTCTGG<br>CAACGGCCGCACAAAAACTAAACTGGGAAGTGATCAAAGAACCCACTTT<br>ACCGAGCGATACAAATAAGGCAAAACCGGACTTAATTTAGTAAGAGAC<br>TCTCATGTCTTGATAGTGGCAGTTCCGTGGGAGTCTCGATCATT<br>GGCACATGCATACGATTTTAAGGTGAAAAAATACGCTACTGACAAAAAA<br>ATGCAAGCATATTTAAAAACTATATATCCGGAAAAAGAAATTAGAAACGG<br>AGGCTTTAATCATATCTGCACGTGGGGGCTGGTGCGCTTTAAATAATAT<br>GGTAACAAAAAAGGTGGGATTGTCAAGTGCATGGGTAAAATTAGCATT<br>GATCAAGGTCATGGAGGGTTCCGTAAAGATATGGCGCTCTTGGAGCAA<br>AGGATAATTTAAGGTAAAATCGTGGGATTGTTTTGATGGCAATCTGCCT<br>AGTCGCGGCCTTCCATTTTGGGTAGGCAGCAGACCCATCTATATAACAA<br>ACTACTTTGCCTTTCATAGGGGTACCCGACCCTACCAACTTTCGGGGAAG<br>TAAAAGAAA (SEQ ID NO: 1092) | TTGCACGG<br>AAAAACAA<br>GTGGCTAC<br>ACGTGTGG<br>CCGTCGTG<br>TCCTGGGG<br>TTTCGCAA<br>CACAACTC<br>CACAAGAT<br>CGACAACT<br>ATGAGGAT<br>GACAATGT<br>ACTTAAAG<br>AACAAAGA<br>GACTGACG<br>CCAAAGG<br>GGATTTAA<br>ACCGCCAA<br>ATCGTACA<br>TTGGGTCT<br>CACTACAA<br>TTTTTTTAC<br>GTGTATTT<br>ATTTTCCT<br>AAGTGTCT<br>GTACTTGC<br>CATTCTTC<br>GCTGCTTT<br>TTCTGCAT<br>TAATTGCA<br>TATCGTAT<br>GCAAATAA<br>GCGAATTA<br>ACCACCAC<br>CGTGCAAC<br>TATATGCA<br>GATGTTAC<br>AGCTGAGC | | CRSINNQISEETLAHIL<br>QKCDRSHYSRIARHD<br>SLVQFLATAAQKLN<br>WEVIKEPTLPSDTNK<br>AKPDLILVRDSHVLIV<br>DVAVPWESRSLAHA<br>YDFKVKKYATDKKM<br>QAYLKTIYPEKEIRTEA<br>LIISARGGWCALNN<br>MVTKKVGLSSAWVK<br>LALIKVMEGSVKIWR<br>SWSKG (SEQ ID NO: 1460) |

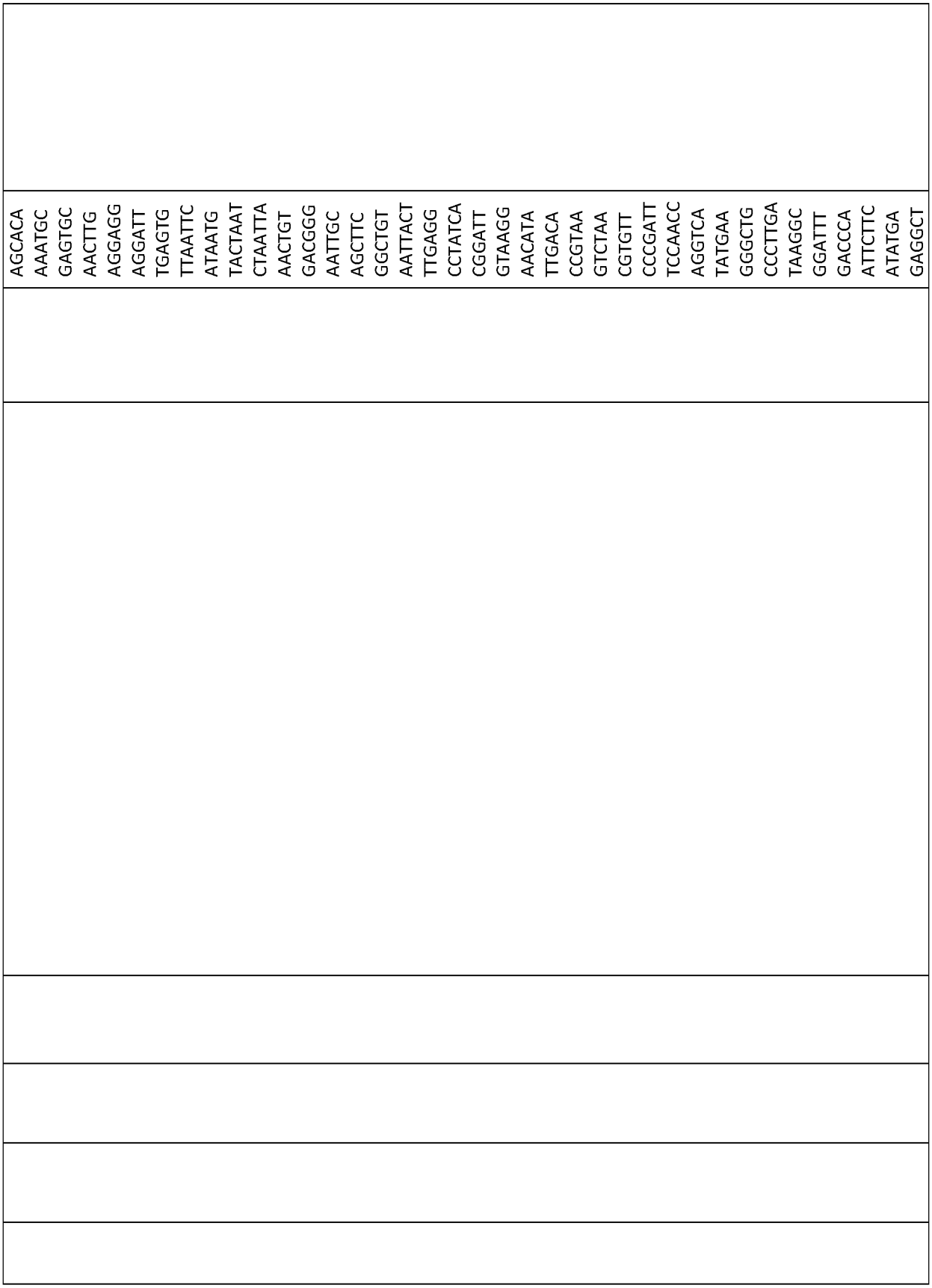

CCTCTATC
ATACCGGT
GTACTAAT
CTGGTATG
GTGTTGGC
ATGCTATG
CTTGCGTA
ACGACCTT
TGCTGATT
GGTTCAGT
CGGCTGAT
GGTGGGTT
CAGGCGA
AACATTTG
TATATTGG
TTTAATCA
AACCGAAA
CACTAAAA
TTTTGAAC
ACAGTTTT
CCATTACA
CCAGTTGT
ATTGCTAG
AAGTGCAA
ATCGAAGG
AGTCAATT
TTGACCGA
CGATTAGC
TGCCGATG
TGCGGTGA
AAAAGCTG
ATCACAAT
AGCATACA
CTTGGGCC
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 |
| | | | | TGCGGCGGTTCAGAGCAGAGTGAAAGCCGTGGTGAGTAAAAAAGTATTA | 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 |
| | | ATTAGACCGCAGACCCATACCACACAAGTGGTGGAAAACGAAGAGTTAAA | 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 |
| | | GCTTTTGATTCTGTAAACCACGAGAGTATTGTTAGAGCTGGAGAAATCA | 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) | |
| | CCAATCAAAGGAGTGCGTTGTAAGAACAAACTCGAGTTCTACAACAAGA | 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 | | |
| | TAAGACGTGTAAAAAAAGTGTTAGAGCCCGTTTGCAGAACCAAGACCGCT | CATTTTAC | GCAGGT | | |
| | GATTGTGATGGTTACGGACCGGAGGAAACCTCCGCCCGACCATGTCATG | TTCCGCCG | CATAAA | | |
| | AGCTTAACTTGAARGAGAGTTCGGGGACGTAAGGCCACGCGCGTCGTC | CGCTATAA | TGGAGG | | |
| | CTTGTTTGATCACTAGTGGATCAACCTTCGACTCCCCGGAACTGTGGGA | AACAGTTT | RGGGCA | | |
| | GTGGCGGAAGAAAGGCCAGAGGATGTCCTGAAACCATATATTTATTTAT | AAAAACTG | ACCACT | | |
| | AGAAGTTTTACTTCATCCTATTTACGTATTTCAGTATGAAAATGAGTAAA | AATAGGAA | GAAACT | | |
| | GTTCTCGACTCGATGAGTTGGGGGGCAACCATTGGGGGTCCTGAAGAGA | TCAAAAAG | TATGAG | | |
| | GGCTCTCACTGTAAAAAAATCTCTTCGTGTCTGTTTATTCCTAGGCACNTG | AACATGGC | CCAGAA | | |
| | CTGCATTATGAAGCGGWGAAAGTAAAGTTAGAGCTGAGAGATAGGTA | AAGCGACT | GAAGCT | | |
| | CTTGCTGCATTATGAAGCGGAGAAAGGCCTCGAATAAATAGGGTGTTA | ATATGTAA | TAACTAC | | |
| | GAGTTATTGATGGAGAGTATACTAGTAAGCTTAAGCTGCCGCCTCGCGCG | 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<br>ACTTTGAGGCCTATCACGGATTGTAAGGAACATATTGACACCGTAAGTC<br>TAACGTGTTCCCGATTTCCAACCAGGTCATATGAAGGGCTGCCCTTGATA<br>AGGCGGATTTGACCCAATTCTTCATATGAGAGGCTTATTCCAGCCTTCCC<br>GTAGTACCGTGAGGTTTTCCCGCCTCGAACGGAACAATGTTGCAGGGTA<br>ATTAAGTACATCGGGCTATATMGCGATATTTAACGTTTTA (SEQ ID NO: 1093) | TGCCATAT<br>CTTTKGAT<br>AAAGATAC<br>AGTTTAAA<br>ACTGCGAT<br>GATACTAA<br>TAGAGATC<br>CTCTTAGA<br>CCTTCGTA<br>AAGAAGT<br>GGGGATT<br>GATGACAT<br>TAGCATTG<br>GAAGAATT<br>AAATCTCC<br>AAGGAAAT<br>GGAGTAAC<br>TTCAATGA<br>AGTCCCAC<br>AACCCCGT<br>TGAAGGG<br>CTGGGTTC<br>GAGTATCG<br>AGAGAAA<br>ACTCTAAA<br>TTCTCTTC<br>GGTTMTG<br>TCCAACGG<br>AGGGGAC<br>ATTACTGT<br>AAAATATC<br>CTCTAAAA<br>ACAACT<br>(SEQ ID NO: 1216) | CAGGTT<br>TCAGTCT<br>ATATACT<br>AAAAGT<br>ATTTTTT<br>GATACC<br>GTGGTA<br>TATAGG<br>CAACTA<br>GTTAGG<br>AAATAG<br>TAAGGG<br>ATGACG<br>CATTGTC<br>TCTCTTT<br>ATGGTA<br>CTGAGG<br>AAACTT<br>ATCGAC<br>TCGCGA<br>GGGATG<br>AAACCC<br>GTAAAC<br>CGATCG<br>ATYTAGC<br>CTATAA<br>GTACCA<br>GCGACA<br>GTTAAA<br>CCATCTT<br>ACGCGA<br>GGGGTA<br>AAACCT<br>GAGGAC<br>CGATTAT<br>GGTATA<br>ACTTCTC<br>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<br>AGAAGGTTGATGAGCGTTCCTCTCCTGGACCGGGGGTGAGGATGTGTT<br>CTACTGAATCATCGTTCCGGTTGTGAGGTCCGCTGCAAATAGGCCTTGG<br>GGTGGTCTACCTCCGGGTCGCTACTCCTTTTGAGCTAGTCGATCCAGGT<br>GAGAGTCGGGGAAGCCCACTTAGGTGGGCCAGCTAAGCAGATCACCCC<br>CCCAGCACGACAGTGCGCTGTAATATAGCCTGTTGAGAGTGCACCCATT<br>TATACTGTTATAAAAGTATAAATGGTTCTTAGCTAACCTATCCAATGATT<br>TTTATTGTTATGTATAATACTGACCACTTTAGTCTAGTGTCATGTAAGGA<br>TCCGAGACAATAGCCTATGTCTCTAGTCTAGCAATAACACAAGAGGGAT<br>AATCCACCTCCCTACTCCAGCCACCACCTTTACCTTTCTTATCTTCCATATA<br>GATGCCAACCAGACGGGTCGATGGGCAGTGAGAAACAAGAATGGGGA<br>TAATTTGGAATTTGACATTCTTTCCTAAAGCAGACCTGAAGGGTTAAGA<br>GTCACAAGCGGAGTCCGGACTGTCTTTATAAACAGACAATATTTTCTTAT<br>TTTACCACCTATATGGGGATATTTCCCAACTCGTAATATAGGGCCCACTG<br>TCAAGTGGCTGGTATAAGTTACCCTGTGGGTAACAAATAAATTCAAACA<br>AAGAATGGAAAACTCCTTTGCGTGGGAGGGTACCAGTTCTGCTGAAGG<br>CAGAACAACAGTTGAGGATTCACCCTCAAGCAGTGACGATTTTGTTAGC<br>AATGTCGGATTCAAAGTAGCTAAGGCGGATCCCACTGTGTGGGAGGAAA<br>GCCAATATGTCAGAAGACAATACAATAATTGAGGACCCACCATCGAGTA | TCTCGCGA<br>CGCGTTCT<br>TCTGCCTG<br>ATGAAGTC<br>ACGTCAGG<br>TAATAGAC<br>TTAGAAGG<br>TTGATGAG<br>CGTTCCTC<br>TCCTGGAC<br>CGGGGGT<br>GAGGATGT<br>GTTCTACT<br>GAATCATC<br>GTTCCGGT<br>TGTGAGGT<br>CCGCTGCA<br>AATAGGCC<br>TTGGGGTG | TATTCCA<br>GCCTTCC<br>CGTAGT<br>ACCGTG<br>AGGTTTT<br>CCCGCCT<br>CGAACG<br>GAACAA<br>TGTTGC<br>AGGGTA<br>ATTAAG<br>TACATC<br>GGGCTA<br>TATMGC<br>GATATTT<br>AACGTTT<br>TA (SEQ ID NO: 1339) | TAAACCT<br>TGCCTCC<br>CCGGGC<br>CCCCCTC<br>AGTGAC<br>TAAGAC<br>AACTTTC<br>ACCGTA<br>ATAATC<br>ATATATT<br>TGTATAC<br>CATGTAT<br>TAATCTA<br>GGTAAC<br>AATTGA<br>AAGTAA<br>CATTGA<br>ACCGTA<br>TTCATAC | MENSFAWEGTSSAE<br>GRTTVEDSPSSSDDF<br>VSNVGFKVAKADPTV<br>WEEANMSEDNTIIED<br>PPSSSDDFANNVGFK<br>VTRADPTAWEEASTS<br>TETEDLPSSSNFIDNV<br>ETQIDMAGPTAWED<br>ADSNEDNIDEGTPN<br>NINNNLAIVRGRADA<br>YACSCCERNFISLKAI<br>GTHLKETHNKKVVFE<br>CAKCQHTFVKAHGL<br>ACHVPKCKEDADTP<br>MLNRLLHGCGECGL<br>AFNTRRGLSQHERHR<br>HPSACLSTRRRSRLD<br>GIARKKSLRNRRDIW<br>TNDEIRLLKQLMIQYE |

| | | | | | |
|---|---|---|---|---|---|
| | | GCGACGACTTTGCTAACAATGTCGGCTTCAAAGTAACTAGGGCAGATCC | GTCTACCT | TTTAATG | HAKKINIKIAEHFNHK |
| | | CACTGCGTGGGAGGAAGCCAGTACATCTACAGAGACTGAGGATTTACC | CCGGGTCG | AGTACA | NAKQVMHKRRSLRE |
| | | 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 |
| | | GTGCATGAACATCTCATTAGCGTCATAATGGACCTGTACGAAGGGGGGA | 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 |
| | | GGTACGGATGGGAAAAACTATTTGCCAGATCTTATTTTCAGTGCGGGAG | CAAATAAA | CCCATTT | IVGLKNSGISRTLQIVL |
| | | CTGGGGAGCCATGTTATGTAGTTGATCCAACAGTGGTGTGGGATGATG | TTCAAACA | CTGCCA | CDTIRMVKAFMAR |
| | | ATCCCAAGAACCTGAGGGAGGCCTGGAGGGGCAAGGTGAGGAAATAC | AAGA (SEQ | AGGACA | (SEQ ID NO: 1462) |
| | | ACACCAATCATTCCGGCGGTTGAAGCGATGCTTCACCCATCATCAGTTCA | | TGCGCA | |

| | | | | | |
|---|---|---|---|---|---|
| | | | | 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 |
| | | | | GTACTTGTCAACATAAAGAGGTGTCTCAATAGTTTGAATCAACAAGGGA | ACAGTCGA | GGTATG | TSRDGVSEVVLPPQP |
| | | | | GAGGAATACCGACCTGCTCCCTTGGGGCGGGGGTACTGGTCTTAGCCC | AGAGAGG | AGCGCT | VHAEGVSPASDSKDK |
| | | | | GGTTCCCCGCAAGTTTCCTTTGCCTGGGATGTGCCTGACTGGCTCCATCC | GCGCTACC | CCAACT | ATKITLLISLPVCDLRC |
| | | | | CCTTTCCCCATTAGGCACGGCTAGATGACGCACCGATGGGCGGGTGTGT | AATCCAAG | ATTGAA | GRCERPLETVGKAVR |
| | | | | AGGTCGCTACCGAAGGGGACTGGGGGTGTCCGGTGAACCAGGACTTCC | TATATGTC | CCCATAT | HFAVAHPTVSVVFKC |
| | | | | CAAAATGGTCTCACATTTTTAAGCGGCTTGAGTATCGCCCAGTATCCTCG | CCAAGAG | GATTCCC | QKCEKSSKNSHSISCH |
| | | | | CGCGGCACTGGGAACCCAGTCAACCGCTCTGTGCCCCGGCGCAGGCGG | 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 |

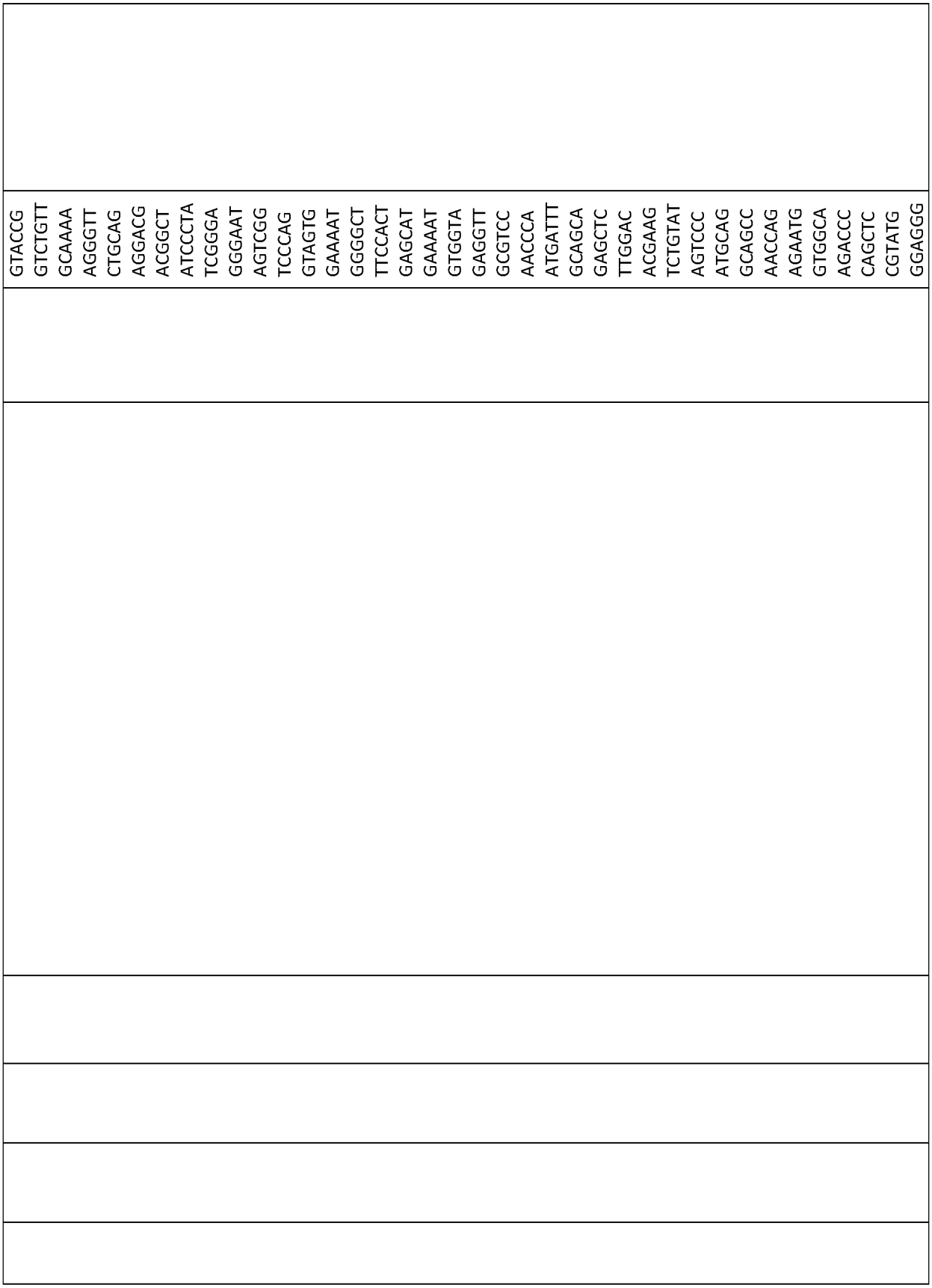
| | | | | | | |
|---|---|---|---|---|---|---|
| | | GGACAAGGCTACGAAGATCACCCTGCTGATATCACTCCCCGTCTGTGAT | ACTTGTCA | AGAACC | ACPINPRQRGFISSPG |
| | | TTGAGATGCGGACGATGTGAAAGACCATTAGAGACCGTTGGGAAGGCG | ACATAAAG | CCAACC | CSENLMILGGLIKRS |
| | | GTGAGACACTTTGCTGTGGCTCATCCGACGGTGTCGGTGGTTTTTAAGT | AGGTGTCT | CAGGTT | WAKGERLAVVLVDF |
| | | GTCAAAAGTGTGAGAAGAGCAGCAAGAATAGCCACTCCATCTCCTGTCA | 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 |
| | | CCCGCCACTCGAGATCGCTTTGGCGTTCAAAGTAGGTGGGAGGAAGT | AGGCACG | GGTGGG | PLCTCDGLLYSRCQD |
| | | GGAAACATTTCATGGGCTTGGCCAGTTCTACTCGAGGGGGGGAAGCGGA | 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 |
| | | ATACTCTCGCATCCTGACACACAGGCTGGAGCGGGCCGTGTCCCATTAAC | GGGTGTCC | CCGAAA | VGVQVFGGDKNSNH |
| | | CCGCGCCAGAGAGGATTCATTTCCTCACCTGGGTGTTCGGAAAACTTGA | GGTGAACC | AAGAAC | WMANPETLGSKERH |
| | | TGATCTTGGGAGGTCTAATCAAGAGAAGTTGGGCGAAGGGCGAGAGG | AGGACTTC | AGGAAG | YIAGLQLRANVYPTR |
| | | CTGGCGGTAGTGTTGGTTGACTTTGCGCGTGCGTTTGACTCTGTGAGCC | CCAAAATG | ATGGTG | EALSRGRPDLPKVCR |
| | | ACTCGCACATCTTGGAGATACTCAGACAGAGAGGGCTTGATGAACATAT | GTCTCACA | GAGTAA | QCLAGTESCAHILGQ |
| | | CATCGGAATCGTAGGTGACTCGTACACCGATGTAACGACCACAATTACA | TTTTTAAG | GTTGAG | CPAVKDSRIRRHHKL |
| | | GTCAGTGGGGAGCAGTCCCCTCCCATTGACATGAGGGTAGGGGTTAAG | CGGCTTGA | AGCGGT | CDLLASEAESAGWTV |
| | | CAGGGAGACCCGATGTCTCCGCTACTGTTTAATCTAGCCCTGGACCCAA | GTATCGCC | TGCTCG | IREMCCRTRAGALRR |
| | | TGATCGACACCCTTGAACGCTACGGCTTGGGGTACAGGATGGGCGAGC | CAGTATCC | GGAAGT | PDLVFVKTGFALVVD |
| | | AGCAGATCACGGCCCTAGCTTTTGCTGACGATCTGGTTCTGGTGAGCGA | TCGCGCGG | TATGTTG | VTVRYEMAYDTLMG |

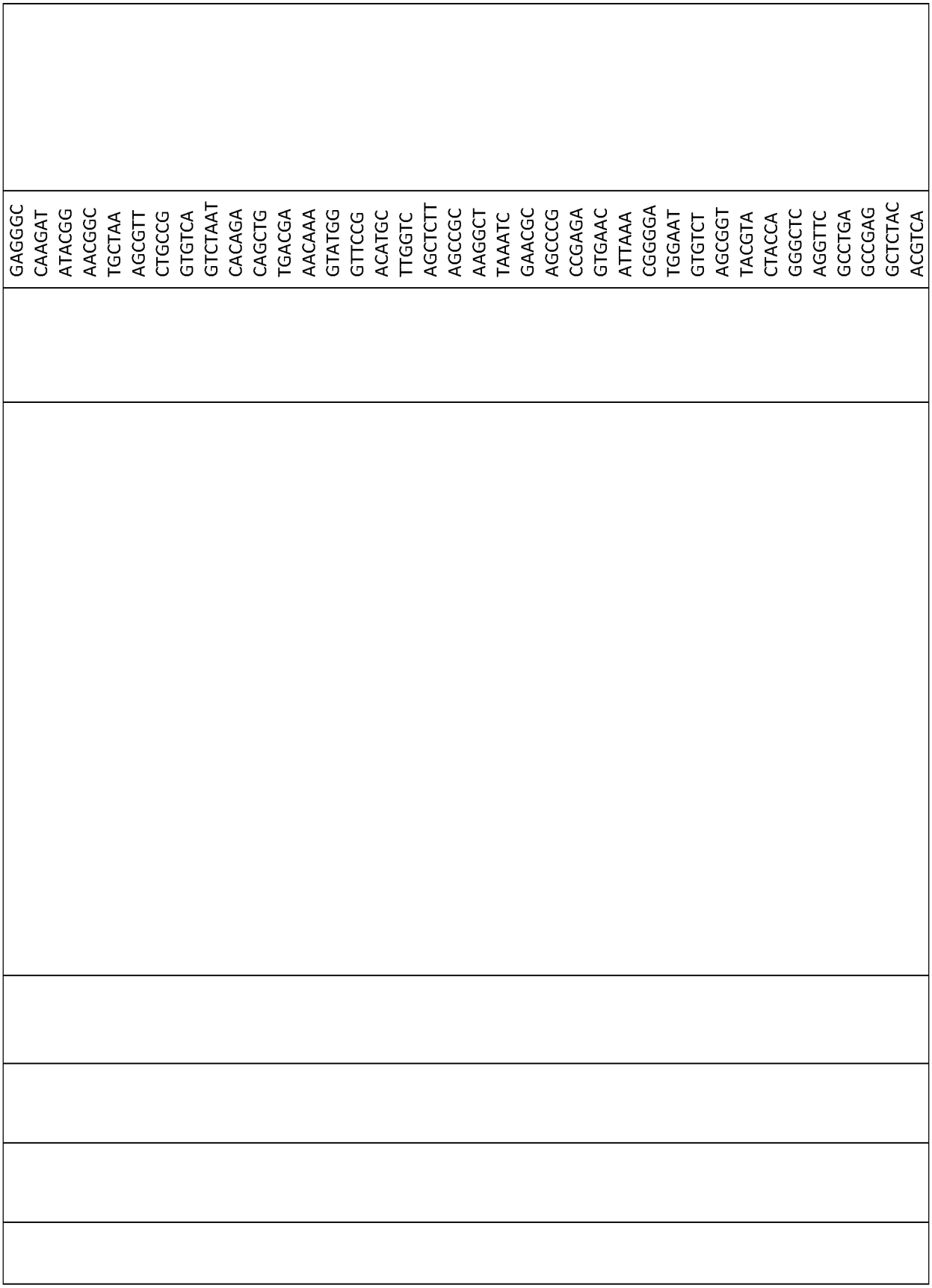
| | | | | | |
|---|---|---|---|---|---|
| | CTCGTGGGAGGGCATGGCGTGCAATATCCGTATTCTGGAAGAATTTTGT | CACTGGGA | TGATAA | AAAEKVARYTPITPYV |
| | CGACTGACTGGGCTGAGGATTCAGCCTAGGAAGTGTCATGGTTTCCTCA | ACCCAGTC | CTCCATT | AMTLKARRVKVFGFP |
| | TACAGAAGATTCAGAGGGCGAGATCGGTAAACCTCTGTAAGCCCTGGA | AACCGCTC | AAGGCC | 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 | |
| | CGCGAGGCACTGTCCAGGGGCGAGGCCGGACTTACCTAAAGTCTGCCGG | TTCGGGTG | ATAAGG | |
| | CAATGCCTAGCAGGAACTGAGTCTTGCGCGCATATTCTCGGGCAGTGTC | AAGGGCA | GTCTCCC | |
| | CTGCAGTGAAGGATAGCCGCATCAGGCGGCATCATAAACTGTGTGACCT | GGTGGGA | CTTATAC | |
| | GCTAGCAAGTGAAGCCGAAAGCGCCGGATGGACCGTCATCAGAGAAAT | AGAGAGG | TTCACTG | |
| | GTGTTGCAGAACTCGTGCCGGAGCTTTGCGGCGTCCAGACCTGGTGTTT | GCTGCCTC | GGAAAC | |
| | GTGAAAACCGGTTTTGCTTTGGTGGTGGATGTTACTGTGCGGTACGAGA | ACGATAAA | CCCATA | |
| | TGGCCTATGATACGCTCATGGGTGCGGCTGCCGAGAAAGTTGCTCGGTA | CACCTAGT | AGGGTA | |
| | CACCCCAATTACTCCATATGTTGCGATGACCCTGAAGGCAAGGAGAGTC | CAATAGCC | TCCCCCT | |
| | AAGGTGTTTGGCTTTCCACTGGGAGCCCGAGGCAAATGGCCGGGAAGC | AGTCGGG | ATATTTA | |
| | AACGACCGGCTGCTGAAGGCTATGGGTGTTGGTGGCGGCAGGAGGAA | AAAAATGT | CTGGGA | |
| | ACAGCTGGCCAAGTTGTTTAGCAGGAGGGCGCTCCTGTACTCCTTAGAT | GGAATGTT | GACCCC | |
| | GTCCTTAGGGACTTCTACCGGGCGGAGGGAGAAACGGGGGGACTTGGAT | AGGACAG | ATAAGG | |
| | GATGAGAGCGTGGATGATCATCTATAGATCCGTTTGTTATGATTGGAGG | GGAGGTA | GTCTCCC | |
| | GAGCCTGCCGAGTGGTATGAGCGCTCCAACTATTGAACCCATATGATTC | AGGGAGG | CCTATA | |
| | CCGAGGCCTGGCCAGACGCCTAGATGCCTGCCACAATTGAACGCAGCCC | CGGCTTTT | GATGTA | |
| | TAGCTTGCTAGGAGATCCATAGGAACTGGCCTATGGGGCGTCATGACG | TCGTAAGG | GAGCGT | |
| | GTTGAAGTTCCTCCATAGCGTGCTTGGGAGGGGACGACAATGACGAGT | CTCCTTCA | AAGGGG | |
| | CATGACGTACCGAGAGAACCCCAACCCAGGTTGGGGGAGAGAGGGTTGAA | ACCCCTAC | TCTCCAA | |
| | AAGAGCGGAGATGCTTGGTATACCAAGCTAGCAGAGAGAGGGTTGAA | CTGTAGTC | AGTACC | |

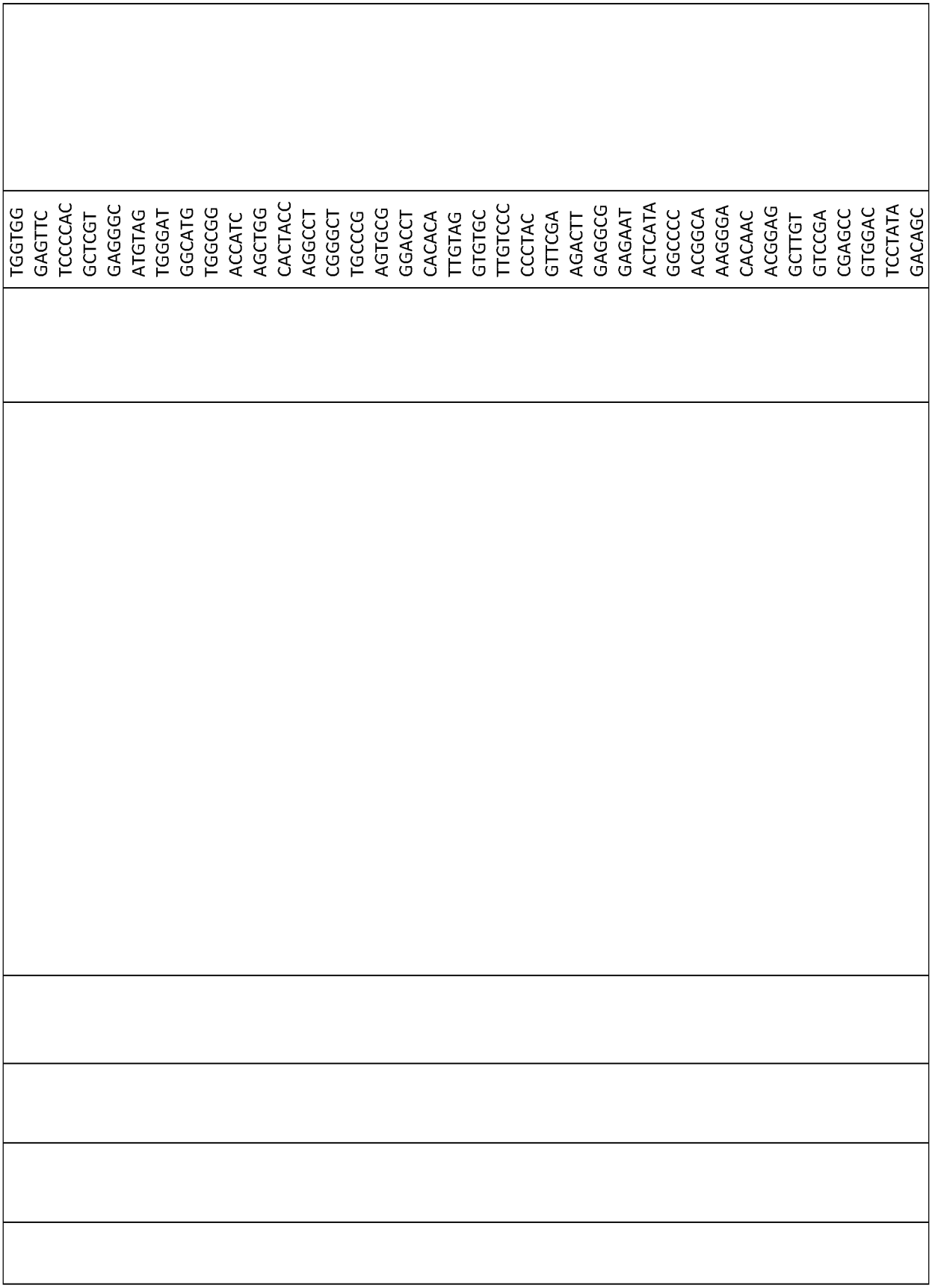
| | | | | | | |
|---|---|---|---|---|---|---|
| | | GAGGATGACTACTGGGCTCAGAGTCATCTCACCCTAAAAGGCGGTGGG | CACCTATT | GGCCGA | | |
| | | GCATCGGTTGAACACCTACCCATACCGGGATGGGAGGTGGTAGGCCGA | GCAGGTGT | TATGGC | | |
| | | AAAAGAACAGGAAGATGGTGGAGTAAGTTGAGAGCGGTTGCTCGGGA | TGACAACA | CTTATG | | |
| | | AGTTATGTTGTGATAACTCCATTAAGGCCGGTGGGCATGGTGCGGATAA | TGCAAGAT | GCAAAC | | |
| | | TGGAAACTATAAAAACAATAAAAAGAAAGACCAAAAAAAATGTTCTGTTA | GACCTGCC | TCTGGT | | |
| | | TGATGCCTTACACATGTCTGGGAGACCCCATAAGGGTCTCCCCTTATACT | TCGTTACG | GGTAGG | | |
| | | TCACTGGGAAACCCCATAAGGGTATCCCCCTATATTTACTGGGAGACCC | GGTCGCGT | GACAAG | | |
| | | CATAAGGGTCTCCCCCTATAGATGTAGAGCGTAAGGGGTCTCCAAAGTA | ATCATTGC | GAGGTA | | |
| | | CCGGCCGATATGGCCTTATGGCAAACTCTGGTGGTAGGGACAAGGAGG | TACAGGTC | AGGGCA | | |
| | | TAAGGGCAGTGCCAACCCCTACTTGATCGGGACCATCCAGGGAATGCCA | GTGTGCCG | GTGCCA | | |
| | | TCCTCCCGCGAAGGTGATGTGGTGAGGTAAGGGGGGAGCCCGTCTTCG | CTTCTAAG | ACCCCTA | | |
| | | AGTTTCCCCAACCCCTACCCACAGGTGAGAGGAGGAAGAGGAATCT | AGGATAGT | CTTGATC | | |
| | | GTCCCCAACGGGAGAGGGTGAGGTGTAAGGGGGAGACCTTCTAGTA | AAGGAGA | GGGACC | | |
| | | GGGTCTTCTCAGTCGCCTGACGTCCTGACTGTGGGGTGGATCAGTACCC | GGTTATAG | ATCCAG | | |
| | | TACAGGTGAGACCGGTGAGGTAAGGGTGTGGCCCTCTTGAGGGCTGCG | GGAGGTCC | GGAATG | | |
| | | CCAACCCCTACTCGAGGTAACCTGAGGGAGTGGTGGAATGGCGGCATG | TGTTAGGG | CCATCCT | | |
| | | TTAGTGCTGGGACTTGATTGCGAGGGTTTAATGAGAGTGGCCTGCTGA | CTTCCTCA | CCCGCG | | |
| | | GAGCAACACTTGTGGTGCTTAAAGCGGGGCGGCCCATGACCACCGTGA | ACCCCTCT | AAGGTG | | |
| | | GATAGGACACTGCACAGTGCAGCCATGAGGTTCCTGGAGGATGATGCG | CTATGCGA | ATGTGG | | |
| | | ATGAGGTGGGGGCCTCATCAGCCCCTCCTGGCAGGGCGTCGGCCAGGG | TTCCTTAC | TGAGGT | | |
| | | AAACTAAATGTCTCTAGCATGTCAGTGCAGTGAGGTAAGGGGGAGAGCA | AGGAGTG | AAGGGG | | |
| | | CTCTAGTAGGGCTCTTCCAACCCCTACCTGTAGGTCACCTGGTCCAGGTG | GATCGAGA | GGAGCC | | |
| | | TCGATGATGTGAAAACAAGAGCTACTTTGGTACCGGTCTGTTGCAAAAA | AGTCCCGG | CGTCTTC | | |
| | | GGGTTCTGCAGAGGACGACGGCTATCCCTATCGGGAGGGAATAGTCGG | ACGTAATA | GAGTTT | | |
| | | TCCCAGGTAGTGGAAAATGGGGCTTTCCACTGAGCATGAAAATGTGGT | CACCCTGG | CCCCAA | | |
| | | AGAGGTTGCGTCCAACCCAATGATTTGCAGCAGAGCTCTTGGACACGAA | AGGTAAG | CCCCTAC | | |
| | | GTCTGTATAGTCCCATGCAGGCAGCCAACCAGAGAATGGTGGCAAGAC | GGAGTGG | CCACAG | | |
| | | CCCAGCTCCGTATGGGAGGGGAGGGCCAAGATATACGGAACGGCTGCT | CCTTCTAG | GTGAGA | | |
| | | AAAGCGTTCTGCCGGTGTCAGTCTAATCACAGACAGCTGTGACGAAACA | TAGGGCTG | GGAGGA | | |
| | | AAGTATGGGTTCCGACATGCTTGGTCAGCTCTTAGCCGCAAGGCTTAAA | CTTCAACC | GAAGAG | | |
| | | TCGAACGCAGCCCGCCGAGAGTGAACATTAAACGGGGATGGAATGTGT | CCTCTGAT | GAATCT | | |
| | | CTAGCGGTTACGTACTACCAGGGCTCAGGTTCGCCTGAGCCGAGGCTCT | GGGAGTG | GTCCCC | | |
| | | ACACGTCATGGTGGGAGTTCTCCCCACGCTCGTGAGGGCATGTAGTGG | TACCGGGA | AACGGG | | |
| | | GATGGCATGTGGCGGACCATCAGCTGGCACTACCAGGCCTCGGGCTTG | ACCTCGAC | AGGAGG | | |
| | | CCCGAGTGCGGGACCTCACACATTGTAGGTGTGCTTGTCCCCCCTACGTT | TTGTAAGC | GTGAGG | | |
| | | CGAAGACTTGAGGCGGAGAATACTCATAGGCCCCACGGCAAAGGGACA | ACAGGTTA | TGTAAG | | |
| | | CAACACGGAGGCTTGTGTCCGACGAGCCGTGGACTCCTATAGACAGCCC | GTATGGGA | GGGGAG | | |
| | | GGGATATCACTGGGCACGCTCATACTGAAGAAATTCGATGAACCGGGGC | GCAGAAG | ACCTTCT | | |

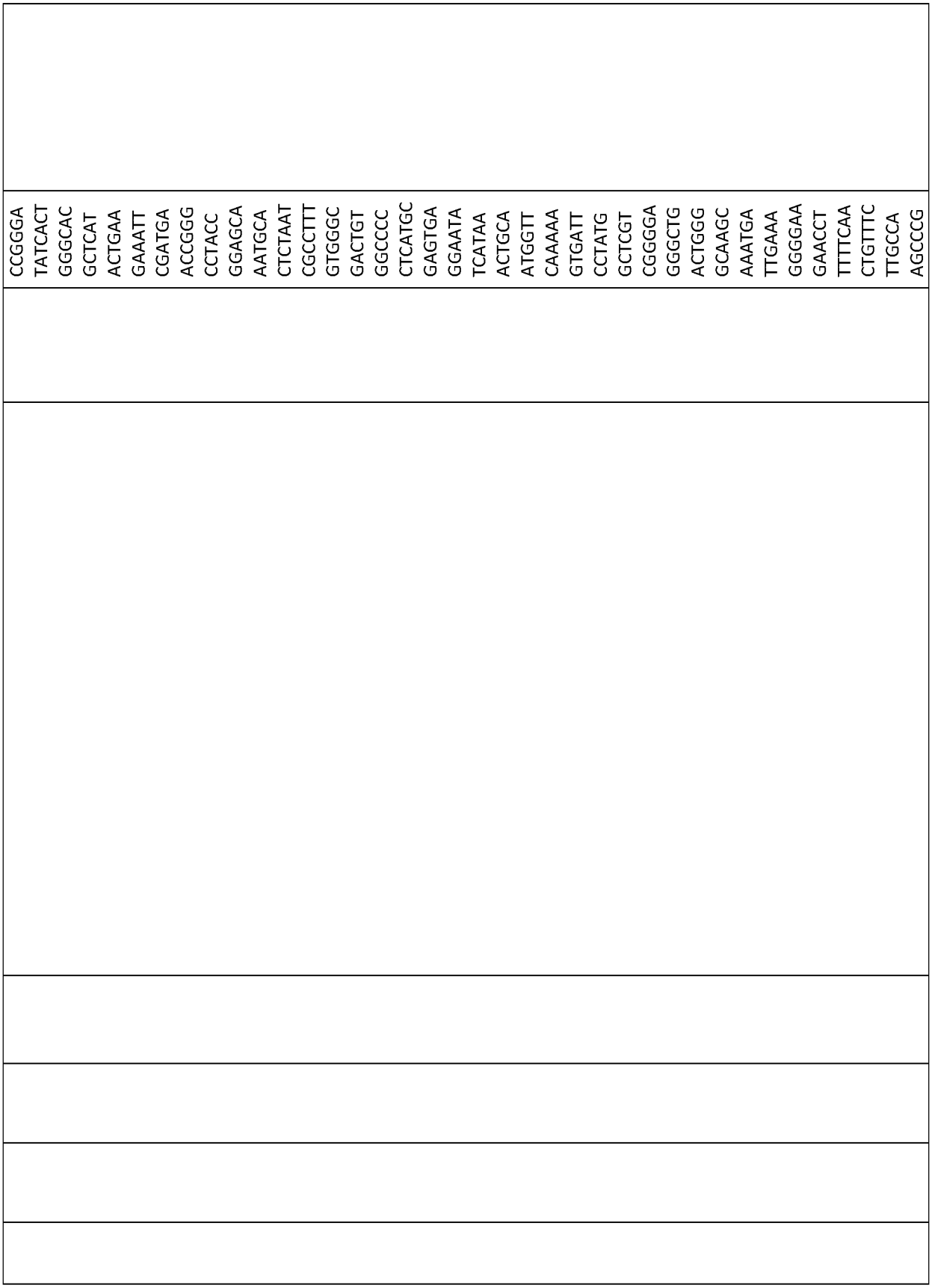
| | | | | | | |
|---|---|---|---|---|---|---|
| | | CTACCGGAGCAAATGCACTCTAATCGCCTTTGTGGGCGACTGTGGCCCC | GGGGAGC | AGTAGG | | |
| | | CTCATGCGAGTGAGGAATATCATAAACTGCAATGGTTCAAAAAGTGATT | CGTAATGG | GTCTTCT | | |
| | | CCTATGGCTCGTCGGGAGGGCTGACTGGGGCAAGCAAATGATTGAAA | GCTTCTCT | CAGTCG | | |
| | | GGGGAAGAACCTTTTTCAACTGTTTCTTGCCAAGCCCGGTTGATGGTGG | TTCACCTG | CCTGAC | | |
| | | CGCTAGTAATTGCGACGGGGAAAATGCGGTTTAAGTCTCCGAAGTAGTGC | CTTACATA | GTCCTG | | |
| | | GTAGCACCGGATGTCGACCGGGTGTAAAAGCCCTTCGTAAAGTCCCTGG | ATACCTGT | ACTGTG | | |
| | | GGAGGTCAGTCCTGGGGCTACTGATGCGCAGTATGTAATTCGCAGAAT | GGTGCATG | GGGTGG | | |
| | | AGGGCCATCGATACCGCCTGCGTGACTCGACTGGGTTTCCACTTGAGGA | TATCTAGG | ATCAGT | | |
| | | TATCCGACCGTAGCGTGCACCCTCTTGTAGTTGCGCCCGGAAACGGCTGT | TCTTGGCG | ACCCTAC | | |
| | | GTTCCCTCACGTATGTGAGGAAACTCAACAATGTGAGTGGGTAAACGGC | 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 | | |

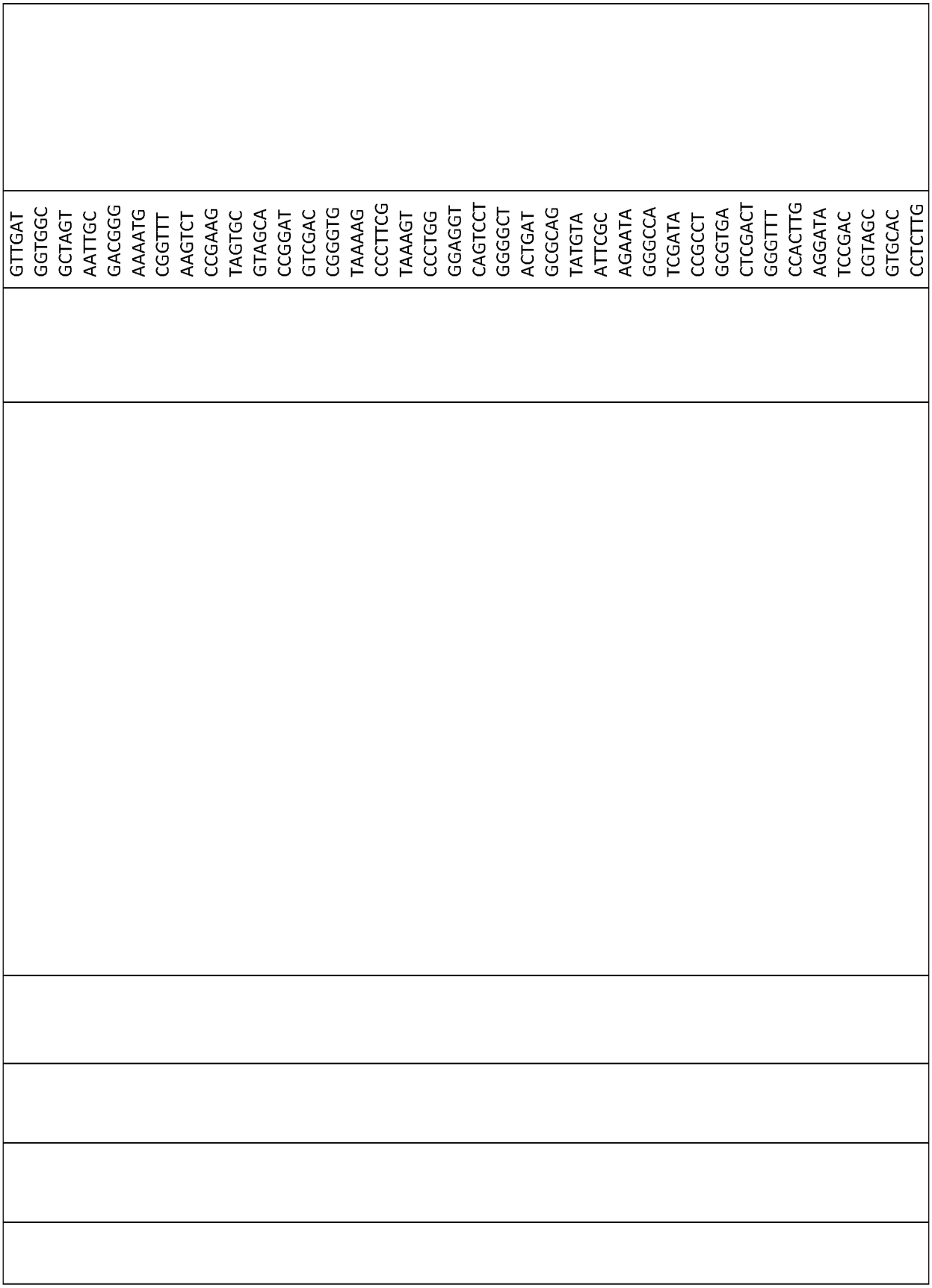
| | | |


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 | . | Triboliu m castane um | AGTCATAGAGCCAGAACCTCCTCGTGGTCCCGCTGGGCACAGGGATTAA | AGTCATAG | TGAAAA | MSRRPGKSNEPPVRS |
| | | | | TTTTTCTGTGGCAAATTTGACTGGCTTCAGAGAGCGTTTTTCGAAGTGG | AGCCAGAA | GAGTGG | RAMGLTTLSGTKTSN |
| | | | | WCTGTGTGACTGCGTTCCCCCCTTAGTTGCTATWTCCGCTKMGATTAAC | CCTCCTCG | CAGTTG | SGAQGPSTSAPMQN |
| | | | | ATCTCACCTCGACGTWTAAGATCATTATGTCGAGGCGCCCCGGGGAAGT | TGGTCCCG | TGGAGA | MAGGFVCDCGRSYA |
| | | | | CTAATGAGCCCCCGGTTCGTAGTAGGGCTATGGGTTTAACCACTTTATCT | CTGGGCAC | CTTCCTT | LKTSLARHKKECGKN |
| | | | | GGCACCAAGACATCAAATTCTGGTGCCCAGGGGCCCTCTACCAGTGCCC | AGGGATTA | CGGACG | NAECRWCGTRFNTL |
| | | | | CCATGCAGAACATGGCTGGGGGTTTTGTTTGTGACTGCGGCCGGAGCTA | ATTTTTCT | TCGGGT | AGTRQHERKAHFVQ |
| | | | | TGCCTTGAAGACATCTCTGGCCCGACACAAGAAGGAGTGCGGAAAGAA | GTGGCAAA | CGGATT | YQSDLAKALPQPESE |
| | | | | CAATGCGGAGTGTCGATGGTGCGGGACAAGGTTTAACACCTTGGCTGG | TTTGACTG | TTCGGA | LMEKIAIVEARSXNGI |
| | | | | CACTCGTCAGCATGAGCGAAAGGCTCACTTTGTCCAGTATCAGTCTGAT | GCTTCAGA | CGCCAG | FYKEMMASTGLTHQ |
| | | | | CTGGCAAAGGCGCTCCCCCAACCTGAGTCGGAGCTTATGGAGAAAATC | GAGCGTTT | GGTACC | QVRSRREKPEYKGFL |
| | | | | GCCATTGTGGAGGCCAGGTCCAKTAACGGCATCTTCTATAAGGAGATGA | TTCGAAGT | TCCACC | ERARRSLAQTNIRAG |
| | | | | TGGCCTCCACGGGTCTGACCCACCAACAGGTCAGATCAAGGAGAGAGA | GGWCTGT | GCTGGG | SISPASTXAGSLESAS |
| | | | | AACCCGAATATAAGGGCTTCCTGGAGCGAGCACGCAGGAGCCTGGCCC | GTGACTGC | TTCACAA | PKAGCSSSASPGPTT |
| | | | | AGACCAATATCAGGGCTGGCTCCATCTCACCAGCGAGTACGAWCGCAG | GTTCCCCC | ACTAGG | RSRAPTKGVPXRSSN |
| | | | | GCTCTTTGGAGTCTGCGTCTCCGAAAGCGGGGTGTAGTAGCAGTGCCTC | CTTAGTTG | CGAACA | SARIVVEAQVHTRAP |
| | | | | GCCCGGACCGACCACCAGGTCCAGGGCACCTACCAAGGGTGTGCCCCW | CTATWTCC | TCTGCC | PNTGETEVALRESRR |
| | | | | ACGAAGCTCTAATTCGGCGAGAATTGTGGTGGAGGCCCAAGTCCATACC | GCTKMGA | GATACC | TVPRLGXNPSRPCGIS |
| | | | | CGGGCGCCACCCAATACCGGGGAAACGGAAGTGGCTCTTCGTGAGTCA | TTAACATC | CTCTTTA | PLMAIAIDEDSVLGG |
| | | | | CGCCGAACCGTGCCCCGGCTCGGGCMTAATCCATCTAGGCCCTGTGGT | TCACCTCG | GGTCAT | LRVQAGPSPTAVHSV |

| | | | | | |
|---|---|---|---|---|---|
| | | ATCTCTCCTTTGATGGCGATTGCCATAGATGAAGACTCTGTGTTGGGCG | ACGTWTA AGATCATT (SEQ ID NO: 1219) | AGGACC ACATGT CTCTGCA CGAGAT TAACCC A (SEQ ID NO: 1342) | EAFPGTSSMTPMET |
| | | GGCTACGAGTGCAGGCCGGGCCAAGTCCCACCGCGGTGCACTCCGTCG | | | DRVHNKSGIDPILEH |
| | | AAGCTTTCCCCGGAACGTCCTCGATGACCCCGATGGAGACCGATCGGGT | | | NGTRQVRREESSTRE |
| | | ACATAATAAGTCCGGTATAGACCCCCATTCTCGAACATAATGGGACACGG | | | 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 |
| | | AAGACCCCCCCCCACTTACATCACCATGGACGAAATAAAGGCGGCAAGA | | | 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 |
| | | ATGTCGACGCCCTAGATTGCATTTAAGTATCTGGGGCTAGAATTTGGCCA | | | 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 |
| | | | ATAAACAACGACCGCTCTGCGGCGTTGCAGATTCCTCCGGTCGTGCGAC | | | | IVIADVQVSWDSESL |
| | | | GAGCTTGCGTAAATTCGGTTGTCAAGTGGGGGTCGTCGATCCATGCTAC | | | | 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 |
| | | | | CCGTGGTTCCCGCCCGGACCCCGTAACATCGGGTGACTGAATCTGTCTCT | 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 |
| | | TACGAGGGTGGTGAAAGTAACGGCGTTGTCAGAACATCTTAGGAGGGC | 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 |
| | | GGCGAAGCAGATCAGCGATAAGAGAAGGCAGCTGGCGGCAGCGACCA | 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 |

| | | | | | |
|---|---|---|---|---|---|
| | | | | TCTCGTC | DPLSPILFNLALDPLLC |
| | | | | CCAAAT | RLEDRGLGYKYGDQ |
| | | | | GCCGCG | QIXSLAFADDLALLSD |
| | | | | TCCAATT | SWEGMQQSIRVVEE |
| | | | | ATACCG | FCQRTGLRVQAPKC |
| | | | | GATTTG | HGFLIRPTKESYTIND |
| | | | | ACAAAG | CDPWTIADMQLDMI |
| | | | | CGGACG | DPGSSEKYLGLGIDP |
| | | | | GCCCGC | WIGLSRPELSEVLTR |
| | | | | TTTATAA | WVKNIGGAPLKPLQ |
| | | | | GCCGGA | KVDILRSYALPRLLFIA |
| | | | | AAAGGT | DHAGLSATCLHSLDL |
| | | | | GCCTTG | SIRSAVKGWLHLPPS |
| | | | | TAAAATT | TCDAIIYVSYKDGGLG |
| | | | | GCAAGG | LPRLASLIPNVQARRL |
| | | | | TTCATTA | VRIAQSEDDVIRSVVL |
| | | | | AATAG | QEGIQEEIRKVWISA |
| | | | | (SEQ ID | GGRPEKVPSVTGEFP |
| | | | | NO: | VMEAQAADEALSEW |
| | | | | 1343) | ERRAPRTIYPIPCKWR |
| | | | | | KREMENWTNLKSQG |
| | | | | | HGIRNFENDRISND |
| | | | | | WLLHYGRIPHRKLITA |
| | | | | | IQLRANVYPTREFLAR |
| | | | | | GLGEGAPRGCRHCP |
| | | | | | AEWESCSHIIGYCPA |
| | | | | | VQEARIKRHNDICGV |
| | | | | | LAEEARKLGWVIFIEP |
| | | | | | HLRDNTNELFKPDLV |
| | | | | | LVKGSCAKVVDVTIR |
| | | | | | YESGLTTLSDAAAEKA |
| | | | | | RKYQHLAGEVRALTS |
| | | | | | ATTVDFLGFPIGARG |
| | | | | | KWYVGNNGLLSDLG |
| | | | | | FSTSRVVRIARALSKK |
| | | | | | ALLSSVDIIHIFASRAR |
| | | | | | QAQTSE (SEQ ID |
| | | | | | NO: 1465) |
| | | AGCATATCTTCCAAGTCCTGGTCCAGAAGGGTGTCGATGGGCATATTAT | | | |
| | | CGACATCCTAAGAGACCTGTATACCAACGCTGGAACGTATCTGGAGTCA | | | |
| | | GGTTCCCAGCGATCGGGATTTATTAAGATCCTCAGGGGAGTGAAACAR | | | |
| | | GGGGACCCACTCTCTCCCATCCTGTTCAATCTTGCATTAGATCCTTTGCTG | | | |
| | | TGCCGCCTGGAAGATCGGGGCCTCGGTTATAAGTATGGAGACCAACAA | | | |
| | | ATAWCATCGTTGGCATTTGCGGATGATCTCGCCCTGCTCAGCGACTCTT | | | |
| | | GGGAGGGCATGCAGCAGAGTATTCGGGTGGTAGAGGAATTTTGTCAAC | | | |
| | | GGACCGGGCTGCGGGTTCAAGCGCCGAAATGCCACGGGTTTTTGATCA | | | |
| | | GGCCAACTAAAGAGTCATATACCATCAATGACTGTGACCCGTGGACGAT | | | |
| | | TGCAGATATGCAATTGGATATGATCGATCCGGGCAGTTCCGAGAAGTAT | | | |
| | | CTTGGCCTAGGGATAGACCCATGGATTGGTCTATCGAGACCGGAACTGT | | | |
| | | CCGAGGTGCTGACCCGCTGGGTGAAGAACATCGGGGGCGCCCCTTTGA | | | |
| | | AGCCACTCCAGAAGGTGGACATCTTGAGGAGCTACGCCCTNCCAAGGC | | | |
| | | TGCTGTTCATTGCGGATCACGCAGGCCTGAGCGCCACCTGTTTGCATTCC | | | |
| | | CTGGACCTTTCGATAAGATCTGCCGTCAAGGGCTGGTTACATCTACCGC | | | |
| | | CTAGTACGTGTGACGCTATTATTTACGTCAGCTACACAAGGACGGCGGGCT | | | |
| | | GGGTCTTCCCCGTCTGGCGAGCCTAATTCCAAATGTACAGGCTCGCAGG | | | |
| | | TTGGTGCGGATCGCCCAATCGGAGGATGATGTCATCAGGAGTGTGGTA | | | |
| | | CTCCAGGAGGGTATCCAGGAGGAGATCCGGAAGGTCTGGATCTCGGCT | | | |
| | | GGGGGGCGACCGGAGAAGGTTCCATCTGTGACGGGGGGAGTTCCCAGT | | | |
| | | GATGGAGGCTCAGGCGGCTGACGAGGCCCTATCCGAGTGGGAGAGGC | | | |
| | | GAGCTCCACGAACCATCTATCCCATTCCCTGTAAGTGGAGGAAGAGAGA | | | |
| | | AATGGAGAATTGGACCAATCTAAAATCGCAAGGCCACGGGATTCGGAA | | | |
| | | TTTTGAAAATGACCGAATCAGTAATGATTGGCTCCTGCATTATGGCCGC | | | |
| | | ATTCCCCACCGCAAACTAATAACAGCTATCCAGTTGCGGGCCAATGTCTA | | | |
| | | TCCCACWCGGGAGTTTTTGGCCCCGCGGCCTGGGCGAGGGCGCACCCAG | | | |
| | | GGGATGTAGGCACTGTCCCGCGGAGTGGGAATCTTGTTCCCACATAATT | | | |
| | | GGCTACTGCCCGGCTGTCCAGGAGGCCAGGATCAAAAGGCATAATGAC | | | |
| | | ATCTGTGGTGTGCTGGCTGAAGAGGCNAGAAAGCTGGGATGGGTGATA | | | |
| | | TTTATAGAGCCCCATCTCAGAGATAACACCAATGAGCTCTTCAAGCCAG | | | |
| | | ATTTGGTTTTGGTGAAGGGATCCTGTGCGAAGGTAGTGGATGTAACCAT | | | |
| | | CCGCTACGAGAGTGGGTTAACCACCTTGAGTGACGCCGCGGCAGAAAA | | | |
| | | GGCTAGGAAGTATCAACATCTGGCAGGGGAGGTGCGGGCCCTAACATC | | | |
| | | GGCCACTACTGTAGACTTCCTGGGTTTCCCTATTGGCGCTAGAGGGAAG | | | |
| | | TGGTACGTTGGTAATAATGGACTCCTTTCCGACCTTGGGTTCTCCACTAG | | | |
| | | CCGTGTAGTGCGGATAGCGAGGGCCCTCTCTAAAAAGGCTCTCCTATCG | | | |
| | | TCCGTGGACATTATACATATTTTTGCGTCTCGCGCTAGACAGGCCCAAAC | | | |
| | | GTCCGAGTAGGGGGGCTTGGCATTTCTCATTGCCTGCTCCTGAAAGGATA | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | TGGGTCCTGCGTCGCGTGGTAGGCAGACCCATTCGTCCGAGTAGGGGG | | | |
| | | | 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 |
| | | CCTGCCTACGATTTCCATGGCGAAAAAAATTTGCATCCTGGCTTTCGCCGA | AGCGGA | GEKICILAFADDLTLV |
| | | TGATCTGACGCTGGTGGCTGACAGCGCAGCTGGTATGAAGATCCTTCTA | ACTGGA | ADSAAGMKILLKAAC |
| | | AAAGCGGCTTGTGACTTCCTGGAGGAATCTGGAATGTCACTTAATGCAG | CCGCTTA | DFLEESGMSLNAEKC |
| | | AGAAATGCCGCACTCTCTGTATTACAAGATCTCCCCGAAGCCCGCAAGAC | AGCCAG | RTLCITRSPRSRKTFV |
| | | TTTCGTCAACCCAGCTGCCAAATTCATCATCAGCGATTGGAAAACGGGT | AAATAG | NPAAKFIISDWKTGIS |
| | | ATCAGCTCAGAAATCCCCTCCCTCCCTGTGTGCGACGGACACCTTTCGTTTCCT | TAAAGT | SEIPSLCATDTFRFLG |
| | | GGGGCACACCTTCGATGGAGAAGGAAAGATCCACACATCGATACGGAGGA | AACAA (SEQ ID NO: 1344) | HTFDGEGKIHIDTEEI |
| | | AATTCGATCCATGCTCAAATCGGTGAAGTCAGCTCCACTGAAACCGGAA | | RSMLKSVKSAPLKPE |
| | | CAGAAGGTGGCTTTGATACGGTCACACCTTCTTCCCCGCCTTCAGTTCCT | | QKVALIRSHLLPRLQF |
| | | GTTTTCTACAGCTGAAGCTGACAGCCGGAAAGCCTGGTTGATCGATTCC | | LFSTAEADSRKAWLI |
| | | ATCATCAGGGGGGTGTGTGAAGGAGATCTTGCACTCAGTGAAAGCTGGT | | 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 |
| | | CTCAAGCCCGACCTCCTACTCATCAAGCCGACACTGCATTCATTGTGGA | | RKEGYQVMIEPKVST |
| | | TGTAGGCATTGCATGGGAAGGTGGACGCCCACTAAAGCTGGTCAACAA | | PVGALKPDLLLIKADT |
| | | AATGAAATGTGACAAGTACAAAACTGCCATCCCGGCAATTTTGGAAACA | | AFIVDVGIAWEGGRP |
| | | TTTCACGTTGGCCATGCTGAGACGTACGGCGTTATTCTGGGCAGCCGCG | | LKLVNKMKCDKYKTA |
| | | GATGCTGGCTCAAGAGCAACGACAAGGCGTTGGCATCAATTGGGCTCA | | IPAILETFHVGHAETY |
| | | ATATCACACGGAAGATGAAAGAACACCTGAGCTGGTTGACGTTTGAAAT | | GVILGSRGCWLKSND |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | TATATTTATAACTCAAATAAGCCGGATTTATAACTCATTCATGAAAAAAT<br>GAGGTTTTTGTTTTCTTTTTTCCTTTTACCATTCTTGTTCCATTGTTGTTATT<br>TGCTTTAATCCTGTATTTTACCGCCGGCAATTCCATTGTTATTATTACTGT<br>TACTGTTATTATTGTTACTATTGTTTTTACTTTTACTTACTACTGTTATTAT<br>ACTTTAATTCGTTAACTTACGTTATTGTTACCACTACTTACTTTGCTCTCTC<br>GCAAACGTTCGTTGTTGTTTCTTTTGGACCAGGTTTAGAGAAATCGCACG<br>CACAGCGGAACTGGACCGCTTAAGCCAGAAATAGTAAAGTAACAA<br>(SEQ ID NO: 1098) | | | | KALASIGLNITRKMKE<br>HLSWLTFEIIFITQISRI<br>YNSFMKK (SEQ ID<br>NO: 1466) |
| R2 | R2-1_TUr | scaffold_6 | Tetranychus urticae | CTCTCTTATTTTAACATATTCGATGTACTCGTACATTGAATATGCTTTTATT<br>TTTTTTCAAAGTTTTTTTGGGTGCATACCCCTGGAAAATTCTGAGATGTAT<br>AAATCTCCCATCAGCTTTGGCTGAAACGTTGGCTAAGTTTTGTAGGTTGT<br>TTGCCCCCTACTACTTAGTCGCAAATGGTATTTGCTAACAGTTGTTAAAT<br>TGTTACATTTACAAGTCCTATCCAGTGCCTCCTCGTGGCGCTACCCGGTA<br>ACACTTAGAGTAATCTGAGTGGCTAAACTGGAAGGGCGGAAAATGCAA<br>ACAGGCGGTTGGTAGATGCTTCGGCATTTTGCCAAAAATCCACGGCTTT<br>TTAGCCCAACAACATCAGGGTGATGGACCCGCCAGCTTGTGGTCAGGAT<br>CCCATCCATGAATAAAGCATGGCTCTGCTTCTGGTGCATCCTCAACGGG<br>ATCGGCTTCGGCTGGATGTAAGTCTTGCGGAGGCATGTGCATCTTGGGT<br>GGCCTAACAAGCCATTCAAGAGAAGGAGGGCTCTCGCGAGGGAGTTCT<br>CAACTTAAGACCGTTAAGCCACAAAACGAAGAAGATAATGGAACTACCC<br>AGTTGAAGGCTGGGAGTGCGGACTCTTTCCCAAGGCCAAGTGGTGATTT<br>GAACCCAGAGGAACCACTGAGCATTGATATCTGTCCAGTTTGTTTCAGA<br>CAGATGAAGAGTTATTTAGGCGTTAGAGTGCATATGCAGAAGATGCACT<br>TAGAAGAATATAATGCGTCTATTCCGGACCCAGTCGTTTCCCACACTAG<br>GTGGTCCGATGAAGAGGCAGCACAATTAGCTTTTACTGAAGCTAAGATT<br>GAGGTCGACAAGCTTTTACCTCGTGGCAAAGGCATAAATAAGTTCCTGT<br>TAGAGCTTTTGCCTGGTCGAACCTTAGAGTCTATCAAGTCTCACAGAAA<br>AAGACAATCGCATAAAGACTTGGTTCGCAAATATGTAAAGGAGTTCGTT<br>GATACTTTAGCTGCAGATAACGATGACGATACCATCATCTGTCAAGATA<br>ATGGTGACATTTTTAATGACCCCATCGTTGGAGCAACTGATTCTCAGTCT<br>GAGACTGAAACTGTGGCTGATCCTGCTGAATTCAAGACTTTTATTGAGTT<br>AGCAGACGATCCCACACAAAACCAAAAGTCGTAGCAAAACTCAGAAACTTG<br>ATAAAAGATAAACCAAAGAGCGAGATCCTTGGAAGTGACATTCTTGTAC<br>GAATCCTGCGAAGAACTCTTCATGGCTTGCCTGTTGAGGACGAGTTGGA<br>CCAGTACTTGGAGGTTTACTTCACTGGGAAGATAAAGCAGAGACGTTCT<br>AAGACACAAACTGCCTTATCTAAGAAACAGATTAAGCAGAGAGATTATG<br>GACGTCTTCAGGAGTTATACTCCCGGTCTAGAAAAAGGTGTGCAAATGA<br>GATCTTGAACCCTACTTCTATGAGTGGTGGTTTTGGTCATCAGGAGCTAT | CTCTCTTAT<br>TTTAACAT<br>ATTCGATG<br>TACTCGTA<br>CATTGAAT<br>ATGCTTTT<br>ATTTTTTTT<br>CAAAGTTT<br>TTTGGGTG<br>CATACCCC<br>TGGAAAAT<br>TCTGAGAT<br>GTATAAAT<br>CTCCCATC<br>AGCTTTGG<br>CTGAAACG<br>TTGGCTAA<br>GTTTTGTA<br>GGTTGTTT<br>GCCCCCTA<br>CTACTTAG<br>TCGCAAAT<br>GGTATTTG<br>CTAACAGT<br>TGTTAAAT<br>TGTTACAT<br>TTACAAGT<br>CCTATCCA<br>GTGCCTCC<br>TCGTGGCG | TGATGT<br>ATCCCTT<br>CAATAT<br>ATTGTA<br>ATCCTCA<br>TTCGTCC<br>CTATCCT<br>TTCATTT<br>GATAAA<br>AGAAAC<br>TTTGTTG<br>CTCCTTT<br>TAATAG<br>TTGGTC<br>CCTCCTG<br>TCCCTTT<br>TCTGGA<br>ACCTGG<br>TTGTATC<br>GATTATT<br>GAAAGT<br>TGCAAT<br>AAACGG<br>ATTTAA<br>(SEQ ID<br>NO:<br>1345) | MCILGGLTSHSREGG<br>LSRGSSQLKTVKPQN<br>EEDNGTTQLKAGSA<br>DSFPRPSGDLNPEEP<br>LSIDICPVCFRQMKSY<br>LGVRVHMQKMHLEE<br>YNASIPDPVVSHTRW<br>SDEEAAQLAFTEAKIE<br>VDKLLPRGKGINKFLL<br>ELLPGRTLESIKSHRK<br>RQSHKDLVRKYVKEF<br>VDTLAADNDDDTIIC<br>QDNGDIFNDPIVGAT<br>DSQSETETVADPAEF<br>KTFIELADDPTKPKVV<br>AKLRNLIKDKPKSEIL<br>GSDILVRILRRTLHGL<br>PVEDELDQYLEVYFT<br>GKIKQRRSKTQTALSK<br>KQIKQRDYGRLQELY<br>SRSRKRCANEILNPTS<br>MSGGFGHQELSEFW<br>TKTFGPDEQPTLGEV<br>EIIPKENCWWDIFSPI<br>SSDEIKASYPSIGKAA<br>GPDNFSAYQLRKVPV<br>WHLECLYNIFAFYKDI<br>PSRLKDAKTILIPKKD<br>NAESPGDFRPITLSSII<br>TRHFHKILATRVNNF |

| | | | | | |
|---|---|---|---|---|---|
| | | 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 |
| | | GAAAAGAAAAGCTGCAGGTCCATGAAAACCCAATATTTAGAATATTGGC | ATGTAAGT | | VWRGRGADEKSLRC |
| | | AAAGCTATCTACAGTAAGCAAGGAGTTGGAAAATTGTAAAAAGATAGC | CTTGCGGA | | RAGCNARETLNHVS |
| | | TTCCAAGACGACAGACATTCAAGAAAAGACTTTCAAGGAAATGCTTGCG | GGC (SEQ | | QSCFRTHRVRTARHD |
| | | ACTTATGATGGTTTGTCATTGAAAGAAGCTCGAGCGGTACCTGAGGTGC | ID NO: | | KILDFICERLDVVGVK |
| | | ACAAATGGGTGGACAGCTATGATAAGCGTTATAAGTTTGCAGGCCGAG | 1222) | | YVREKPISFPGKKLIPD |
| | | ATTTCGTTCAAGTTATACAGGCTCGTTTCAATGCGTTACCAACGAGATCT | | | LIVENTDQALVLDLQI |
| | | CGAGTATGGCGGGGGCAGAGGTGCTGATGAAAAATCGTTACGTTGCAGA | | | VGDNSELPLDERGKN |
| | | GCAGGCTGCAACGCTAGGGAAACGTTGAATCACGTTTCACAATCTTGCT | | | KVIKYNCSEMQELYK |
| | | TCAGGACGCATCGAGTACGAACAGCTAGACACGACAAAATCCTGGATTT | | | RKKKTLAVKALTLHYK |
| | | TATTTGTGAAAGGCTGGACGTAGTCGGTGTTAAATACGTACGAGAAAA | | | GLMAPETSNILRSFG |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GCCAATCAGCTTTCCTGGCAAAAAGTTGATTCCTGATTTGATTGTGGAAA<br>ACACGGACCAAGCTCTTGTGCTCGATCTCCAGATTGTTGGCGATAACAG<br>TGAACTTCCACTTGATGAAAGAGGCAAGAACAAGGTTATCAAATACAAT<br>TGTTCGGAGATGCAGGAATTGTACAAGAGGAAGAAAAAGACTTTGGCG<br>GTTAAGGCACTCACGCTACATTATAAAGGGTTGATGGCACCGGAAACAA<br>GCAATATCCTAAGGTCATTCGGATTCAAGTCTAAGGATTTGGAAAAGAT<br>GGCGTATATGGCGTTGTTTGGTACGGTGGCTGCTTGGGGCATTTTCAAT<br>CGCTCGACGGAGACAATGCGTTCAGTCGCTAACTGGCCTAGACCTGAG<br>GAATTATGATGTATCCCTTCAATATATTGTAATCCTCATTCGTCCCTATCC<br>TTTCATTTGATAAAAGAAACTTTGTTGCTCCTTTTAATAGTTGGTCCCTCC<br>TGTCCCTTTTCTGGAACCTGGTTGTATCGATTATTGAAAGTTGCAATAAA<br>CGGATTTAA (SEQ ID NO: 1099) | | | | FKSKDLEKMAYMALF<br>GTVAAWGIFNRSTET<br>MRSVANWPRPEEL<br>(SEQ ID NO: 1467) |
| R2 | R2-2_DWi | . | Drosophila willistoni | GAAGCTGGGAAGCTGGGTCGGATGAGCGCAGAAGGGGTGTTCTTCGG<br>AGCACTGTAATTCATAAGTCGTAAGTCTGATCAAGTCGACTCGGAACCT<br>CTTCGTGGTGTTTCCTGGGTGCTGTTGAGTTCCTAGTCTCTAGGTTCTCT<br>CCAGTAGCTAATTCGAGCGGCGAAGCAACTCTTGGGGTTACCAGAACCT<br>TGAGCCAAGTAACGTTGGTCAAGATATGAATACTGTGCCTCGGATAAAC<br>AACACAACAACTCCAGCGACTTCCCGTCCTGGAGACCAACCGAGAG<br>AGGCTATAGCAGTGGTAAATCTCGCGGGAGAGATTCCCTGTGCAGTAT<br>GTGGGCGCCTCTTCAATACTAGAAGGGGGTTTGGGGTTCATATGTCACA<br>TCAACACAAAGACGAACTAGATACGCAACGTCAGCGTGAAGATGTAAA<br>ACTCCGATGGAGTGAGGAAGAAGCGTGGATGATGGCGAGAAAGGAGG<br>TGGAGCTCGAAGCAAGTGGTAATTTGAGATTTCCTAATAAGAAGCTAGC<br>GGAAGTATTTACTCACCGTAGCTCCGAAGCAATTAAATGTTTTCGGAAG<br>AGGGGTGAATATAAGGCAAAACTGGAGCAGATCAGAGGGCAATCTACT<br>CCCACCCCAGAAGCGTTGGACTCAATTACCTCACAGCCTCGCCCTAGTTT<br>ACTCGAGCGAAACCACCAAGTATCATCGTCGGAAGCGCAACCAATCAAT<br>CCATCAGAAGAACAGTCGAACTGGGAAATCATGCGGATACTACAGGGC<br>TATCGCCCCGTAGAATGTAGTCCCCGGTGGAGAGCCCAGGTCTTGCAAA<br>CTATCGTAGATAGGGCGCAGGCCGTAGGGAAGGAAACCACTCTCCAAT<br>GCTTATCCAACTATCTCCTGGAAGTATTTCCATTACCAAACGAACCACAC<br>ACCATCGGTCGGAGCAATTTGCGAAGACCTCGAACTAGGAGACAGTTA<br>AGACAACAAGAGTACGCACAGGTTCAGCGTCGTTGGGATAAGAATACT<br>GGGAGATGCATTAAATCCTTGCTTGATGGAACAGATGAGTCGGTTATGC<br>CAAACCAAGAGATAATGGAACCCTATTGGAAACAAGTAATGACGAATCC<br>CAGCACATGCTCTTGCGAAAACACAAGATTCCGTATGGAACATTCGCTT<br>GAGACGGTTTGGTCAGCGATAACGCCACGCGACCTGAGGGAAAATAAG<br>TTAAAGTTGTCAAGTGCTCCGGGTCCTGACGGTATCACTCCAAGAACAG | GAAGCTG<br>GGAAGCT<br>GGGTCGG<br>ATGAGCGC<br>AGAAGGG<br>GTGTTCTT<br>CGGAGCAC<br>TGTAATTC<br>ATAAGTCG<br>TAAGTCTG<br>ATCAAGTC<br>GACTCGGA<br>ACCTCTTC<br>GTGGTGTT<br>TCCTGGGT<br>GCTGTTGA<br>GTTCCTAG<br>TCTCTAGG<br>TTCTCTCC<br>AGTAGCTA<br>A (SEQ ID<br>NO: 1223) | TAGATG<br>TACTAAC<br>CTCTAGT<br>TTCTCTA<br>TACTTTT<br>GCCTGC<br>TACCTTG<br>GCATTA<br>CATCTAA<br>AAAGGT<br>ACAAAC<br>ATCGCA<br>TTGGCA<br>AAAAGA<br>GGTGGT<br>TTTAGTA<br>CATAGG<br>CGCTGT<br>GGGACT<br>TCATTGT<br>CCCGAT<br>GATGCA<br>GCGAAT<br>CGTGCA<br>TACGAG<br>ATTGTCC | FERRSNSWGYQNLE<br>PSNVGQDMNTVPRI<br>NNTTTTPATSRPGDQ<br>PREAIAVVNLAGEIPC<br>AVCGRLFNTRRGFGV<br>HMSHQHKDELDTQR<br>QREDVKLRWSEEEA<br>WMMARKEVELEAS<br>GNLRFPNKKLAEVFT<br>HRSSEAIKCFRKRGEY<br>KAKLEQIRGQSTPTPE<br>ALDSITSQPRPSLLER<br>NHQVSSSEAQPINPS<br>EEQSNWEIMRILQGY<br>RPVECSPRWRAQVL<br>QTIVDRAQAVGKETT<br>LQCLSNYLLEVFPLPN<br>EPHTIGRSNLRRPRTR<br>RQLRQQEYAQVQRR<br>WDKNTGRCIKSLLDG<br>TDESVMPNQEIMEP<br>YWKQVMTNPSTCSC<br>ENTRFRMEHSLETV<br>WSAITPRDLRENKLK<br>LSSAPGPDGITPRTAR<br>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 |
| | | ACAAGAGCCCCTCTGAAACCACACAACAGAAGTTATTTGCACATAGGACTG | | | | GPKLLRLTRAPLKPQ |
| | | TCCTTATCCCACAACTCTATCACCAACTAACACTTGGGAGTGTGATGATA | | | | 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 TCCTCGGCGATCAAAANAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1100) | CGGTGCGT TCCCTTGG GTAAGGA ACACGAGT CTTAGTGG CTTGACCTC CACGTGGT CCCGCTGG TAACATCA TCTCTTGA TGATGGCT AACAAGGCT ATCTGTGC TTTGTCT GTTTGTCT GTTCCGCT GTTTATCT GTTTCTCT GTTTCTCT GTTTCT

| | | | | | |
|---|---|---|---|---|---|
| | | 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 | | | SFCGLSGLEVQARKC |
| | | GAACAACTGCGACCCATGGACTATCAAAAATAAAGATGTCCATATGATC | | | QGFMISPTKDSYTVN |
| | | CAACCTGATGAATCAACGAAATACCTTGGTCTAAAAATTTGCCCTTGGAC | | | NCDPWTIKNKDVHM |
| | | TGGCATTATACGGTCGGATCTACATGTTCAACTAAAGACACGGATCTCG | | | IQPDESTKYLGLKICP |
| | | AAAATCGATGAGGCGCCTCTGAAACCGACTCAGAAGGTCGAACTCCTCA | | | WTGIIRSDLHVQLKT |
| | | ATGCCTACGCCCTACCCAGATTATTGTACCCTGCTGACCACTCGGACTGC | | | RISKIDEAPLKPTQKV |
| | | AAGCAATCAACTCTCCGTGTGTTGGACCAAGAAATAATAAAGGCGGTAA | | | ELLNAYALPRLLYPAD |
| | | AAGGATGGCTCCATCTTCCCGCGTCAACCTGTGACGGGCTGTTGTACGC | | | HSDCKQSTLRVLDQE |
| | | CAGAGCCCGAGACGGAGGCCTTGCCATCTTGAAACTGGAAAATGCAATT | | | IIKAVKGWLHLPASTC |
| | | CCTTCGGTTCAAGTTAGAAGGCTGCAACGTATTGCAAACTCCTCTGACG | | | DGLLYARARDGGLAI |
| | | CTATCGCTCGAAACATTGCGTCCTCGCAGGGTGTGGAGGAAGAGTACC | | | LKLENAIPSVQVRRLQ |
| | | GAAGTCTGTGGGTACGGGCAGGGGGTGACAGCGAAGCAATCCCAACG | | | RIANSSDAIARNIASS |
| | | TTCTTCTCAGGGGTTCGGAATCAAAAGAGCCCGTGTATCCGAGACCCT | | | QGVEEEYRSLWVRA |
| | | GCGATTGGAGGAAACGCGAATCTCGGAGACGGTGCGAAAAGCCGGTTC | | | GGDSEAIPTFFLRGSE |
| | | AAGGAAGGGGCATTGTAAACTTTGCGCAAGATAGAATCAGTAATGCAT | | | SKEPVYPRPCDWRKR |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GGTTGGGGCCACGGTGCGGCTTTAAACAATGCTTCTTTATCGCAGCATT<br>ACAATTAAGGGCAAATATTTACCCAACAAGAGAAAGCATAAACAGAGG<br>CAGAGATGGTGCCTCACGGTCCTGCAGGAAATGCTCTGCCAGGCTGGA<br>GTCTCTCTCGCACATTCTTGGTCAATGTCCCGCAGTACAAAAATTCAAGG<br>ATTGCGCGACGCAATAAGATCAGCGACATTCTAGCTGACGAAGCGGCG<br>AGACTGGGCTGGTGGGTGTACAAAGAGCCACGGTTCACATCTGAAGCC<br>GGAGAGCTAAGGAAACCTGCCCTTGTGTTTGCCAAAGGTGAGGAAGCG<br>CTTGTTATTGATGTCACCGTCCGGTTTGAGCTCTCGAGGAAAACCTCATC<br>AGAGGCTGCCTCGCACCAAGTTGCGTACTACACCCCCCCTTGTGATCAA<br>GTCAAAGTGCTGACGAAGGCAAGCAATGTCACATTCTTTGGATTCCAGG<br>TTGGGGCAAGAGGGAAAGTGGCCCCTTGAGAATAATGAGGTGCTAACC<br>TCCCTGGGCCTGACCAAACCCAGAACACATCACTGGCCAAGATGATTTC<br>CCGCAGCACGTTGCTTTTCTCTCTCGATACCCGAGAATGTTCTGCGGAAC<br>TACGCAGTCTATGAACAGTCACAGACAACCTCTGATCCAAG (SEQ ID NO: 1101) | | | | ESRRRCEKPVQGRGI<br>VNFAQDRISNAWLG<br>PRCGFKQCFFIAALQL<br>RANIYPTRESINRGRD<br>GASRSCRKCSARLESL<br>SHILGQCPAVQKFKD<br>CATQTEAEMVPHGP<br>AGNALPGWSLSRTFL<br>VNVPQYKNSRIARRN<br>KISDILADEAARLGW<br>WVYKEPRFTSEAGEL<br>RKPALVFAKGEEALVI<br>DVTVRFELSRKTSSEA<br>ASHQVAYYTPPCDQ<br>VKVLTKASNVTFFGF<br>QVGARGKVAP (SEQ ID NO: 1469) |
| | | | | | | | |
| R2 | R2-2_SMed | . | Schmidtea mediterranea | AGTCATAGGGTGAACTGCAATTCTGACACGATGACCGAGCTGTGTCAGT<br>TTGCAGCTAGTCGCTAAAGACTCGATCAGTCCGCCAAGTGAGGTGGCCG<br>GGTATCTGCAGCACTAGAGCCACTGGTATCAAGAGCAGAGATACGCGA<br>GTGGAAGTTGAGTACGACTACCTTCACGGGGTCCTCCTGATAACCACAG<br>TGGACTGTGGGAACTAAATGTGTGCTCAGCGTTCCCTACTTTCTCGTAG<br>GGTAAAGGGTATGATGATAACCCAGAGAATATCCCATGGGAGATATCCATG<br>GAAAAAGCACCACGTTAGACAATCCGATGGTCTAACTCGGCTCCGAGG<br>GGCTAACTATCCCAAAGGGCTTAAAGAAAAGAACTGGTTACAATCAAAA<br>ATCTATTTGAAGAGTCAGGTGCCACCGCGCCTGCACCAGTTCCCCTAGA<br>AGTTGCAGTTGAGGTACACCAGTCATCGAGTGTTCCCGAGATAACCGAC<br>GAATCTACGACTACTCAGGAAGGAAGCTATTCAGAACCACCGATACACC<br>GATGTGAGAACTGTGGAAGAGAATTCAGAACAAGAGCAGGAGTTCAAC<br>AGCACCGGAGAAAAGCTCACACCAACGAGTTTATGGAGGAGAAAGAGA<br>AGGCAGCTCCAACCAAGAAGCTTCGATGGACAGACGAAGAAAAAGAGA<br>TCCTCATCGAAAGCGAGATAAAGATCATCAAGGAAGGATCCCTTAAAGA<br>ACAACACGAAATAAACAAGATTCTAGCTTCAAGAATGCCTGGAAGATCA<br>CAAGACGGAATAGCCAAAATTCGCCAGAAACAAGAGCACAAGGCTGAA<br>ATACAGAGAAGGCTACATGGGACCGTCACTACCAACGAAACCAGAGGA<br>AACAGAACCAGCGAAATTACCGAACCAATAAGAAGCTTACCAATAAACA<br>CCAAAACCTGGAGCGAAGACGAAATGAAAAGAATGCTAGCCGAAGAA<br>GTGAAGTTGAGAACCAAGAATGAAAAGGATATCAACAAGAAGCTAGCA | AGTCATAG<br>GGTGAACT<br>GCAATTCT<br>GACACGAT<br>GACCGAGC<br>TGTGTCAG<br>TTTGCAGC<br>TAGTCGCT<br>AAAGACTC<br>GATCAGTC<br>CGCCAAGT<br>GAGGTGG<br>CCGGGTAT<br>CTGCAGCA<br>CTAGAGCC<br>ACTGGTAT<br>CAAGAGCA<br>GAGATACG<br>CGAGTGG<br>AAGTTGAG<br>TACGACTA | TAGAAG<br>GGAAAC<br>AAAGGA<br>AAAACG<br>AAATGA<br>CTGGAA<br>ACTATG<br>AAGGAT<br>ATAGCT<br>GAAAGC<br>CGCAAG<br>GAAGGC<br>TAAGTC<br>CTGAAA<br>CCGATCT<br>ACATCTT<br>CGATCC<br>CAAGAG<br>GAACTG<br>TGGGTT<br>AAGCTT | RKELVTIKNLFEESGA<br>TAPAPVPLEVAVEVH<br>QSSSVPEITDESTTTQ<br>EGSYSEPPIHRCENC<br>GREFRTRAGVQQHR<br>RKAHTNEFMEEKEKA<br>APTKKLRWTDEEKEIL<br>IESEIKIIKEGSLKEQHE<br>INKILASRMPGRSQD<br>GIAKIRQKQEHKAEIQ<br>RRLHGTVTTNETRGN<br>RTSEITEPIRSLPINTK<br>TWSEDEMKRMLAEE<br>VKLRTKNEKDINKKLA<br>EIFPNRTMGSIKSKRT<br>KDKDYQDLVKLTMQ<br>TISENPDNETDFNTS<br>NTENNSTDAEKEVKN<br>YLNMLLLTINEEEWL<br>TSTLKEAATLALQGKK<br>TEASEKLNEYASKTLF |

| | | | | | |
|---|---|---|---|---|---|
| | 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 |
| | AAGGCAAGAATACGCTGAAATACAGAAACTCTACAAAAAAGAATATTTCA | CTACTTTCT | CTCCTCG | | SMAILFNTCLLTSFLP |
| | AGCGCAGCCGAAAAAGCAATCAATGGGAAATGGTCCATAAAACCCGAA | CGTAGGGT | TGGTATT | | LPLKIARTILIPKQENP |
| | GAAGAATACCACACAACAATAAGGATCTCATTAAAGCATGGAAACCAATAC | 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 |
| | GATTAATACACAAATTGGTATTAGGAAGAGTGGCCAAGAAGTCATTAAA | | | | YTKQVLGKWSRPTKI |
| | AAGAATCGACCAGAATATAAGAAAGAAAGTGAGGAATTGGCTACATCT | | | | GETLATNKSQLKEAF |
| | CCCTAAAGACACGACAGCAGCATTCATACACGCTGATGCAGGAGATGG | | | | RKQMLITLDGKGLKD |
| | AGGGCTTGGAGTACCAGCGTTAGAACACACAACAATTCCTCTACTGAAAAGA | | | | 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) | | | | |
| R2 | R2-2_TCas | . | TGGAAGACCCCGCCCATGAGGCTTGGAGAGTGTGATCCTGATCACACTT | Triboliu m castane um | TGGAAGAC | TGATGC | MKSRSFRRIGDCAAG |
| | | | GAAAAGTTATGCTGAGTACGTCGTGAGAGTCGGTAACTGTCCCAGGAT | | CCCGCCCA | TCCTTTG | SSRRGVRLTGKAGRE |
| | | | GGTCTGGGATAGGCTAAACCTCAGCAGGGGAAAGTTGTAGGGGCCTGC | | TGAGGCTT | GTTTTAC | GRFAASPHLSPRYLA |
| | | | CACCCCTACACTTTTATAGATATGGCATTCGATACCTCAAATAGAGCCTC | | GGAGAGT | CATCTGT | GSVSGNVPSVPPGP |
| | | | GGACTTGGAGGAGCATGGTTCCCCTCCTCCTCGTACTAGACCTGGAACC | | GTGATCCT | GGGGGC | GLGAGAPAFAAGRN |
| | | | AACGGTCTTGACAACCCCATTGGACCTACGGGAGCGGACCATGCTCATG | | GATCACAC | ATCGGT | ADGGPAQNPCPYCA |
| | | | GACATGGATTCCGAAGACGAAGCGGGGGGAACACGGACCCCCCCGCCGAT | | 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 |
| | | CCCCGGCCCCGGCCTGGGGGCCGGGGCTCCTGCATTTGCTGCTGGTCG | 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 |
| | | TTTCGGGCCAGTTGCTGGCTGCGGGATGTCTCTCCCCGTCAAGAG | 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 |
| | | TCCCGACTTGTTCACACAAAGTATCAAAGGGGCTTCTTAAGCGGCGTGGAC | GACCCCCC | | LLLNPNKCQALSLVAS |
| | | GGGATCGCAGATAATCTCTCGGTCCTTGACACAATGCTGACGATGTCCC | GCCGATAA | | GRDHKVKLVTKPTFK |
| | | GGAGATGCTGCAAGCATTTGCACTTGGCCGCGTTAGACGTGTCAAAAGC | TGCTCACT | | VGQNTIHQVDASSI |
| | | ATTTGACACTGTCTCTCACTTTGCCATAGTGCGTGCTTGCCGGTCCATATT | TAACGTCA | | WKYLGIQFRGSGMC |
| | | TGGGAGTGCGGAGACCGTGTTGGAGGAGGGCGGGAGGAGGCACTTCG | GGCGAACC | | GCGSEGVAAGLKRIT |
| | | TGCAAGTGCGGYGTGGGGTGAGGCAGKGAGACCCACTCTCCTCTCCT | CATCGAAA | | CAPLKPQQRMHLLR |
| | | CTTTAACCTGGTCTTGGACCGGGCTCTGAAGAGGCTGTCCACCGATGTG | TCATCTTG | | VFFLPKFYHAWTFGR |
| | | GGCTTTCGTCTAACGGATGCGACCAAAGTTACGGCACTCGCTTTTGCGG | ATGTTACC | | LNAGVLRRLDVVVRT |
| | | ACGACGTCGTTCTGTGCGCGACCACCGCTAGAGGACTCCAGACCAATCT | CTTTCAAA | | SVRTWLRLPHDIPVG |
| | | GGATGTTCTCGAAGCGGAGCTTCGCCTAGCCGGTCTCTTGTTGAACCCC | GCAGGTCA | | YFHAPTKSGGLGIPQ |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | AACAAGTGTCAGGCCTTATCCTTGGTGGCTTCCGGACGGGACCACAAAG | TGCGGCAT | | LSRFIPFLRLKRFDRLG |
| | | | TGAAGCTCGTGACAAAACCCACCTTTAAGGTGGGACAAAACACGATTCA | ATGTCTCA | | RSAVDYVRECAFTDI |
| | | | CCAAGTGGATGCGTCCAGTATTTGGAAGTACCTGGGAATCCAGTTTAGG | ATGCCGGA | | ADRKIRWCRERLSGI |
| | | | GGTAGTGGGATGTGCGGATGCGGTTCGGAGGGTGTGGCTGCTGGGTT | AAGGGTG | | VDQVAGGRDALDAY |
| | | | GAAGCGGATCACTTGTGCTCCGCTGAAGCCTCAGCAGAGGATGCACCT | GCTCCCGG | | WTAQLHQSVDGRAL |
| | | | GCTGCGAGTCTTTTTCTTGCCCAAGTTTTATCATGCTTGGACATTTGGGA | GCGTAACG | | RESASVASSTQWLRC |
| | | | GACTCAATGCGGGTGTACTGCGCCGTCTGGACGTTGTGGTCCGAACTTC | 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: | | |
| R2 | R2-7_MR | . | Megachile | 1226) | TAAATG | SGPATSTFGETKSRLC |
| | | | TCTAGTTAGCAAGCGGCCCCCTCTAATCGGGTCCCGCCACGAGTACCTTC | TCTAGTTA | TCGAAC | EPTSALGCRPGAVVI |
| | | | GGGGAAACCAAGAGCCGCCTCTGCGAACCTACCTCCGCATTGGGTTGCC | GCAAGCG | CGAATTT | QWAQIHKEKRKRIVG |
| | | | GGCCCGGTGCGGTTGTAATTCAGTGGGCCCAAATTCACAAGGAAAAGA | GCCCCCTC | | |

| | | | | | |
|---|---|---|---|---|---|
| | 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 |
| | | AGCCGGACCCTCGAATCTGTCAAAGGAACTCGGAAAAACCCCAAATACA | | 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 |
| | | GTGCTCAGTATGGGGTCCAGTAGAGAAAGAAGAAACTTCTCTCCTCCGTG | | 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 |
| | AAGTGTCGGGAGCACTATAAATTCCCAAAGAAGAAAAGAAAAAGAAAAA | 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 | | |
| | | ACTACTGGTCGACATTAAAAAAGACGAAGCAGCTGCCAGCTGATAAACAA | 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<br>CAGGTGTT<br>AAAATGAA<br>TCTGATAG<br>CTGGCGGA<br>TCGTCGAC<br>CCTCTTTG<br>ATGGCTCT<br>GCGCCAAC<br>GACTGGAA<br>AGAATAG<br>GAACGGA<br>AGTCTAAT<br>GGAAGGA<br>AAGTGTCG<br>GGAGCACT<br>ATAAATTC<br>CCAAAGAA<br>GAAAAGA<br>AAAGAAA<br>AAAAATAA<br>AAAACCCA<br>AATTAA<br>(SEQ ID NO: 1228) | TGACCT<br>GAACAA<br>AACGTG<br>TTGTCTT<br>GTCTTGT<br>CTAAAA<br>CTATTTA<br>TTCGAA<br>ATAAGG<br>GGAGGC<br>TAACTG<br>CCTGCA<br>AGTTGA | TFAPTHPMVRSGPC<br>RKTKRPGSDYRESLI<br>MDSGNNVASEPRGA<br>VDVTSAAPIGAELNA<br>EPCEGRNQRREAALS<br>AQTRRRNXARRARN<br>AQQADEPGDDEEIET<br>HGPLTIRTXEPMEIVA<br>IAKNPQACPKCLQGG<br>TQLLCMGSWELSRHI<br>NKEHPSVDVTWVCG<br>ACQRRCTTLRSWSC<br>HVLHCKGRQEPKDLP |
| R2 | R2B_NVi | . | Nasonia vitripennis | GACTAGACTATGGGTTCAGTCAGTCCCAAATAGCCGATCCTGGCGCGTC<br>CGGCAGTAATGCCACGTATGAGTCGGTTACCCATCTCTAAACGCGTAGA<br>GGTGGGGAGCTAAAGGCCAGGCGGTTTACCCGACGTCGAATTTCTCCA<br>GGTCTGTGTCAGTCGACGGAATAAAGGTACTACAACATCTACTATCTAT<br>CGGGATCGGAAGACGCCTTACAGCGTTTTCCGATTTTTGCTCTTTGAGCA<br>TTTTTCTTCAAATTGCGATAACCGACCCGATCACGCGGGGGCTTTGACAAA<br>GCAATGCGTGGTCGGTAAGATGGTTGCAATCTTTTCCACCTCGTTTCTTT<br>TACGGAACGAAAGCAATGCGTGTGGGGAACGTTAAAAACTCCCTTCATG<br>CATCCCAGGATTTATCCTGCTTACTGCAAAGCAATGCGTGTGGAGCGAC<br>TTTACCACGAGTCGCTCCACCGCAAAGCAATGCGTATCGCGCAAAAGCA<br>ATGCGTGTGGGGGACTTGTCAAAGATCCCCCGCCGCAAAGCAATGCGT<br>GTCGGCACCACGTAGAGCAAAGCGTGTAGGCAGACTTTGTCAAAAGTA<br>GTTCTGCCGCAAAGCAATGCGTGTGGAGATCTTCGCCGGTGAAAGCAAT | GACTAGAC<br>TATGGGTT<br>CAGTCAGT<br>CCCAAATA<br>GCCGATCC<br>TGGCGCGT<br>CCGGCAGT<br>AATGCCAC<br>GTATGAGT<br>CGGTTACC<br>CATCTCTA<br>AACGCGTA<br>GAGGTGG | | |

| | | | | | |
|---|---|---|---|---|---|
| | | 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 |
| | | GGACTAGACTCCCTGTATGACCAGCTCATCCAGATTATGCGGCACCCTTC | 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 |
| | GAACGAGGAGCCGGGGACGTAAAGATTGGAATGAAGAGAGAGGAGTCAAG | 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 |
| | CCAGGGGGTCGAGGAGTTCGCACACAAGATCCCGTCGGCAACTCATGGCT | 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) |
| | AAACCCGATCTCGTCGTGAAACGAATGAGGGAAGGGTGCACGTAATC | | CTTGATA | |
| | GACGTAACTGTCCGCTACGAGCACAGAACCTATCTGGATGAGGGCCGT | | TCAGTT | |
| | ACTGAGAAAATTGGCAAATATCGCCAAATTCTCAGCACGCTTCGGCGAG | | GTTTTGA | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | ATCTGCACTCGAACGCCGAGGAGGTCATTCCAATTGTAATCGGGTCGAG<br>AGGTGCAATTCCAAGGGAAACGCGGAAAGCCCTCTCGAAACTCGGAAT<br>TGGCAAGAGTGATTGGCTTACGATCTCACTAATAGCGCTGCGTAGCTCG<br>CTAGAGATCGTCAACGCCTTCATGGATGACTGACCTGAACAAAACGTGT<br>TGTCTTGTCTTGTCTAAAACTATTTATTCGAAATAAGGGGAGGCTAACTG<br>CCTGCAAGTTGAACGCGAAAGTTAGACCTTCCCACCTAAAGCCCAAAAG<br>TGATCGGGGAATGAATCCGCGGGTGACCCCAGAGTTGGGTAAACCCTT<br>GAAACGTTGGAGAAGCGGAAGAGAGTCCCGCCACCGAGCATCGAGTGC<br>TGCGGCGCCCGAATGAAACCGATCGCGGATGGTGCAAGTCGTAGGACG<br>GGGCACGACCTAAGCCTCTGTCACGGCGCGGCGAAGCCAGGAATCACCAT<br>GCAAAGGTGTGAACTGGGGCGGATACCTCCACGGGGTTTCCCTGGGCA<br>TCGCGCGAGCGATGGCCAAAGTCCGCTTTCTCAGCTACAAAACAAAAAT<br>GGTATGAGACTTCGTTAACACTAATTTTTCCGAGCCTAGCAGGCTCCCTT<br>GACAACGCTTATGAATCTGGAAAAGGACACAAAGTGGAAAAAGCGCTG<br>ATGGTGGACAAAAGTCAGTTGAGACTTGATATCAGTTGTTTTGACTAAG<br>AATTTTATTATCGTTGACTTTTAAATATTTTATTATTGACTGTTAATATACT<br>GACTTGGGACCAAGTCATCTCTGTTACCCGGTACCGGTTCCTGTCATCAA<br>ACCGGAAAGTCCGTCCCACGTAATGTGGTAGACGCAGGAG (SEQ ID NO: 1106) | | | CTAAGA<br>ATTTTAT<br>TATCGTT<br>GACTTTT<br>AAATATT<br>TTATTAT<br>TGACTG<br>TTAATAT<br>ACTGAC<br>TTGGGA<br>CCAAGT<br>CATCTCT<br>GTTACCC<br>GGTACC<br>GGTTCC<br>TGTCATC<br>AAACCG<br>GAAAGT<br>CCGTCCC<br>ACGTAA<br>TGTGGT<br>AGACGC<br>AGGAG (SEQ ID NO: 1352) | |
| R2 | R2Ci-B | AB097122 | Ciona intestinalis | CGACGGTGAACCACCTTGTCGCGGTGTAAGAGCTTTAGTGTCTCGAACA<br>AGAAATAGCTTGTGTGCTGTCCTTCTGGGCGGTGCACATACTTCTTAACC<br>TCCCGAGGCCATGCCGGCGGGGGGCTTTAGCCCCCGGCAGGTTTTACCAT<br>GCCGGACGGGTTCGAGAGGTAGAGGCCAAACTAAGAGTTCACCAGCAG<br>ACTTCGCACGCGGCTGGCCACTGGCCGAAGTTTAAACAACAGGGCCGC<br>ATCTTCCCAAACTCAATATATGGTGTTAAGTGAACCGTGCCGATGGGTG<br>AATGGCCATGGGTAAGTTGGTCTCTAACGGTCCTTGTGGAGAAGTGGC<br>GACCCTTCACAATACTTCAGCCTTACCCTATGCCCGGGCAACTTAGGGTG<br>GACGTATATCTCCCGCGGAAAACATCTTATCTCATGGATAAAAATATATA<br>CGAAAACACTACTTCTCCAGGAGGGGGACCGCTTTGCGGTGAGAAGAC<br>GCATCGTTCCGATGTGATCATCCCTCCACCGGGGTTCGCCCCGTCGACG<br>GACACTGCTTCCAATACATTAGGGGAAAATGTGGATGCATCTGCAACCA | CGACGGTG<br>AACCACCT<br>TGTCGCGG<br>TGTAAGAG<br>CTTTAGTG<br>TCTCGAAC<br>AAGAAATA<br>GCTTGTGT<br>GCTGTCCT<br>TCTGGGCG<br>GTGCACAT<br>ACTTCTTA | TGACAG<br>TAATAT<br>GAAAAC<br>ATCACAT<br>CTGACC<br>GGCACA<br>GAATCA<br>CCATGC<br>CGTAAT<br>GCACCC<br>AACTAA<br>GGATTC | MGEWPWVSWSLTV<br>LVEKWRPFTILQPYP<br>MPGQLRVDVYLPRK<br>TSYLMDKNIYENTTS<br>PGGGPLCGEKTHRSD<br>VIIPPPGFAPSTDTAS<br>NTLGENVDASATTSS<br>ANPLSQEPGWCESCS<br>KLFKSQRGLRVHQRS<br>KHPELYHSQNQPLPR<br>SKARWSDEEMVIFA<br>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 |
| | | | TGTGGATGGGCGTGGACTAAGCAATTGCAGTGATGTCCCGCGCCCAAAG | | | | 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 | | | | |
| | | | GATTCCAATGGGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA | | | | |
| | | | AAAAAA (SEQ ID NO: 1107) | | | | |
| R2 | R2C_NGi | . | Nasonia giraulti | CGGGTTCCCCCGACTTCGGCTTGCCGTGGTCTGGGGCTCACTGCTTTTTG | CGGGTTCC | TAGCGG | WVTSPRRPRYVGPQ |
| | | | | TGGAGTCATGGTTACATGGTGACCCTGGTTCCTCGCACCCCCGCTGGAA | CCCGACTT | ACTGGA | KKKASDGNDGRAAA |
| | | | | ACTATCTGGGGAGGCCATGATTGGGTAACGATAAAGGTCCTGGTCGTG | CGGCTTGC | CTGTCT | RAEPTNPGGPDRAD |
| | | | | TCCTCCTGAGATAGGCTGAATGGGTCACTAAGTGGCACCTAATGGGTTA | CGTGGTCT | GGAGGA | DDEGDVKFWCEFPG |
| | | | | CCAGCCCTAGGCGGCCGAGATACGTTGGTCCACAGAAAAAGAAAGCCT | GGGGCTCA | GTGTTTA | CDRFFMTRSGRGLH |
| | | | | CGGATGGAAATGACGGACGAGCCGCTGCTCGTGCCGAACCAACGAATC | CTGCTTTTT | ACTCGG | HKKGHPDWNDQRN |
| | | | | CGGGCGGACCAGACCGCGCTGACGACGAAGGGGATGTTAAATTCT | 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 |
| | | AGAGGCGGAGGGGGGGAGTACAAAAAATGCCAGGATCTATGGCGACGA | ACCTAA | | GECLPRELMEPFWN |
| | | AATAAGTCAACCTGTGTTCAGCGGATCCTTAAAGAGGATCTTTCGCAGG | (SEQ ID | | ATFTQNPGTAPVLPP |
| | | GTGAATGTTTGCCTCGAGAGCTGATGGAGCCCTTTTGGAATGCGACTTT | NO: 1231) | | PTEVYSSVWEPIRPE |
| | | CACCCAGAATCCTGGCACGGCTCCGGTGCTCCCTCCTCCCACGGAGGTT | | | NIKGNYPPQNTAAGI |
| | | TATTCTAGTGTTTGGGAGCCCATTCGGCCCGAGAATATCAAGGGCAACT | | | 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 | | | LISDAKPLIKDLELLT |
| | | 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 |
| | | | GGAGGAGTGTTTAACTCGGGTTCTCATGGGAACCCGACAACGTTGTTAT | | | | 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 | |
| | ATCGTACCGTGTTGATACCGAAAATCGACAATCCGACACATCACCAGGCGA | | 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<br>AGGCTATATACGTCATTTGATGGTAAAACACTTGCCTATAGCGAGTACTT<br>TACTCAAGGCGGCGGCTGGATCCGTGAAGATAAGATCCTTCAACCTGCC<br>CATGTTTTCGCTGAATGCATCAAGTTACGTATTAATGCTTTACCCACTAA<br>ATCTAGAGTTGCACATGGCCGTCCAACTAAAGATCGTTCCTGTCGTGCC<br>GGATGTCTTGACGTTCAAAAAGTGCCAACTATCGAAACAATCAATCATA<br>TAGCACAAGTTTGTCCTCGGACGCATGGTGCAAGAATAAAAAGACACG<br>ATCGGCTCGTTCAATTCCTTTCTCTCAACCTTAGGAAAAATCCAAAAAGA<br>AATGTTTTAGTGGAGTACAATTTCCGTACAGTCGCTGGTATCCGTAAACC<br>GGATATCATTGTTATTGAAGATACCCGTGCAGTTATCCTTGACGTCCAAG<br>TTGTTGGTGATTCGTCCAATTTGGAGATGGAATACCTGGAGAAATCCAG<br>GAAATATTCTAATGACGCTAACTTTATCAATGCGCTTCAAAAACTTTATC<br>CAACGGTCACTAACCTAACTTTTCATGCAGTTACCTTCAACAATAGAGGC<br>CTGATTGCCAAATCCACAGTGGCGGCCCTCCGGATGCTCGGCGTCCCAC<br>CTAGATGCATAATGATCCTTTGTGTGATTAGCCTTGAGAAGACCCTGGA<br>AGTCTGGAGGATGTTTAACCAATCAACTGCTTCCGCTCGGAAATGATAA<br>TCGCTCCATCCTGCAACTAATTATGAATGCAAATCTGTTAAGTGACATTA<br>GTGATACTTACCTGATACTTACCCTGGTATTTATTTGACCTATACTTACCC<br>TGGTATCTACCTGACATATATTTATCTAACCACCTACCTATGATGACTCCC<br>GCGGAAACTCTCACTTACCTTATTACCCACTTGGTCTTTTATTTTCTCGTT<br>CCTTATTACTTTGTTCCTTTGGTGTAGGGTTCTCTGGTTTTTGGAACGGCT<br>TCCTTAGCCGGAATTTTGTCTGATGTATCTTGCTTGTGTCCTTGAAATATA<br>CGACCCAGGCTTGCGTCATTTAGGCTCTGGGAAA (SEQ ID NO: 1109) | | TTGCGT<br>CATTTAG<br>GCTCTG<br>GGAAA<br>(SEQ ID<br>NO:<br>1355) | KFWSERLYTSFDGKT<br>LAYSEYFTQGGGWIR<br>EDKILQPAHVFAECIK<br>LRINALPTKSRVAHG<br>RPTKDRSCRAGCLDV<br>QKVPTIETINHIAQVC<br>PRTHGARIKRHDRLV<br>QFLSLNLRKNPKRNV<br>LVEYNFRTVAGIRKP<br>DIIVIEDTRAVILDVQ<br>VVGDSSNLEMEYLEK<br>SRKYSNDANFINALQ<br>KLYPTVTNLTFHAVTF<br>NNRGLIAKSTVAALR<br>MLGVPPRCIMILCVIS<br>LEKTLEVWRMFNQS<br>TASARK (SEQ ID NO:<br>1477) |
| R2 | R2LcA | Lepidurus couesii | TTTGGGGGTAGCAATTGATCGATTCCCGCCTCCTCGTGGCGCTACCCCGG<br>GATAGCCTCAAAGAAATTTGACGGTAAAGCAAAGAGGAATTGATCACC<br>CAAGGCAGTACATCGGCCTTCCTGCAGGAGCTCTGATAAAGATATTAGT<br>GAGTTATTCTGTTGAAGCTCGCTATTTCATTCCCCTGATGTCAGAGGAAA<br>GCAGACCAAAGCAGACGGCCTCCAAGCGTGGAGCCGCCGTGGAGAAG<br>ACTATGATGAGTGGGACCTACGTGTGTACTTTATGTGGTAGGAGTTTTG<br>AGAAGTCGGTGGGTTTGTCACTCCACACGAACCGGATGCACCCTGAAGC<br>TTACAATAAGCTGAAGGAGGCTAAGAAACCTGTTTTGAAGAAGGCGCG<br>ATGGTCAGAGGAGGAGGTTTTCCTTCTAGCGCAGAAGGAAGCGGAGCT<br>CTCATTCATTGGAGGAATTAAGTTCATGAACATCGAGCTGCATAAAATTT<br>TTCCAGAGCGTGAGTTGGAGGGAATCAAAGGCCAAAGAAAGAACCCAA<br>CGTACAAAGCGCAAGTCGTTAGCTTATTAGCTGAGATCCGAGAGTCTAA<br>GGCCAATGATTCCTCTTCATCTTCCTCATCCTCCTCCTCTTGTGATTCGGC<br>CTCTTTGGGCATTAGCAATTGGCTTGAATTTCTGCTAGCTCTCCCGAAGA<br>CCTCCAACCAATTCCAGGAAGGTAGGCTAGATAGGCTAATCTCTGATGC | TTTGGGGT<br>AGCAATTG<br>ATCGATTC<br>CCGCCTCC<br>TCGTGGCG<br>CTACCCCG<br>GGATAGCC<br>TCAAAGAA<br>ATTTGACG<br>GTAAAGCA<br>AAGAGGA<br>ATTGATCA<br>CCCAAGGC<br>AGTACATC<br>GGCCTTCC | TAGTGC<br>TTGAGT<br>GATGCC<br>TATCCTT<br>TCTTTGA<br>TTAACTC<br>TTACCAT<br>ATACTTA<br>CCAGTTC<br>TTACCCG<br>TACTTAC<br>CCTGTAT<br>ACTTACC<br>TGTGTG<br>CGTACCT | MSEESRPKQTASKRG<br>AAVEKTMMSGTYVC<br>TLCGRSFEKSVGLSLH<br>TNRMHPEAYNKLKE<br>AKKPVLKKARWSEEE<br>VFLLAQKEAELSFIGG<br>IKFMNIELHKIFPEREL<br>EGIKGQRKNPTYKAQ<br>VVSLLAEIRESKANDS<br>SSSSSSSSSCDSASLGI<br>SNWLEFLLALPKTSN<br>QFQEGRLDRLISDAL<br>RGVDVLENLDAYLLE<br>VFAKPMAQNPCPKP<br>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 |
| | | GCTTTTTTGCCTGTTGATGGCTGTGGGGAGAAATATTTTCCTTTTGGATTA | | TGTAATT | AVLDVKKAFDSVSHH |
| | | TATTCTTCGTAGCTCTAAGAAAAAGTAGTAAATCAGTGGCGATGGCAGTT | | 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<br>TCGACCTAATAAGGAAATGTGCTGTAGAGCGAAATGTTTTGATTCCCAG<br>AGAATGCCAGCGTTTGAATCCTTGAACCACATAACGCAAGTGTGTCCGC<br>GCACTCATGGCTCGCGAATCCAGCGGCATGATAAGATAGCCAAGTTCCT<br>TTTTAAGAACCTGAACAATTGCCCATCTCGCTCGGTTCTGTATGAACCCC<br>ATTTTGTGACGGTCGATGGCCTGCGGAAGCCTGACATTATCATTTACGA<br>TGATAGCCACATGGTGGTTCTCGATGTCCAGGTGGTTTCCGACTCTGCTA<br>ACCTGGAGAAGGAATTTGAGTGCAAGGCGAAGAAATATGCCAATGATG<br>TTGCACTTCGATCTGCAATGCTAATTAAATACCCATTCATTAAGTCATTTT<br>CTTTCGTTGCAGCTACCTACACAACAACCGTGGTCTGATAGCTAAATCCTCT<br>GTCCAGGTCCTTCGCCAGCTGGGTCTCTCGCCGCGCTCCATAATGGTCTC<br>CATCCTCATCTGCCTTGAGGGTACTTTGGAAACATGGCGCATCTTCAATC<br>AGTCTACAATGAATGCCCATTAGTGCTTGAGTGATGCCTATCCTTTCTTT<br>GATTAACTCTTACCATATACTTACCAGTTCTTACCCGTACTTACCCTGTAT<br>ACTTACCTGTGTGCGTACCTGTGTACTTGTCCTTTAGCCGCCTTGTGTTTT<br>TACCATTGGTACTTACCTTGTGTGGTTGCCCGATACTTACCTTGTATTTGC<br>CTTGTAATTCTGCATGATATTTATTGTGTAGGTTCCTGATGCTTACCTGAT<br>TTGTCCCCCTCATCATCTTTAGTTTCGTTCTATTTCACTCCATTATGGAGTT<br>CCGTTTGTTTTTTGGTGGAGGTACAGCACCCTTTAAGCTGGAATTGAGT<br>GAGTTTATGTACTTTGGATGGTTGTAATAAACTACCCGGAGGCAT (SEQ ID NO: 1110) | TTTTGGGG<br>TAGCAATT<br>GATCGATT<br>CCCGCCTC<br>CTCGTGGC<br>GCTACCCT<br>GGGATAAC<br>CTCAAAGA<br>AATTTGAC<br>GGTAAAGC<br>TAAGAGG<br>AATTGATC<br>ACACCGTG<br>ACGAATAT<br>CATCAGTC<br>ACAGCTGC<br>ACGAATCC | GGATGG<br>TTGTAAT<br>AAACTA<br>CCCGGA<br>GGCAT<br>(SEQ ID NO: 1356) | LRINAIPTRVRVSRGR<br>PNKEMCCRAKCFDS<br>QRMPAFESLNHITQV<br>CPRTHGSRIQRHDKI<br>AKFLFKNLNNCPSRS<br>VLYEPHFVTVDGLRK<br>PDIIIYDDSHMVVLDV<br>QVVSDSANLEKEFEC<br>KAKKYANDVALRSA<br>MLIKYPFIKSFSFVAA<br>TYNNRGLIAKSSVQV<br>LRQLGLSPRSIMVSILI<br>CLEGTLETWRIFNQS<br>TMNAH (SEQ ID NO: 1478) |
| R2 | R2LcB | JN937619 | Lepidurus couesii | MSGKSSKPRTVSSGS<br>SSQETPPSGSNACDI<br>CGKCFMKPVGLSLH<br>MSKVHPTQYHARLE<br>KNQPKAKKFRWTDE<br>DLYFLAKKEAELLLLG<br>GIKFMNKELAEFFPE<br>KSVDQIKGQRRSETY<br>KQQVVSIHSELLKLQ<br>AVADSPPPSRIPAKEV<br>SAWLDFLLALPKTKN<br>KFSEDKLDQLIRTAQE<br>GTPVLNDLDLYLREVL<br>VQPTRQGERQAKPL<br>PPPKSSREKRDREYA<br>RVQNFYRKNKTACV<br>NAILDGNKKCENKIP | | | |

| | | | | | |
|---|---|---|---|---|---|
| | | 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 | RVSEGVPINPRQKAF |
| | | CTATCACGATTTCTTCCATAATTGTTCGGCAGCTGAATAAAATCATCGCT | | ACTTACC | RQIDGCAENVFLLDFI |
| | | GCTCGCGTCTCAGAGGGCGTTCCCATAAATCCTCGTCAGAAGGCCTTTC | | TTATTAC | LRDAKTKIKSLSLATV |
| | | GGCAAATAGATGGATGTGCTGAGAATGTTTTCCTTTTAGATTTCATTCTT | | CCACTTG | DIKKAFDSVSHHSIFR |
| | | CGTGACGCAAAAACGAAAATTAAGTCCCTTTCGTTGGCTACTGTCGACA | | GTCTTTT | AIRGARCPENLVNYI |
| | | TTAAGAAAGCATTTGACAGTGTGTCACATCATTCCATTTTCCGAGCGATT | | ATTTTCT | QNSYSGCTTQJSVGG |
| | | 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<br>AAGTTTGTCCTCGGACGCATGGTGCAAGAATAAAAAGACACGATCGGCT<br>CGTTCAATTCCTTTCTCTCAACCTTAGGAAAAATCCAAAAAGAAATGTTT<br>TAGTGGAGTACAATTTCCGTACAGTCGCTGGTATCCGTAAACCGGATAT<br>CATTGTTATTGAAGATACCCGTGCAGCTATCCTTGACGTCCAAGTTGTTG<br>GTGATTCGTCCAATTTGGAGATGGAATACCTGGAGAAATCCAGGAAATA<br>TTCTAATGACGCTACTTTATCAATGCGTATCAATGCGCTTCAAAAACTTT<br>ATCCAACGGTCACTAGCCTAACTTTTCATGCAGTTACCTTCAACAATAGA<br>GGCCTGATTGCCAAATCCACAGTGGCGGCCCTCCGGATGCTCGGCGTCC<br>CACCTAGATGCATAATGATCCTTTGTGTGATTAGCCTTGAGAAGACCCTG<br>GAAGTCTGGAGGATGTTTAACCAATCAACTGCTTCCGCTCGGGAAATGAT<br>AATCGCTCCATCCTGCAACTAATTATGAATGCAAATCTGTTAAGTGACAT<br>TAGTGATACTTACCTGATACTTACCCTGGTATTTATTTGACCTATACTTAC<br>CCTGGTATCTACCTGACATATATTTATCTAACCACCTACCTATGATGACTC<br>CCGCGGAAACTCTCACTTACCTTATTACCCACTTGGTCTTTTATTTTCTCG<br>TTCCTTATTACTTTGTTCCTTTGGTGTAGGGTTCTCTGGTTTTTGGAACGG<br>CTTCCTTAGCCGGAATTTTTGTCTGATGTATCTTGCTTGTGTCCTTGAAATA<br>TACGACCCAGGCTTGCGTCATTTAGGCTCTGGGAAA (SEQ ID NO: 1111) | GGTTGGGG<br>GCCTTCTC<br>GTGGCGG<br>AGTCGTGA<br>GTAAGGG<br>GTATAGGG<br>GTAAGGG<br>ACACCACG<br>GACCGAG<br>AACGGTTA<br>CCGCTCAA<br>GGCGAGT<br>GGTGGAA<br>GGCATAAA<br>ATCGTAAC<br>GCCGCCCT<br>CCGACCTG<br>CTCCTGAA<br>ACTAATGC | TGATGG<br>TGGGTT<br>ACTCGC<br>CTCTGT<br>GTAACA<br>GGCAAA<br>TGAAAG<br>CTGCGC<br>AAGCAG<br>TCGATG<br>AGCCAA<br>AGCCGC<br>ACAGCC<br>CCCGAC<br>TGGGGT<br>ACAGGG<br>CAGCCC<br>TGGGCT<br>ATGCCC | TINHIAQVCPRTHGA<br>RIKRHDRLVQFLSLNL<br>RKNPKRNVLVEYNFR<br>TVAGIRKPDIIVIEDTR<br>AAILDVQVVGDSSNL<br>EMEYLEKSRKYSNDA<br>TLSMRINALQKLYPT<br>VTSLTFHAVTFNNRG<br>LIAKSTVAALRMLGV<br>PPRCIMILCVISLEKTL<br>EVWRMFNQSTASAR<br>K (SEQ ID NO: 1479) |
| R2 | R2Nvec-A | Nematostella vectensis | MLRGTGNMNDKRD<br>GSATADPTSALLGAV<br>GDGSLVCNLCGLACK<br>SRGGLSIHRRSKHAT<br>VYHAERQPAPRAKA<br>RWTNDEMILVARKQ<br>IASEKSRCSAVVEGM<br>REAVPHRTFDAVKSL<br>KTKNRNYTRILEQIRA<br>ECSEEEVIESGVLKDR<br>TENVCVQTTSNVPGS<br>AGRAASVELEGNIQV<br>GHQLAQKTMAGNN<br>SRKQPANHTNWAEF<br>NIEEGNITLRKSKRKA<br>NGMPDATHRPGPPT<br>VDSLKHPVCLLQGAA<br>DKRDEPHTVEQLYYN<br>IEEGMPLAEEQQWS | GGTTGGGGCCTTCTCGTGGCGGAGTCGTGAGTAAGGGGTATAGGGGTA<br>AGGGACACCACGGACCGAACGGTTACCGCTCAAGGCGAGTGGTGG<br>AAGGCATAAAATCGTAACGCCGCCCTCCGACCTGCTCCTGAAACTAATG<br>CCAACCAACTGACTGTGGGGCTAACCTCCCCAGAGTCAGGATGCTGAG<br>GGGTACTGGGAATATGAATGACAAGAGGGACGGATCGGCCACTGCTGA<br>TCCTACCTCAGCGCTTTTGGGGGCTGTCGGTGATGGCAGCCTGGTATGC<br>AATCTGTGTGGCCTTGCTTGCAAGTCTCGTGGGGGGCTCTCGATTCATC<br>GAAGGAGCAAGCATGCCACGGTGTACCATGCGGAAAGGCAGCCCGCTC<br>CGCGGGCTAAGGCGAGATGGACGAATGATGAAATGATACTTGTGGCCA<br>GAAAACAAATAGCTTCGGAAAAGAGTAGGTGTAGCGCAGTAGTTGAAG<br>GGATGCGAGAAGCAGTTCCACATAGAACTTTTGACGCCGTCAAAAGTCT<br>CAAAACCAAGAATAGGAACTACACACGGATTCTGGAGCAGATTCGCGCT<br>GAATGCTCCGAGGAAGAGGTGATTGAATCAGGAGTCCTTAAGGATAGG<br>ACGGAGAACGTGTGTGTGCAAACAACATCTAACGTTCCCGGTTCGGCAG<br>GACGGGCAGCATCGGTGGAGCTAGAGGGTAATATCCAGGTGGGCCATC<br>AACTCGCCCAAAAAACCATGGCAGGCAATAATAGTAGGAAACAGCCTG<br>CCAATCATACCAACTGGGCTGAGTTCAATATTGAGGAGGGAAATATTAC<br>CCTCCGAAAAAGCAAAAGAAAGGCTAATGGTATGCCGATGCGACCCA<br>CCGCCCCGGCCCACCGACTGTCGACAGTCTTAAGCACCCAGTATGCTTG | | | |

| | | | | | |
|---|---|---|---|---|---|
| | TTGCAGGGTGCAGCCGACAAAAGGGACGAACCGCACACAGTGGAACAA | CAACCAAC | GAAGTC | | EKLFDAIDSSLLSVEVE |
| | TTGTATTATAACATTGAGGAAGGAATGCCTCTTGCCGAAGAGCAACAGT | TGACTGTG | TTTTGAC | | LGRIVPGCPDEETRQ |
| | GGTCAGAAAAGCTTTTCGATGCAATCGACTCATCTTTGTTGTCTGTCGAG | GGGCTAAC | AGGTCC | | LIDREFLDFIHSYSREK |
| | GTCGAGCTGGGGGAGAATCGTTCCTGGTTGCCCCGACGAAGAGACCCGA | CTCCCCAG | AAACTTC | | PPQRGLAKSKPPPKG |
| | CAGCTCATTGATAGAGAATTCTTGGATTTTATCCATTCCTATTCGAGGGA | AGTCAGG | ACCCTTG | | PKSLRRQQYRQLQRL |
| | AAAGCCGCCACAAAGAGGTCTCGCTAAGTCCAAACCACCACCCAAAGG | (SEQ ID | CCGCCA | | WDKNRSAAAEQALT |
| | ACCAAAGAGCCTCAGACGCCAGCAATATCGGCAACTCCAGAGATTGTG | NO: 1235) | GTAGGG | | GKWQEVRTAAGVPL |
| | GGACAAAAACCGATCCGCAGCAGCCGAACAGGCCCTAACAGGTAAATG | | CACCAG | | SLMEVPWREIFETPS |
| | GCAAGAAGTTCGGACTGCTGCTGGGTCCCATTATCGTTGATGGAGGTC | | GGCCAG | | TTDVREPAPAGPVL |
| | CCGTGGCGGGAGATTTTCGAAACACCATCTACGACGGACGTCCGAGAG | | GTGCAG | | WQLLRPVTIAEVEDA |
| | CCGGCCCCAGCTGGGCCGGTTCTGTGGCAGCTTCTCAGACCAGTGACTA | | GTGGGC | | ISSKKSASGPDGVPC |
| | TCGCAGAAGTAGAGGACGCCAATTTCCTCCAAGAAATCAGCTAGCGGTCC | | 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 |
| | ACGTACATCAATGAACTGTACTCTCGTTCGGAGACAGTCTTTGAAGTTG | | TTTTAAC | | FEVGGESSGSVKVSQ |
| | GGGGCGAGAGCAGTGGCTCTGTTAAAGTATCCCAAGGGGTAAAGCAAG | | ACCAATT | | GVKQGDPLSSTLFNC |
| | GGGATCCATTGAGTTCGACGCTCTTCAACTGCGTCATTGACTGGGCGGT | | TTTTTTA | | VIDWAVSDLDPHIGV |
| | CTCAGATCTCGACCCCCACATAGGCGTCTTACTAGGGGAGTCTCGAGTC | | CAACCCT | | LLGESRVSFLAYADDL |
| | TCCTTTCTAGCTTACGCTGATGACCTGGTTCTCCTATCGGAGACCGAGGC | | CTACTCA | | VLLSETEAALTSQLNS |
| | TGCACTGACAAGCCAGCTGAACTCAATTGAAAAGAGTCTAGCGCACTGT | | ATAATCC | | IEKSLAHCGLKLSTGD |
| | GGTCTCAAGCTCAGTACGGGGGGACAGTGGTAAATCTGCTTCGCTGAACA | | AAATGA | | SGKSASLNIVIDGKAK |
| | TCGTGATTGATGGTAAAGCTAAGCGGTGGGTTGTAAACCCTACCCCGTT | | ATAAAA | | RWVVNPTPFLRASG |
| | CCTACGTGCGTCTGGTGGCGAAATTCGATCTCTCGTCGCAAATGAGACG | | GCGGCA | | GEIRSLVANETYKYLG |
| | TACAAGTACTTGGGCATTAACATCGGAGCCCAGGGCGTGAAAGCTGCT | | TAACAG | | INIGAQGVKAAEYNA |
| | GAATACAATGCATTCAAAGAAGCACTCGATAACTTGTCTCGAGCTCCACT | | GTGAAC | | FKEALDNLSRAPLKP |
| | TAAGCCGCAGCAGAGGTTATTCCTGCTGAAGACCTATCTGTTGCCGCAA | | AC (SEQ | | QQRLFLLKTYLLPQLH |
| | CTACACCACTCACTGGTATTGTCAAGGACAACTGGCAAACTATTGAACA | | ID NO: | | HSLVLSRTTGKLLNSL |
| | GCCTCGATGCTCTCGTAAGAAAAAGCTGTACGGGGATGGCTTAAATTGCC | | 1358) | | DALVRKAVRGWLKL |
| | ACACGACACTCATAGGGCCTTCTTTTATGCTCACCAGGCCGACGGGGGA | | | | PHDTHRAFFYAHQA |
| | CTCAATGTCCCAAGTCTGTACCATCTAATACCTCTCCTAAGGAGGAGTAG | | | | DGGLNVPSLYHLIPLL |
| | GTATGAACGCTTGACGAGGGTCGAAGACCCTGAGATAAGGGAAGTGTC | | | | RRSRYERLTRVEDPEI |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CCGAACCGACTACTTCAAAGAGTCTTGGGTGCCGCGGCGGCGGCCAC | | | | REVSRTDYFKRVLGA |
| | | | CACGGTTGCAGGGCATCGAATTGATTCCAAAGTAACGTTGCGCCTAGCT | | | | AAAATTVAGHRIDSK |
| | | | TGGCGCGAGGCCCTTTACGCTTCAGCTGACGGGCGGGGTCTCTCCCAAT | | | | VTLRLAWREALYASA |
| | | | GCCCGTTGGTACCCGAGGTTCACTCCTGGGTTACCGATGTTTCCGGTCTT | | | | DGRGLSQCPLVPEVH |
| | | | CAGACCGGATCGCAGTACATCTCTGCTGTTCGCTTAAGGGGGGGCCTTGC | | | | SWVTDVSGLQTGSQ |
| | | | TTCCCACGGCGGTAAGGAAATCGAGGGGTCGCGGGGGGTGTTAACTCGA | | | | YISAVRLRGALLPTAV |
| | | | ATTGCGACTGTTGCGGTCGTGGACAACCAGAGTTTCTTGGGCACGTTCT | | | | RKSRGRGGVNSNCD |
| | | | CCAGACGTGTCCTCGGACCTGGGGGGTCTCGCATCAGTCGTCACAACCAC | | | | CCGRGQPEFLGHVL |
| | | | GTCCTCTCCCTTATCGCTAAAGCGTGCCGTTCGCGAAGATGGCAGGTGC | | | | QTCPRTWGSRISRHN |
| | | | TGGAAGAGCCGATCATCCAGACTCCTGCGCAGTTGCTCAAGCCAGACTT | | | | HVLSLIAKACRSRRW |
| | | | GGTGATCTGGAACCATCAAGCGGCGTACGTCGTTGACGTTAGCGTACCG | | | | QVLEEPIIQTPAQLLK |
| | | | GGTGATAACACCCCGCTCAGTACGTGTCACAACAGGAAAGTCGCATATT | | | | PDLVIWNHQAAYVV |
| | | | ACTCTGGTGAGAGCGTTCGTGAATGGGTGCGTTCCAAGACTGGGCATA | | | | DVSVPGDNTPLSTCH |
| | | | ATCCTACAGTGTCCTCTGTAGTGATTAACTGGCGCGGTGCCATGGCAAA | | | | NRKVAYYSGESVRE |
| | | | GGAGTCCTACCGCCTTTGACGAAAGACCTTCGTCTAGCGAAGACCCTG | | | | 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 |
| | | | GAAGGCTCAGAAACTACCCACTTTTTGCTGTTTATCCACACAACAGTTG | 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 |
| | | | ATGTCTTGTGTTCTTTGCGCCGCCGCTTTTCATCTAACAGGGGACTAAC | CAACAACA | TCTGGG | | TTSSNNHTTTDIPNTE |
| | | | ACAACACATTCGCCACCGGCACATCTCCGAATATAACGAACTAATCAGA | 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 | |
| | ACACATCCAGACACTCTTCCACACCGTCCCAAAGATGCCCGCCAACACCG | CGGTCCGT | TCTACCT | AGGIDRLTSYDLMRF | |
| | TTCTAGACGGTCGGTGGAGAAAACCCCTATGTCGCAAACCATTCAATGAT | 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<br>CAACTACCTCACGGAAATATCGTGCGCTCCACTTAAGCCGCAGCAGCGC<br>ATGAAGATCTTGAAAGATAATCTACTCCCTCGACTCCCTCTACCCCCTGAC<br>TCTAGGAATAGTACACCTGAAAACCCTGAAGTCAATGGACCGAAATATC<br>CACACGGCCATAAGGAAATGGTTGCGGCTACCCTCCGACACCCCGCTAG<br>CATATTTTCACTCACCCGTCGCTGCCGGAGGCCTAGGGATCCTCCATCTG<br>TCCTCATCGGTTCCATTCCACCGTCGAAAACGTCTAGAAACCCTCCTATC<br>TTCACCGAACCGCCTACTGCACAAGTTGCCAACTTCCCCAACACTAGCTT<br>CTTATTCACACCTTAGTCAACTGCCAGTTCGAATTGGGCACGAGACCGTA<br>ACGTCTAGAGAAGAGGCTTCCAACAGCTGGGTGAGACGATTACATTCGT<br>CCTGCGACGGGAAGGGACTACTCCTAGCACCACTAAGCACCGAGTCCCA<br>TGCATGGCTGCGCTACCCCCAGTCTATTTTTCCAAGTGTTTACATCAACG<br>CCGTTAAATTACGAGGTGGCTTACTATCCACCAAAGTCAGGAGATCTCG<br>CGGAGGTAGAGTGACGAATGGCCTGAACTGTCGAGGCGGTTGCGCCCA<br>TCATGAAACAATCCACCACATTCTGCAACATTGCGCGCTCACCCATGACA<br>TCAGATGCAAACGCCATAACGAACTATGCAACCTTGTGGCAAAGAAACT<br>GCGTAGGCAAAAAAATCCATTTCTTACAGGAGCCCTGCATTCCTCTAGAA<br>AAAACTTACTGCAAACCTGATTTTATAATTATACGTGATTCAATTGCTTAT<br>GTTCTAGACGTCACTGTATCGGACGACGGAAACACCCACGCCAGCCGCC<br>TGTTAAAAATATCAAAATACGGCAATGAGCGAACCGTCGCATCGATCAA<br>GCGATTCCTCACATCCAGTGGATATATCATTACCAGTGTTCGACAAACAC<br>CAGTCCTTACATTCAGAGGTATTCTGGAGAGAGCAAGTTCACAATCCCT<br>ACGACGCCTATGTTTTTCGTCCCGTGACCTCGGTGACCTTTGCCTGAGTG<br>CGATTCAAGGCTCAATTAAAATATATAATACCTATATGAGAGGAACCCA<br>ACGGCTGAACGAATAGCCCCCTTCACTCTTAGACATTCCCCCACTGTTGT<br>TGCTTATCTTCATGTTTTTGTGTTAATTGACTGCTCTCTTCTGGGTTGATG<br>TCTGATTGTCTCTCTCTTTCCATATTGCTTGCTCTCCCCGCTTACTTCCA<br>ATAGTTGTCATATTATGTCTTTGTTTACTTGCCATGTCTAACGACAATTAC<br>TTTATCTACCTTAGTTGGTCCTCTTGGTTTGGTTGCCTTCATGTGTTCATG<br>GCGGAATCTGATGTTTATAATGACTATTCCTACTACCACCATTACAACTA<br>TTATTATTATCACTATTATTAACATTATTATTACTTCTACAATTAGTATTAT<br>GGCTACTCCTTTCAGCACACCAATAAAAATCTCAATCAAACATCTCACTTAT<br>TAAACTCTCTATTTCCCCTTCGTTATAAACTTACAATTCAGTTTAACCGAA<br>TATCTCTCTTTTACAAATCTTAAGTATGTAATTTTGTGCCAAGCCCATTTG<br>GGTCTGTACAATTTGATACTTAAAAATAAATGTTAT (SEQ ID NO: 1113) | TTTTTGGT<br>CTGGCATT<br>TGATCGTT | ACTCCTT<br>TCAGCA<br>CACCAA<br>TAAAAT<br>CTCAATC<br>AAACAT<br>CTCACTT<br>ATTAAA<br>CTCTCTA<br>TTTCCCC<br>TTCGTTA<br>TAAACTT<br>ACAATTC<br>AGTTTA<br>ACCGAA<br>TATCTCT<br>CTTTTAC<br>AAATCTT<br>AAGTAT<br>GTAATTT<br>TGTGCC<br>AAGCCC<br>ATTTGG<br>GTCTGT<br>ACAATTT<br>GATACTT<br>AAAAAT<br>AAATGT<br>TAT (SEQ ID NO: 1359) | MDRNIHTAIRKWLRL<br>PSDTPLAYFHSPVAA<br>GGLGILHLSSSVPFHR<br>RKRLETLLSSPNRLLH<br>KLPTSPTLASYSHLSQ<br>LPVRIGHETVTSREEA<br>SNSWVRRLHSSCDG<br>KGLLLAPLSTESHAW<br>LRYPQSIFPSVYINAV<br>KLRGGLLSTKVRRSR<br>GGRVTNGLNCRGGC<br>AHHETIHHILQHCALT<br>HDIRCKRHNELCNLV<br>AKKLRRQKIHFLQEPC<br>IPLEKTYCKPDFIIIRDS<br>IAYVLDVTVSDDGNT<br>HASRLLKISKYGNERT<br>VASIKRFLTSSGYIITS<br>VRQTPVLTFRGILERA<br>SSQSLRRLCFSSRDLG<br>DLCLSAIQGSIKIYNTY<br>MRGTQRLNE (SEQ ID NO: 1481) |
| R2 | R2Tc | EU854578 | Triops cancriformis | TTTTTGGTCTGGCATTTGATCGTTTCCGCCTCCTCGTGGCGCCAGACTGG<br>GTAAGCTGATTTATCAGTGAGCTAAGAAGAAACGATCACCGCAGGAGTCC<br>ATCTACCTGCGCTGCACGTGATTTCATTGTGCACTTGGCGAGTTATCCCT | TAGATG<br>ACTGCC<br>CTACCCC | MSQKRRPEKAVPDE<br>GATAHDVAQPDKSK<br>CSVCGETFKGPASVT | | |

| | | | | | |
|---|---|---|---|---|---|
| | TGTGGAGCTCGCTCGTTAGCTTTTGAGATGTCTCAAAAGCGCAGACCGG | 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 (SEQ ID NO: 1360) | | QDIIVRTIAESETLGVE |
| | | | GACATTATTGTCCGCACTATTGCCGAATCGGAGACGTTGGGGGTCGAGA | | | | IRKLHDLLSIDGTILDT |
| | | | TTCGAAAGCTTCACGATCTGTTGTCCATTGATGGAACAATTCTGGACACG | | | | SVKIHSFWAERLYSSY |
| | | | AGCGTTAAGATCCATAGCTTCTGGGCAGAAAGACTCTACTCGTCCTATG | | | | DGKCLCNSANFPPG |
| | | | ATGGTAAATGTCTTTGCAATTCTGCAAATTTTCCCCCCCGGAAACAAGTGG | | | | NKWIGEDSLNQRSHI |
| | | | ATCGGTGAAGATTCTCTGAACCAACGCTCGCATATCTTCGCTGATTGCCT | | | | FADCLKLRINALPTRS |
| | | | AAAGCTCCGGATAAATGCGCTCCCAACACGTTCCAGAACAGCACGTGGA | | | | RTARGRPLKDKPCRA |
| | | | AGGCCACTAAAAGATAAACCATGTAGAGCTGGATGTCGGAACAGTGAT | | | | GCRNSDGVKVIETLN |
| | | | GGTGTAAAGGTCATTGAGACACTGAATCACATTACACAGGTGTGTGAG | | | | HITQVCERTHGARVK |
| | | | AGAACCCACGGAGCCCGCGTCAAGCGGCATGATCGCCTAGTGGACTTC | | | | RHDRLVDFAVKGLQ |
| | | | GCGGTGAAAGGACTACAAAGACCCCATCGTGTGGTTCTAAAGGAGCCA | | | | RPHRVVLKEPHYKTV |
| | | | CATTACAAGACAGTAAACGGTGTTCGGAAACCGGACATCGTCATTAAAA | | | | NGVRKPDIVIKIPDHT |
| | | | TCCCCGACCACACATACATTTGTGATTTCCAAGTTGTATCTGATACCTCTT | | | | YICDFQVVSDTSCLEL |
| | | | GTTTAGAGTTGGAGTTCCGGAAGAAAGCCCTCAAGTATGCGGAGGACA | | | | EFRKKALKYAEDKGL |
| | | | AAGGCCTCTGCGATCAACTTACCCGCGACCATCCTGGGGAGCTCTCATT | | | | CDQLTRDHPGELSFT |
| | | | CACCGCAATTACTTTCAATACAAGAGGGCTTATAGCCAAGTCTTCAGTCA | | | | AITFNTRGLIAKSSVT |
| | | | CCGCCCTCCGGAAATTGGGGATGCCACCCAGATCTATCATGACCCTCCA | | | | ALRKLGMPPRSIMTL |
| | | | GAAAATATGTATGGAGGGTTCTCTGGAGATCTGGAGGATTTTTAACCAG | | | | QKICMEGSLEIWRIF |
| | | | ACTACTGCTATGGCCCGCAATTAGATGACTGCCCTACCCCTTTGCTGCCG | | | | NQTTAMARN (SEQ ID NO: 1482) |
| | | | AAGAACTACTGAAGACTATTGAACTAACCAACTGTGTTAAGTAAGAACT | | | | |
| | | | AATGCCTCTTTTTCCCTGCATGTATCCCCTGATCAGTGACTTATTTTCTTTT | | | | |
| | | | CCTGTTGCGCCCTTTGTTTAGTTATTTCCTTTAATTACTAGAATTATCTTTT | | | | |
| | | | CGTTCTCCGTCTAATTGCTTTTTCGGTGTAGGAACGGCTACCTAAAGCTG | | | | |
| | | | GAAGTGGGAAGTGTTTTCAATGTACTTTGTGATTATAGAAATATATGAC | | | | |
| | | | CCGAGGTGCATTGTTTGGCATTTCCTCGAGAAA (SEQ ID NO: 1114) | | AGAATATG | TAGCCA | FERRKDPWGYRPPG |
| R2 | R2_DAn | . | Drosophila ananassae | AGAATATGGATTTGATTGTGCAGAGGGGGTGCTATACCGTAACTCGTAA | GATTTGAT | ATGCAC | TLKQIGATENNEPRN |
| | | | | GCCATGCAATCAGATCAAGTCGACTCAAAACCTCCTCGTGGTATTCTCTG | TGTGCAGA | GGGTTC | LNRFVRGESTASSLES |
| | | | | GGTGCCAGTATTTACTGGTAGCTGATTTGAGCGGCGAAAGGATCCTTGG | GGGGGTG | CAGATT | TQFGTSAEVNLAGRV |
| | | | | GGTTACCGGCCCCCTGGAACCTTAAAACAAATTGGTGCAACTGAAAATA | CTATACCG | AAGCTT | PCTICEMTFSSKRGLG |
| | | | | ACGAGCCTCGGAACCTAAATCGTTTTGTAAGAGGAGAATCCACGGCTTC | TAACTCGT | GCTGCC | VHMSHRHKDDLDA |
| | | | | CAGCCTGGAGAGCACACACAATTTGGAACCAGTGCAGAGGTTAACCTTGCA | | | |

| | | | | | |
|---|---|---|---|---|---|
| | | 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 |
| | | GATTTAGTTTTGAGGGAACATCATCATAAGCGATTTAAATCGTGCTACATCG | | CATAAA | TTIVDLVLREHHKRFK |
| | | GGACCCTCGATGTTAGTAAGGCCTTTGATGCTGTAGCTCACGAAGCGGT | | TATCCCG | SCYIGTLDVSKAFDAV |
| | | CTACAACACATTGGCTTCATATGGTGCCCCGAAAGGCTTCATCAACTACT | | TCCTTAC | AHEAVYNTLASYGAP |
| | | TACGGAAGGCGTACGAGGGCGCGGCACACAATGCTCGCTGGGAACGGG | | GTATCTT | KGFINYLRKAYEGGG |
| | | TGGGTTTCAGAGGCGTTCATTCCTGCCCGAGGAGTGAAGCAGGGTGAC | | TGAAGA | TMLAGNGWVSEAFI |
| | | CCTCTGTCTCCCATACTATTCAACTTGGTCATTGACCGGTTGCTTAGGTCC | | TTTCCAT | PARGVKQGDPLSPIL |
| | | TTACCCAGTGAGATTGGTGCCAAAGTCGGAAATGCCATGACAAACGCA | | CCTCAG | FNLVIDRLLRSLPSEIG |
| | | GCAGCATTCGCAGATGATATAGTCCTTTTGCGGAAACTCCGATGGGGC | | 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 | TTGGGGATCATGGGGTATTTGAGAGCAGAGGGGGAGTATTCTTCTGTA | 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 |
| | GGCCGCGTTTGCAGATGATGATTTGGTACTATTTGCGGAAACTCGGATGGGG | | AAAAAA | | GQPKQKCTVLEAQSF |
| | CTTCAAGTATTGTTGGACAAGACGTTGGATTTTCTATCTATCGTCGGCCT | | AAAAAA | | YVGSSEIPSLKRTDEW |
| | CAAACTTAATGCCGACAAATGTTTTACCGTTGGCATTAAGGGCCAGCCG | | (SEQ ID | | KYLGINFTATGRVRC |
| | AAACAGAAGTGTACCGTGTTAGAGGCACAGAGCTTCTACGTAGGCTCG | | NO: | | NPAEDIGPKLQRLTK |
| | AGTGAGATTCCATCACTGAAGCGCACGGACGAGTGGAAGTACTTAGGC | | 1362) | | APLKPQQRLFALRTV |
| | ATCAACTTCACTGCAACCGGGAGGGTTCGATGCAATCCGGCCGAGGAC | | | | LIPQLYHKLALGSVAI |
| | ATTGGTCCAAAGCTACAAAGATTGACAAAGGCCCCCCTCAAACCACAAC | | | | GVLRKTDKLIRYYVRR |
| | AGAGGTTGTTCGCCCTTCGGACTGTCCTTATCCCACAGCTCTACCACAAG | | | | WLNLPLDVPIAFVHA |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | TTAGCCCTTGGGAGTGTGGCGATAGGCGTCCTAAGAAAAACTGATAAA<br>CTTATAAGATATTATGTGCGAAGATGGCTAAATCTTCCGCTGGATGTGC<br>CGATAGCATTTGTTCATGCACCCCCAAAAAGTGGAGGTCTCGGAATTCC<br>ATCACTAAGATGGGTAGCTCCAATGTTAAGGCTAAGACGCTTGAGTAAC<br>ATTAAATGGCCTCACCTCACGCAAAACGAGGTAGCCAGCTCTTTCCTCG<br>AAGCCGAAAAACAACGGGCCCGAGATAGATTATTAGCTGAACAAAATG<br>AACTGTTATCGCGTCCGGCAATAGAAAAATATTGGGCGAACAAGTTGTA<br>CCTCTCAGTTGATGGTAGCGGACTCCGTGAAGGCGGCCATTATGGCCCG<br>CAACACGGGTGGGTTAGTCAACCCACGCGTTTATTAACAGGAAAGGAAT<br>ATATGGACGGTATTCGTCTCGCGGATAAATGCCCTACCCACAAAGTCTCG<br>TACTACAAGGGGAAGGCACGAATTGGAACGACAGTGTCGTGCAGGATG<br>TGATGCTCCCGAAACAACAAACCACATAATGCAAAAATGCTACCGCTCG<br>CATGGGAGGCGGGTAGCTAGACACAACTGCGTAGTAAATCGAATCAAG<br>CGGGGACTTGAGGAGAGAGGCTGCGTGGTCATTGTTGAACCAAGTCTG<br>CAGTGCGAATCCGGCCTTAATAAACCAGACCTGGTGGCACTACGACAAA<br>ATCACATTGATGTGATCGACACACAAATTGTGACAGACGGACACTCTAT<br>GGATGATGCGCACCAGCGCAAAATCAATAGATACGACAGACCGGACAT<br>ACGAACTGAATTGCGTCGCAGATTCGAAGCCGCAGGTGACATTGAATTC<br>CATTCTGCCACCCTGAACTGGAGGGGGATCTGGAGTGGTCAATCCGTTA<br>AAAGATTGATAGCAAAGGGTCTCCTCAGCAAATATGATAGTCATATCAT<br>TAGCGTCCAGGTTATGAGAGGCAGTCTCGGTTGTTTTAAACAGTTCATG<br>TACCTGAGCGGGTTTTCCCGAGATTGGACTTAGCTAAATCGTTTGGTTCA<br>AAACATTTGCTTGCTGTCTTGGCATAACATCAATAAAGGCATAAACATCG<br>CAAAATAATGGTTATAATTAAATGGCTATGAGGATGGTTTTAGTACGTA<br>GGCGTTGCGGAACTTCGGTTCATATAGAGCAATGAATCGTGCATGCTAG<br>GAAAACTGACCACACACAGTGTTGGCAGACCTAGTATCTTTCGAAGATT<br>TCCATACCTCCGCGATCAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1116) | | | | PPKSGGLGIPSLRWV<br>APMLRLRRLSNIKWP<br>HLTQNEVASSFLEAE<br>KQRARDRLLAEQNEL<br>LSRPAIEKYWANKLYL<br>SVDGSGLREGGHYG<br>PQHGWVSQPTRLLT<br>GKEYMDGIRLRINAL<br>PTKSRTTRGRHELER<br>QCRAGCDAPETTNHI<br>MQKCYRSHGRRVAR<br>HNCVVNRIKRGLEER<br>GCVVIVEPSLQCESGL<br>NKPDLVALRQNHIDV<br>IDTQIVTDGHSMDD<br>AHQRKINRYDRPDIR<br>TELRRRFEAAGDIEFH<br>SATLNWRGIWSGQS<br>VKRLIAKGLLSKYDSHI<br>ISVQVMRGSLGCFKQ<br>FMYLSGFSRDWT<br>(SEQ ID NO: 1484) |
| R2 | R2_DPe | Drosophila persimilis | AAGATATGGATCTGAATAATAGCGTAGAAGGGGAGTCATTCCGTAATTC<br>GTAAATCGTAAAAATCAGATCAAGTTGATTCAAGACCTCCTCGTGGTAT<br>CTTCTGGATGCTATTAGACTGAAGTTCTTTTGGTCTAATAGTAACTAACT<br>TGAACAGCGAAACAGTCCTATGGGGTTGCCAGCCCCTTGGACAGTTCAG<br>TTTGATTGGCACTAATATGCAAAATACAACGCCTCGGATAATAAACACTA<br>ATTCGTTGACGAACCAAATCCCTACGGTCTCTAGCCTAGGGGCCCAATCT<br>GAACATAGTGCACAGGTTAACCCAAACAGTGGTTACCAATGCACGATAT<br>GTGAATCGTCTTTCCGTAGCAAAAGCGGACTAGGCGTTCACATGTCACG<br>TCGGCACAAGGACGAGTTTGATCAACTTCGTCTGCGTACCGACCGTAAG<br>GCACAATGGAGTGAGGAAGAGTTGAGTATGATGGCAAGAAAAGAGAT | AAGATATG<br>GATCTGAA<br>TAATAGCG<br>TAGAAGG<br>GGAGTCAT<br>TCCGTAAT<br>TCGTAAAT<br>CGTAAAAA<br>TCAGATCA<br>AGTTGATT | TAGCCT<br>ATACACT<br>ATGTTG<br>GAGAGA<br>AGACGC<br>TTGCTAC<br>CTAGGC<br>AAAATG<br>TGAAAT<br>TAGGTA | SSFGLIVTNLNSETVL<br>WGCQPLGQFSLIGT<br>NMQNTTPRIINTNSL<br>TNQIPTVSSLGAQSE<br>HSAQVNPNSGYQCTI<br>CESSFRSKSGLGVHM<br>SRRHKDEFDQLRLRT<br>DRKAQWSEEELSM<br>MARKEIELAANGERY<br>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 | | | DPLSPIIFNLIIDQLLRS |
| | | ATCGGTGATCACACAACAAACGCGGCCGCGTTCGCAGATGATATTGTCT | | | YPNEIGATIGDHTTN |
| | | TATTTGCGGAAACTCGTTTAGGCCTTCAAACAATGCTAGACACGACTGTC | | | AAAFADDIVLFAETRL |
| | | GATTTTCTATCTTCAGTCGGGCTTACCCTTAACTCGGATAAAATGTTTTACA | | | GLQTMLDTTVDFLSS |
| | | GTTGGAATAAAGGGGCAACCGAAACAGAAGTGTACTGTGGTCATCCCA | | | VGLTLNSDKCFTVGIK |
| | | GAGACCTTCCGTATCGGTTCGCGCTCGTGTCCTGCATTGAAGCGCACAG | | | GQPKQKCTVVIPETF |
| | | ACGAGTGGAAGTATTTAGGCATAACATTCACTGCACAAGGGAGGACCA | | | RIGSRSCPALKRTDE |
| | | GGTACAGTCCAGCCGACGACCTAGGTCCGAAGCTGTTGAGGCTGACAA | | | WKYLGITFTAQGRTR |
| | | GGTCCCCCCTAAAACCACAACAGAAATTGTTCGCACTCAGAACACAGTTCTT | | | 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 | . | CAATTGGAAAGATATGGGTCTGAATAATAGCGTAGAAGGGGAGTCATT | Drosophila pseudoobscura | 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 GGATGCTA TTAGACTG A (SEQ ID NO: 1241) | AACTTG AGGTGG GTTTTTA GTACGT ATGCGT GATTACT TCGTAAT CATGAA TCGTGC ATGCTA GTGGGG TTTGGC CTCCACT AGTATCT TTGAAG ATTTTCC TTCCTCA GCGATC AAAAAA AAAAAA AAAAAA AAAAAA (SEQ ID NO: 1364) | TPPNTLGTTPTELSNR |
| | | CCTAGTATACGCGAGCGAGATCTCCTAGTAACGCCACCTAACACTCTCG | | | EILAVLQGYPPVVCN |
| | | GAACCACTCCAACCGAACTGTCGAACAGGGAAATCCTGGCAGTACTACA | | | DQWRVEVLQSIVDG |
| | | GGGGTACCCACCTGTAGTATGCAATGACCAATGGAGAGTTGAGGTTTTG | | | AQASGKEITLQRLSTY |
| | | CAATCCATCGTAGATGGTGCGCAGGCCTCGGGTAAGGAAATTACTCTTC | | | LMEVFPSQNDRPIQT |
| | | AGCGCTTGTCTACTTACCTTATGGAAGTATTTCCCTCACAGAATGACCGC | | | RPPRRPRNRRQGRR |
| | | CCCATTCAAACGAGACCTCCACGGAGACCTCGTAATAGGAGACAAGGTA | | | QQYALTQRNWDKH |
| | | GGAGACAGCAGTACGCCTTAACACAGCGTAACTGGGACAAGCACAAAG | | | KGRCIKAILDGTEGTA |
| | | GTCGTTGTATAAAAGCCATTTTGGATGGAACTGAGGGGACAGCAACTAT | | | TMPSQGIMGSYWR |
| | | GCCAAGTCAAGGTATCATGGGGTCCTATTGGAGACAAGTCATGACACAA | | | QVMTQTSPTYSGTN |
| | | ACAAGCCCAACATATAGTGGTACGAACACCACGTTCCGGACGGAACACC | | | TTFRTEHPLEGVWSP |
| | | CACTTGAAGGGGTTTGGTCCCCGATAACACTAGGGGACCTAAGGGTAC | | | ITLGDLRVHRVSLTKS |
| | | ACAGAGTGTCATTGACGAAATCTCCAGGACCTGATGGGATTACTCCAAG | | | PGPDGITPRTVRSIPS |
| | | AACTGTCAGGAGTATTCCGTCAGGAGTTATGCTTCGCATAATGAACCTG | | | GVMLRIMNLILWCG |
| | | ATACTTTGGTGCGGAAAGTTGCCTGTCTCCATCCGACACAGGCACGAACCA | | | KLPVSIRQARTIFIPKV |
| | | TCTTCATTCCGAAGGTGGGGAATGCTTCTCGACCGCAAGACTTTCGTCCA | | | GNASRPQDFRPITVQ |
| | | ATTACGGTACAATCTGTTATGGTAAGGATTTTAAATGCCATTTTGGCTTC | | | SVMVRILNAILASRLT |
| | | CCGGTTGACCTCATCAGTCGACTGGGATCCGCGTCAGCGAGGTTTCCTT | | | SSVDWDPRQRGFLP |
| | | CCAACCGACGGATGTGCCGATAATACGACGATAGTCGACTTAATCTTAA | | | TDGCADNTTIVDLILR |
| | | GGGATCACCATAAACGTTGTAAATCACTTTATATCGCAACTTTAGATATA | | | DHHKRCKSLYIATLDI |
| | | AGCAAAGCATTTGACTCGGTGTCTCATGCAGCAGTTAGCGCCACTCTAA | | | SKAFDSVSHAAVSAT |
| | | CTGCATATGGTGCCCCTAAAGAATTCGTTGACTACGTACACAAAATTCGTAC | | | LTAYGAPKEFVDYVQ |
| | | GAGGTCTGTGGCACAACGCTCAATGGGGACGGATGGAGATCAGAGGA | | | NSYEVCGTTLNGDG |
| | | ATTCATACCTGCTCGAGGTGTCAGACAGGGTGACCCGCTATCTCCCATA | | | WRSEEFIPARGVRQG |
| | | ATATTCAACTTGATCATCGATCAGTTGCTTAGGTCCTACCCCAATGAGAT | | | DPLSPIIFNLIIDQLLRS |
| | | TGGTGCCACAATCGGTGATCACACACAACAAACGCGGCCGCGTTCGCAGAT | | | 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 |
| | | | CTTACGTGAAGCGGGGACTTTTTCGTCCCCAACACGGCTGGGTGTGTCAG | | | | 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) |
| | | | CGATCAAAAAAAAAAAAAAAAAAAAAAAA (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 |
| | GACCTTCGGGCATCAAGAAATCTCATTATCTTCATCTCCGGGGCCTGACG | | 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 |
| | AACTTCACTGCAACCGGGAGGGTTCGATGCAATCCGGCCGAGGACATT | | | | VRCNPAEDIGPKLQR |
| | GGTCCGAAGCTACAAAGATTGACAAAGGCCCCCCTCAAACCACACAACAGA | | | | LTKAPLKPQQRMFAL |
| | GGATGTTCGCCCTTAGGACTGTCCTTATCCCACAGCTCTATCACAAGTTA | | | | RTVLIPQLYHKLALGS |
| | GCCCTTGGGAGTGTGGCGATAGGCATCCTACGAAAAACTGACAAACTA | | | | VAIGILRKTDKLIRYYV |
| | ATACGATACTATGTGCGAAGATGGCTAAATCTTCCGCTGGATATACCGA | | | | RRWLNLPLDIPIAFIH |
| | TAGCATTCATTCACGCACCCCCAAAAAGTGGAGGTCTCGGAATTCCATC | | | | APPKSGGLGIPSLRW |
| | ACTTAGATGGGTAGCTCCAATGTTAAGGCTAAGACGTTTGAGTAATATT | | | | VAPMLRLRRLSNIKW |
| | AAATGGCCTCACCTCACGCAAACGAGGTAGCCAGCTCTTTCCTCGAAG | | | | PHLTQNEVASSFLEA |
| | CAGAAAAACAACGGGCCCGAGATAGATTATTAGCAGAACAAAATGAAT | | | | EKQRARDRLLAEQNE |
| | TGTTATCGCGTCCGGCAATAGAAAAATATTGGGCGAACAAATTGTACCT | | | | LLSRPAIEKYWANKLY |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CTCAGTTGATGGTAGCGGACTCCGTGAAGCAGGCCATTGGGGCCCGCA | | | | LSVDGSGLREAGHW |
| | | | ACACGGGTGGGTTAATCAACCCACGCGTTTACTAACAGGAAAGGAATAT | | | | GPQHGWVNQPTRLL |
| | | | ATAGACGGTATTCGTCTGCGGATAAATGCCCTACCCACGAAGTCTCGTA | | | | TGKEYIDGIRLRINALP |
| | | | CTACAAGGGGAAGGCACGAATTGGAACGACAGTGTCGTGCAGGATGTG | | | | TKSRTTRGRHELERQ |
| | | | ACGCTCCCGAAACAACAAACCATATTATGCAAAAATGTTACCGATCGCA | | | | 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 | . | GGGATCTGGGGTAATTGCGAGCAGAGGGGGAGTATTTTTCTGTAATTC | GGGATCTG | TAGCTA | | TCLAANLSGKNFSDG |
| | | Drosophila simulans | GTAAGTCATATCATATGGTGTGCGGAAGGGGAATTTTACTCTGTAACTC | 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 |
| | | CGAATAATGAACCTAATTCTATGGTGCGGTAATCTACCACACTCTATCCG | | 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 |
| | | | TGGATGATGCACACCAGCGCAAAATCAATAGATACGACACAGACCGGACA | | | | VALRQDHIDVIDIQIV |
| | | | TACGAACTGAATTGCGTCGCAGATTCGAAGCCGCAGGTGACATTGAATT | | | | TDGHSMDDAHQRKI |
| | | | CCATTCTGCCACACCCTGAACTGGAGGGGGGATCTGGAGTGGTCAATCCGTT | | | | NRYDRPDIRTELRRRF |
| | | | AAAAGATTGATAGCAAAGGGTCTCCTCAGCAAATATGATAGTCATATCA | | | | EAAGDIEFHSATLNW |
| | | | TTAGCGTCCAGGTTATGAGAGGCAGTCTCGGTTGTTTTAAACAGTTCAT | | | | RGIWSGQSVKRLIAK |
| | | | GTACCTGAGCGGGTTTTCCCGAGATTGGACTTAGCTAAAACGTTTGGTT | | | | GLLSKYDSHIISVQVM |
| | | | CAAAACATTTGCTTGCTGTCTTGGCATAACATCAATAAAGGCATAAACAT | | | | RGSLGCFKQFMYLSG |
| | | | CGCAAAATAATGGTTATATATAAATGGCTATGAGGATGGTTTTAGTACG | | | | FSRDWT (SEQ ID |
| | | | TAGGCGTTGCGGAACTTCGGTTCAGATAGAGCAATGAATCGTGCATGCT | | | | NO: 1488) |
| | | | AGGAAAACTGACCACACGCAGTGTTGGCAGCCCTAGTATCTTTCGATAG | | | | |
| | | | ATTTCCATACCTCCGCGATCAAAAAAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1120) | | | | |
| R2 | R2_DYa | . | 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 | | LTRPPRRPRNRRRESR |
| | | | 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 | | | | WRGIWSGQSVKRLI |
| | | | TTCATGTACCTTAGCGGGTTTTCTCGTGATTGGACGTAGCTTAAAACGTT | | | | AKGLLSKYDSHIISVQ |
| | | | TGGTTCACATACATCTGCCTGCTGCCTTGGCACAATATCAAAAAGGCATA | | | | 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 |
| | | | GTCACTGAGACTACCTGCGGCTAAAGGCGCGCGCGAGGGATTGTAAGT | 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 |

| | | | | | |
|---|---|---|---|---|---|
| | | GTTATTTTAGAGGGCTCTGACGCCGAGGGGGACGAGGATGTTTTAAT | 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 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GGCATGACCCTCAATCCACGAAAATCGTCCTCCTTCCTCATCACTGCGAT | | | | VPGARPSSALKYLGV |
| | | | GAGGAAGACCTGGATATGTAGGGACCCGGGCTTGTCAATAGGGGAGA | | | | NYTLSEGLESGALIDK |
| | | | CTAAGGTGCCTGGCGCCAGACCCTCCTCTGCACTTAAATATCTCGGAGT | | | | LMQAVNRARGLALK |
| | | | AAACTACACACTCTCGGAAGGCCTGGAGAGTGGTGCCTTGATAGACAA | | | | PLQKVNLILERIIPKFL |
| | | | GCTCATGCAGGCCGTTAACAGAGCACGAGGTTTGGCCCTTAAACCGCTG | | | | YGIILGGPSLTRLHAA |
| | | | CAAAAGGTCAACCTCATACTTGAACGAATAATCCCCAAATTTTTGTATGG | | | | DKCVRMAVKEILHLH |
| | | | TATCATCCTTGGCGGGCCTAGTTTGACTAGGCTGCATGCAGCCGACAAG | | | | PSTTDHVLYARKKDG |
| | | | TGTGTCAGGATGGCTGTAAAGGAGATCCTGCACCTCCATCCTAGTACAA | | | | GMGIPRLAHLVRLAS |
| | | | CAGACCATGTACTATATGCCAGAAAAAAGGATGGGGGAATGGGGATTC | | | | LRSGLALLASGDVAV |
| | | | CCCGACTGGCTCACCTTGTCCGCCTAGCCAGTCTCAGGTCGGGGTTAGC | | | | QAAGMAGDLEGRCK |
| | | | TCTGTTAGCGAGTGGGGACGTTGCCGTTCAAGCAGCGGGGATGGCAGG | | | | KVANDLRLNWPVTL |
| | | | CGACCTGGAAGGCCGCTGCAAGAAAGTCGCAAATGATTTGCGGCTTAA | | | | RDVVRASNKFKSQES |
| | | | TTGGCCGGTGACGCTGAGAGATGTGGTGAGAGCCAGCAATAAATTTAA | | | | KDWERLASQGHGVK |
| | | | ATCTCAAGAATCAAAAGATTGGGAGCGTCTGGCCTCCCAGGGTCACGG | | | | DFRNDRLGNCWLYD |
| | | | AGTCAAGGACTTCCGGAATGACCGTCTTGGGAATTGCTGGCTCTATGAT | | | | PTVLSSSRYTDALRLR |
| | | | CCGACGGTGTTGTCCAGCAGATACACTGATGCTCTCAGGCTGAGGA | | | | TNTFGVNVALRRADK |
| | | | CTAACACGTTCGGCGTCAACGTTGCCCTAAGGAGGGCCGACAAGGACC | | | | DLEVNCRRCHGKPET |
| | | | TTGAGGTCAACTGTCGGCGATGCCACGGGAAACCTGAAACCTTAGGGC | | | | LGHVLGECVAGKGM |
| | | | ACGTGCTCGGAGAGTGTGTGGCAGGTAAGGGTATGAGGATACAACGAC | | | | RIQRHDKMAAFVAT |
| | | | ACGACAAGATGGCGGCCTTTGTGGCTACAAAATGTGAAGAAAAGGGGT | | | | KCEEKGYQTTREQLF |
| | | | ACCAGACGACCCGCGAACAGCTATTCTCCATCGAGCAGGGAAAACTCAA | | | | SIEQGKLKPDLVVIDG |
| | | | ACCTGACCTCGTCGTAATAGACGGGGAAAGGCTCTCATTGTCGACGTA | | | | ERALIVDVTVRFESG |
| | | | ACAGTCCGCTTTGAGAGTGGGAACGCCCTCTCACGGGGAGCCTCTGAG | | | | NALSRGASEKIEKYQP |
| | | | AAGATCGAAAAGTATCAGCCTTTGGCTGACTATTTTGTCTCTCAGGGGG | | | | LADYFVSQGAVREAN |
| | | | CTGTTCGAGAAGCCAACGTGTTACCGATTGTTGTCGGTTCCAGGGGAGC | | | | VLPIVVGSRGAITQAT |
| | | | AATCACCCAGGCCACATTGAAGTCCCTAGCTACGCTAGGGTTAGATGTG | | | | LKSLATLGLDVERVGK |
| | | | GAACGTGTTGGCAAGTACCTCGCCATCTGTGCCGTTGCCTCCAGCGTTG | | | | YLAICAVASSVEIACM |
| | | | AGATCGCATGCATGCACCTGGATTACACATAATGTCCCTTTTGGCTTGCC | | | | HLDYT (SEQ ID NO: 1490) |
| | | | CCCACCTGCTTAAAGGAACTGGCAGGAAAGAGAGTGATCCGTGCCATA | | | | |
| | | | GAAATATGGTTATCCGGGGCAAGTCACTAGCAATATGGGACTTCTCCGG | | | | |
| | | | GTCCGTGCGGTCCTTCCAACATGAGCTGGACGTAGTCCACTCTATGACTT | | | | |
| | | | GAACGATACGGGGGGCGTATCTCCCCCGGAAGAGGTCGCCTAGGCGACT | | | | |
| | | | TAGAA (SEQ ID NO: 1122) | CTCACTGC | TGAGTT | | MMADYNNSVDHAL |
| R2 | R2_RL | Reticulitermes lucifugus | CTCACTGCTGTCATGCCATTGTTAATGCGTTGGGTGATGGCGGGTGATG | TGTCATGC | AGATAT | | EDNTRLIFARDAVLA |
| | | | GGAGACGAGTTACAGCAGAGCTGGCTTCACGGGGGGGGCTGTAGTTGC | CATTGTTA | ATATGA | | RVCGPFDNLECGLCG |
| | | | CCGAACCAGTCTGCTTCTGAGTGCCTCCACGTGGCCTCGCTGGAAACGT | ATGCGTTG | CGACTG | | VLLTSLQGVREHCHR |
| | | | CGGAGCTGCTTACGGCTAACCCGTAGACAACTCCGGCGGCCAAACTCAG | GGTGATG | ATTCTTC | | AHHNLDLTFQCTKCD |
| | | | AAAACGGCCTTACTACCAGGGTCATCCCTCGCGGCGGCGCTAAGACCTT | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | | ACTGTATGTACTACGGACTACAGTTGAGTAGCGCGGGAGAGCACTTGAT | GCGGGGTG | 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 |
| | | CCTCGTTCTTGTAGGGGGGCGAAGGACGAGGAATGCATCCGTCCCTCCAG | 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 |
| | | AGGAATGCTAAGAGGCTCAAAGATTTTAATCCGCGGGGGGGGGGGTGGT | CCTTACTG | CAATAG | LESNTIEEKLPSSSLFD |
| | | AAAACTATTCACGGCAATACTAAGTGGACTGAAGAGGAGGTCCAGTTG | TATGTACT | ACAGCA | CLPPITPEEVEGRIHKI |
| | | CTTGTATCACTTAGCAAGAGATTTGAGGGGTATAAGTCTATCAATAAAG | ACGGACTA | CTAAGA | RPSSAPGLDGVRKIH |
| | | AAATTAGCCTTATTTTAACATCAAAAACTTGTAAACAAATTTCTGATAAA | CAGTTGAG | AATTCTG | LVGKGITLVLVKLYNL |
| | | CGTCGCTATTTAAATTTGCATAAATGGAAATGGGGGGCTAGCTGCAGCGG | TAGCGCGG | CAAACT | LFLTGGYPECWKRNR |
| | | AAGCTGTATTAGAATTTTGTGAAGATAGTCACCCTGAAGTTACTGAGAG | 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 |
| | | TTCCCCCCATCACGCCGGAGAGGGTTGAGGGAAGAATCCATAAGATTC | TGGATAAA | | YDAGQRLISAAKRGS |
| | | GACCGTCTTCCGCCCCAGGGTTGGATGGGGTACGCAAGATCCATCTTGT | TACTCCAA | | QLSLKPQQKINLITAF |
| | | CGGTAAAGGGATAACCCTGGTCCTTGTTAAACTCTACAACTTGCTATTTT | AATGGCTT | | LLPKFLYILIEDPPSPA |
| | | TGACTGGGGGATATCCTGAGTGTTGGAAGCGGAACAGGACCGTCTTTA | AGCCACCT | | YLKSIDHDLRQIYKNIL |
| | | TACCGAAGATTGGGAAAGATCTGAGTGAGGTGGGAGGATGGAGACCG | CACGTGGA | | HLPNCVSTAFMYSPK |
| | | TTGACAATTGGTTCCCTATTAGCTCGGATGTATTCAGCCTTCCTAGAGAG | GAGAGGG | | RDGGLGLPRLSCLVPL |
| | | GAGGATTAGGAGGGTGACCTCCTTGTCCTTGTCGCAGAGGGGCTTTACG | GAATTAGG | | AHLKAGIKLGSLQDSL |
| | | AATATCCAAGGTTGTCACGTTAATTTGACTATCCTCAAAGAAGGCATTCG | GGTTTACT | | VREITTSDRFVRTMG |
| | | ACAGGCCAAAGTCAAGAACGGGGGAGTTATAGTCTCAGTAGACATAGA | 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 | | |

| | | |

| | | | | | |
|---|---|---|---|---|---|
| | | 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 |
| | | CGTATAGGGGAATTTGTTGTCATTTTTTCGAAGTGTAGAGGAGCTCGTAT | 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 |
| | | AGCTGTATTAGTATTTTGTGATGATAGTTCACCTTGAAGTTACTGACAGCG | 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 |
| | | TTTGGATGACAAATTGACTAGTGAATTATCTTCCGACAGCCCCGTGTGG | CAAGTCCA | CTTAAA | NIAIDPVLRSLEEFQG |
| | | AATGACAACTGCGAACTTATTGTGTCGGACTTATGTCACGTCTTATGCAA | TGGGGGC | (SEQ ID | GLPLGNSAIKILAFDD |
| | | AAATAAATATGAGTTAGGTAGGCAGCATCACGTTAGGAAAGGAAAGAG | GACCCGGC | | DIILGASSAGQAQQ |

| | | | | | |
|---|---|---|---|---|---|
| | | GCATCGGGGTATACATCATAAGAGGGAGAAATTTAGGGAATGCCAGAA | TGTGGGA | NO: 1370) | MVDMLGIGLTSCGL |
| | | GATATTTAGAAAGTCCCCCAGAAAGCTTGCCGAATATCTCTACCGCGAT | GTTATCCT | | GVSHRKCFGFQIVNK |
| | | AAGGACCTGTCTCATATCAGTAAAGACGTTAGCACTCCGCAGGGTATAG | GCACCCGG | | NKTWAIVDPMITLN |
| | | AGCAATACTATTCTCAGTTATGGGGGAAACCTGAATTACTAGAATCCAA | 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 | | |

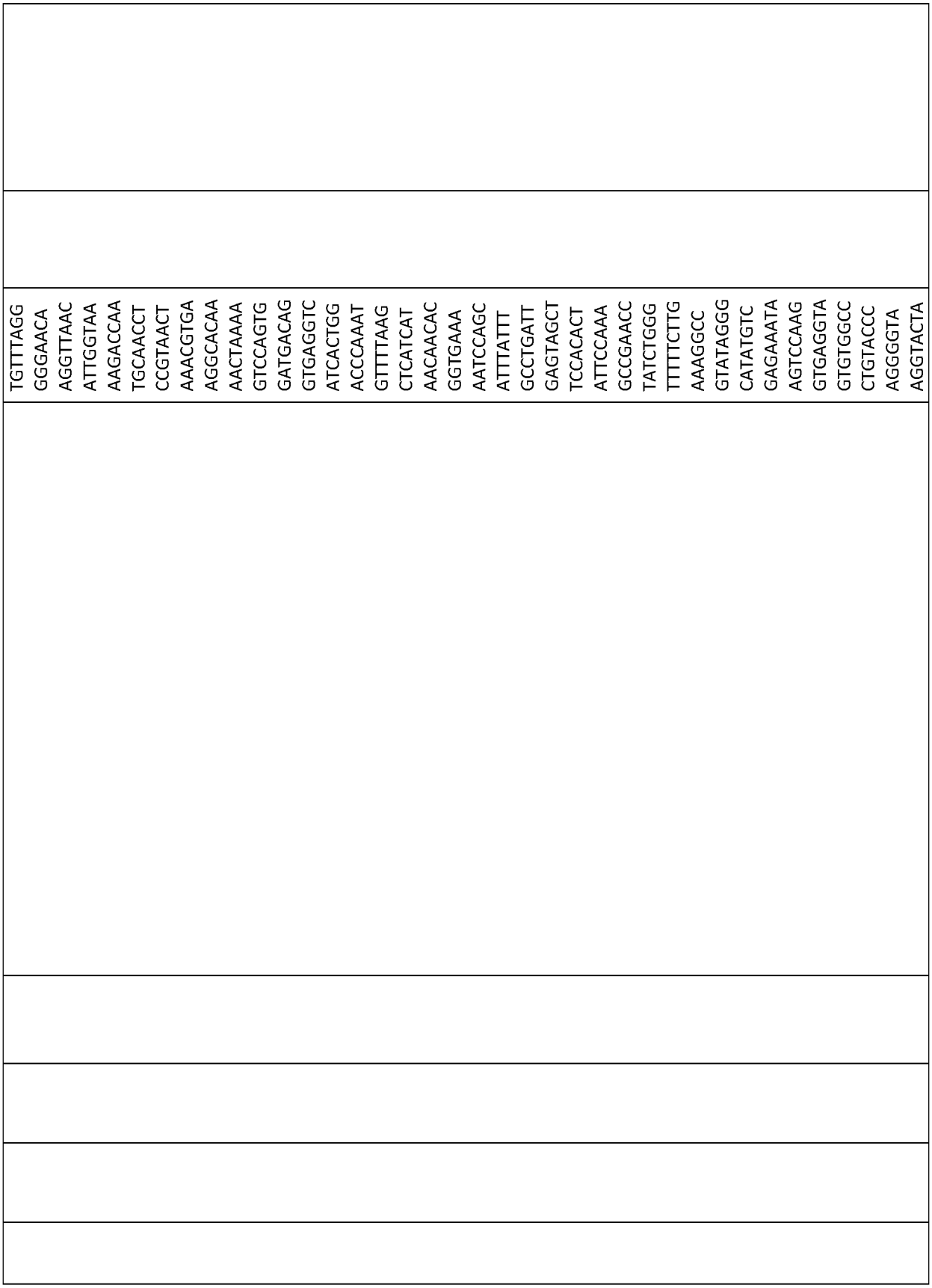
| | | | | | | |
|---|---|---|---|---|---|---|
| | | | CTTGTAGGTTCTGTAAGGATATACCTGAAACGCAGGCTCATGTTCTTGG | TATATTTTC | | |
| | | | ACTTTGTCCAAAGACCAAGGGAATGAGAATACTAAGGCACGACTCAATT | CACCGATC | | |
| | | | GTGAATAGGGTGCGAGATAAACTTAAGACTAAGTCGCCCGTTGCTCTCA | GGTTCGCT | | |
| | | | TGCATGAGCAAAATTTCACTGTTGAGGAAGGCCAGGTGTTCAAACCAGA | GTGGGGG | | |
| | | | TATAGTGACGATTTTGGGAGAGGTGGGTTACGTAATTGATGTGACGGT | GATGTCGG | | |
| | | | ACGGTATGAAGATAGAGATTACATCAAGGATGCCTCTGTGGAAAAAATC | GGCCTGGT | | |
| | | | AGGAAATACGAGGCCCTTAAAGGGTACTTGAAGGATCTGTATCCACAGC | GGCCAGG | | |
| | | | TGAATAAGGTAGAGGTATTGCCTCTTGTCTTTGGTAGCAGAGGGGCCGT | 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<br>TGTCGAAG<br>AGGACATA<br>TGGGGTAA<br>CCCATAGA<br>ACCTAGGC<br>AGGAGTA<br>ATCCCTTA<br>GCTGGGG<br>GGGTCAGC<br>TGGTGGGT<br>CCTGTCAA<br>TTTATCCTC<br>CCCTCCAT<br>GCCAGAGC<br>CGGTCCGA<br>GGTTAGAG<br>ACGGACTC<br>ATTTTCTCC<br>TTTTTATAT<br>GTC (SEQ ID NO: 1247) | TAGAAA<br>TGTGTG<br>CGATAA<br>GGTGTG<br>AATAGA<br>AGGGTT<br>CACCAA<br>GAGGGA<br>GACCTA<br>GTTTGG<br>ACCTCA<br>GAAATG<br>GGGTCA<br>TAGGAG<br>TGATAG | MSNYNETNTSGGDN<br>PRMATQTTGSLSSGP<br>INQHTCELCCRTFGT<br>RAGLGQHVRKTHPIE<br>SNQSINVERKKRRWS<br>PEEIRRMANMEAQA<br>TINNIKHLTQYLATYL<br>PQRTLNAIKGRRRDA<br>EYKELVTGIIANLRSN<br>SSTQQTNQVCNESE<br>MSQRSKILQSIRESVR<br>DLRSRRNKYAKALQE<br>LGEAALCGKMLNEE<br>QLIHCIKSMFNTAKC<br>PKGPRFRKTATHSGT |
| R2 | RaR2 | FJ461304 | Rhynchosciara americana | CAGAACGTGGAGAAACGGAATAACTACCCAGATCCGTTGGTTAACCGG<br>TGGCAAAGTTAATCAAGGTTGCCATAGGCTTAATAACCCTATGGAAATG<br>TTTCCACACACCTCCACGTGGTGCCTGCCGGAAATTGTTCTAGGGTGAA<br>CAAGCTAAGTTGTGAGAAACGGGCTCCACCACAATATGGAGCCTGCCA<br>GGGCGCGAGACTCAGGACTCTCCATGTACAAAGTGGTTAGTTGCAAAA<br>AGAGTGCGCCTAGCATGACTGATAATTTTTCACTGAAATAACGTTGAAC<br>TTTATCTGTGTCATGTGCACAACACTATGGTGTCTGATCAAGCACCATCA<br>GTGGTGGACCTGCTAATGTATTAGTAGAACGTGTCCAGGCGATAATGCA<br>CACACGGCCTCCGGGCCCATCGCTTTTTTTTGAGATTCCCTAGAAACTTCA<br>GTGTGTGCGACAACTGTATAACCCATAAGGGATGGACAAAGGTTATACT<br>AGGGGGTAAAAACCCTAATCGGCTAATGGCAAATGGGATGTAGAAATG<br>CCAAAGATTACTCGCACCGAATAATGGTGGCCGAAAAGCGGGTAATCG<br>AATGAAATGGTAATGCTGTAGCGGAAACATGATCACATTCTGTGACAGT<br>AAACCATTAGACCTAGGGGGAACTATGATTAACAAGATACCAGCTTACA<br>TGGAAGCAATGAAATGTAAGTCACAGTAGTGATAAGTGGTGAAGAGTC | CAGAACGT<br>GGAGAAA<br>CGGAATAA<br>CTACCCAG<br>ATCCGTTG<br>GTTAACCG<br>GTGGCAAA<br>GTTAATCA<br>AGGTTGCC<br>ATAGGCTT<br>AATAACCC<br>TATGGAAA<br>TGTTTCCA<br>CACACCTC<br>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<br>ATAGGAGTGATAGGTTGTAAAGCCGTTGGGGAATCCGGCTACACATGG<br>TATCTCAGGAGCCATTCATGCGCTGATCTCATTAAGGCGTAATAAACTGT<br>GAAACAGATCCTGATAATGCCGTGCTACCAAATGATGTAACGAGGCGG<br>AAATAAAATTAATCTGGGGCGTTCTGCGGAATGACTACTAAATATAGCG<br>ATGCTATATATACAAACGACTGATGGTAACACCGGCCTTA (SEQ ID NO: 1125) | GGTGAAG<br>AGTCTTGT<br>AATCACCG<br>TAACTAGG<br>CCAGGTTC<br>TGGGGAT<br>GCCATGAA<br>CTTAGGGG<br>GAGTATGG<br>TTAGCAGA<br>TCTACCAG<br>CTAACACT<br>ATTACTGA<br>TAATATGT<br>AAGCCGCA<br>GTAGCGCT<br>AAGTGGTG<br>TACAGATT<br>TGCAACCA<br>CCGTAACT<br>AAGTTCTG<br>TTTCGATG<br>GACTAGG<br>GGGAATCA<br>TGATTAAC<br>AAGATACC<br>AACTTACA<br>TGGAAGTT<br>ATGAAATG<br>TAAGTCAC<br>AGTAGTGA<br>TAAGTGGT<br>GAAGAGTC<br>TTGTAATC<br>ACCGTAAC<br>TAGGCCAG<br>ATGTAGTC<br>AAAGCACA | | |
| | | | | | |

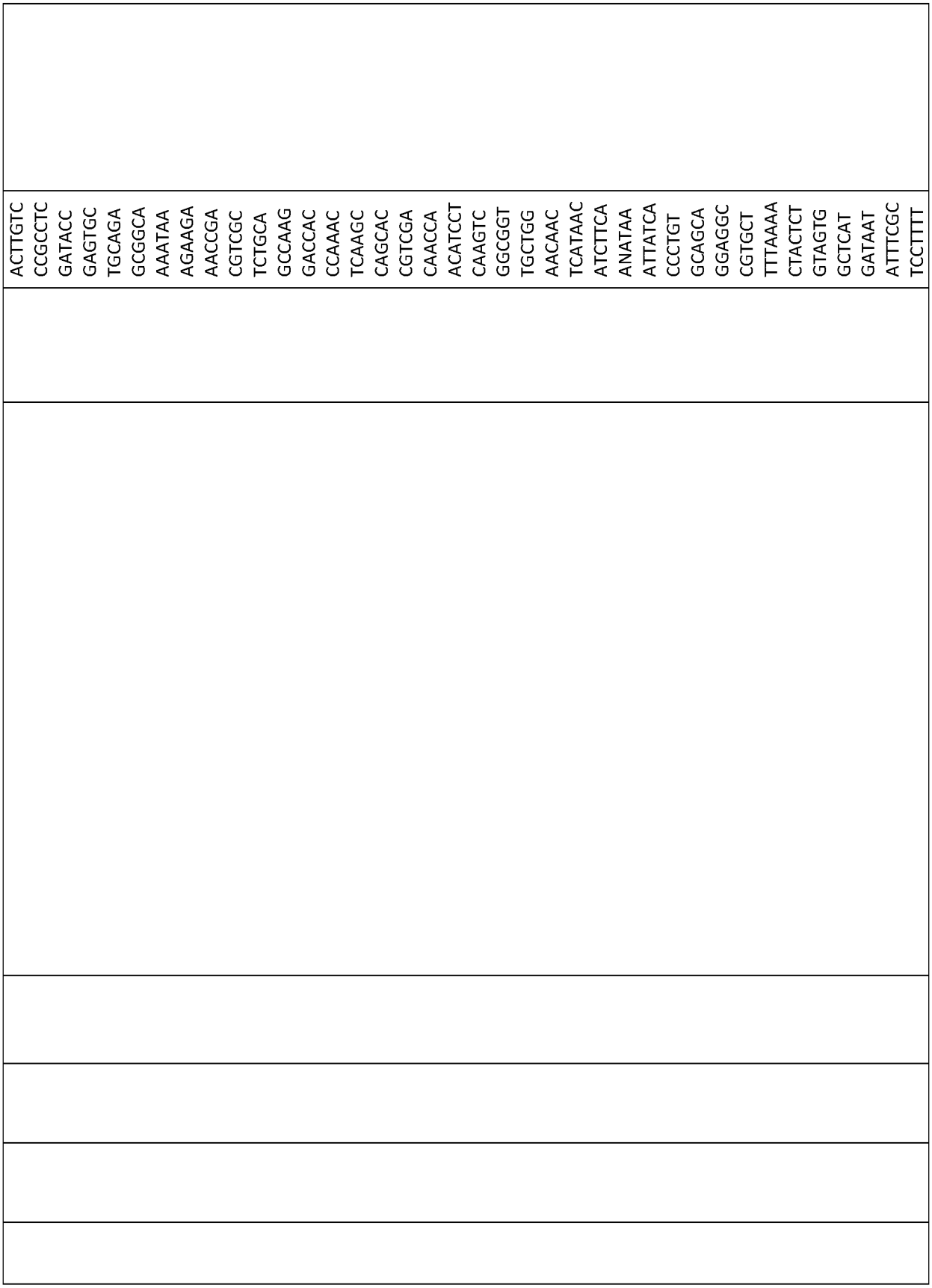

TGTTTAGG
GGGAACA
AGGTTAAC
ATTGGTAA
AAGACCAA
TGCAACCT
CCGTAACT
AAACGTGA
AGGCACAA
AACTAAAA
GTCCAGTG
GATGACAG
GTGAGGTC
ATCACTGG
ACCCAAAT
GTTTTAAG
CTCATCAT
AACAACAC
GGTGAAA
AATCCAGC
ATTTATTT
GCCTGATT
GAGTAGCT
TCCACACT
ATTCCAAA
GCCGAACC
TATCTGGG
TTTTTCTTG
AAAGGCC
GTATAGGG
CATATGTC
GAGAAATA
AGTCCAAG
GTGAGGTA
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 |
| | | | | CCAGACTACTACCCGGATACAACACCCTTGGAAACGAGAGGAATGGCCGG | GAATTTTC | CTGGGC | PPKPTPFFSPRTLERS |
| | | | | TCACAAGATCACAATGTCGGAGGGGAAGCTGCTGGAGGTTGCTGTACG | CGCTGCTC | TGCACA | ASWPALSEVVDPLPK |
| | | | | GTATGGTGGTGTGAGGAATGTCTCGTACGAGTGTCCAGTCCCGGATTGC | TCCACTCT | TCCCACA | LKEKHPSLPCALDKCP |
| | | | | ACAAAAACCTTCTCACAGGCAAATAACTTGATACGGCATTTAAATAACTT | GCTCCAAC | CGTAGG | PSPRIKPSTLVPPCHT |
| | | | | TGGCAATACAAAACACCGAGCACACAATTTCACCTATTTTTTCACCTGTG | ACCACCTC | GACCTG | ANNSPKPTSPESPSTL |
| | | | | AGAAGTGCAAAATCCAAATTCACAGCAACACACAAAACATAATATTTCAAA | TCCTGCTC | TTTATAT | KPLPRPIRPSKPLEDW |
| | | | | TCATTACAAACAGTGTTGTGCAACCGGCGCGGCGGACCCTCGTGCGAGAC | 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 |
| | TCTCGCCAAGAACACTTGAGAGGTCGGCGTCGTGGCCCCGCATTGAGCG | GTGGATAG | CATTTTT | MKATVRRIIDGERGD |
| | AGGTTGTCGACCCACTGCCCAAGCTCAAGGAAAAACACCCCTCGCTGCC | AGCAGCG | ATATTTT | ARCEIDPKTIEGRFRD |
| | ATGCGCCCTGGATAAGTGCCCACCCTCACCCCGGATCAAACCCTCGACC | GCCTCTAC | GTAAAA | ELSPPVREGPECSLPP |
| | CTAGTCCCTCCCTGCCATACAGCAAACAATAGTCCCAAACCAACTTCACC | AACCAAGT | CACAGA | WMAEAQAGEHAPS |
| | CGAATCACCCTCTACCCTAAAACCCTTACCCCGCCCCCATCCGACCATCTA | TCTCCACT | TTTTTAT | NDSQPGDAYDGPIT |
| | AACCCTTGGAGGATTGGTTAACTGTGCGGAGCGTGGGACCGGATCGGG | 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 |
| | GCGCTTTCGTCGACCAACCGAAGGCCAAAGCCGAGGCGCGGCGCAGGTAC | 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 | |
| | CTCCGGAACGCGTTCGGCAGCGTCGACCACACCACCATAGTCGAGGCGT | CACCTACA | ACAGGG | VRRCVKTALRLPKGS | |
| | TGTCTCGCTTCGGCGCCCCCCGCGGGCCTCGTGGAGATGATATCGGACAT | AACTCGTC | ACTCCCT | CRAFFHTSPDAGGLG | |
| | CTATACGGGTGGGTCCTGCCGAATTAGAACTCGTGCGGGATTCACCCCG | TGAATAGA | TCCGGA | ITSVVAECDILTVTQA | |
| | GACATCCCTGTCGGTCGTGGAGTGCGACAGGGGTGTCCCTTGTCCGGTA | CGCCCCCG | GCCCCT | FKMLSSPDHLVSLVA | |
| | TCATCTTCAACCTCGTGATGGAGGTCCTGCTGCGGGGCGTCGAAGCGAA | CGTGGGA | GCACCC | KGRLGMHAARMGR | |
| | CAACGCTTGTGGATACCGGCTCTCCTGCGCCGGCGGCGCGTCCGTCAGG | 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 | |
| | GCGCCATGGCGGACTACCTGAGCGGGGACTCGGTAATGGGGCACANGT | 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 | |
| | | AGTTTTTTTTCGTTTTTTTCGTTTTTTCACCCCCACCCCACTCGCCTCTGGGC | GCTACAGA | TCGAGA | |
| | | TGCACATCCCACACGTAGGGACCTGTTTATATTATTTGCCTTTTATATGTA | GCATTTTA | AGCGTC | |
| | | CCACTTTTTAAATATATTTTTGTACCCCACAAGATGCTTTTCGCCAAAAAA | TTTAATCA | ACATCG | |
| | | AAAAAATTTTTTGTATCACATTTTTATATTTTGTAAAACACAGATTTTTTATA | ATTTGTTA | CCCCACT | |
| | | AACTTTGCACTATTTTTATATAAACTTCGCACTTATTTAAAATGAATCGCA | CCGACCTC | GACCAC | |
| | | TCTTTTTTATATACACCAACACAAACAGGATGTGCAGCTCAGGGGAACC | CTGCTGCT | GACCTTC | |
| | | AATCCTGCGTCCCTCCTAGCGGCGGGAGGGGCGCCCACCTACCCCCACG | TCTTTTTAC | CCCGAC | |
| | | CTCCTCTTGAGCACCAACAGGGACTCCCTTCCGGAGCCCCTGCACCCTCA | TTTCTCCA | ATTGGA | |
| | | ACTTTTCTTATTTTTAAAAAAAAAAAAATCATATATATTGATCTTGACGACG | GACTACTA | GTCCTT | |
| | | GGGGCTACATTCAGCCCCCAAAAACCCCACCCACCATCCCCAACGAGTGC | CCCGGATA | GGCGTC | |
| | | CGGGGCATTGAAGAGCTCCGGCACAATTAGCACTTAGCTTATTTTATTATT | CAACCCTT | TCCCAG | |
| | | TTTTGTCAACATTTTTTGTTTTTTTCAAATTTTTTTCACCCCTCACCCCCACCCT | GGAAACG | GTCGAA | |
| | | AATAGGTCCCTCGGGCTTGGGCCCCTTTTTCGTGCTCGAGAAGCGTCAC | 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 | GCTAGCTC | TTGCCCC GTGTAA ACTTAGT NGATGC GAATAA AATCAG TTGAATC A (SEQ ID NO: 1372) | TAAAAA | MLRRGRIFLPASTKA |
| | | | | TGGGGAGCGGAAACGCGTACTATCGAAACTTACGCGGCTAACAAGGTA | CCTAAAAT | | CTAGCA | GKTRGRMKWSREV |
| | | | | AAGGTAACCCATTAATATGGAGACAAGACTAAAAAGAAAATTGAGAGG | CCTACCTT | | TAATTAT | NLFIMRTYYVVTKLET |
| | | | | GCCGCTTCCCGGGGGCGATCGCCGGGGCACACCTGGAGCTGGCGCTGG | ACGTCCGA | | TAACTCA | DLTIYRKKLHEHFSLK |
| | | | | GTGTTCCAGCATAGGATCCGTCAGCGGCGGGAGAGTTGAGCAGGCGGGC | GGCGAAC | | TAACTA | YPNVIISQQRISDQKR |
| | | | | TCTCGCCGGGGATTTACAACCCGAAAATGCTACAGCCGCCAACAGAAAT | ATCTGTCC | | ATGTAT | AIERNKLLSQETLDRL |
| | | | | CACGATGTAGAAAGTAGGAGCAATAGCCCATGTGAACCTTACAGCCCG | ACGTGGG | | ATTACTT | KEEVRKQLEDEQTNN |
| | | | | AGTACCGGTTCATACAACCCCTCGGTACAATCATCACCATCATCCTCGGG | GAGCGGA | | GGCCAA | VENEKLNSETYSHEYT |
| | | | | TCATAGAGGCTCACCAACGTCGACTATGGATGTTGCGCAGGGGGCGAA | AACGCGTA | | AAGCCC | TLTPQTILTKKTQQHT |
| | | | | TCTTTTTACCTGCGTCCACCAAAGCTGGCAAAACAAGGGGCCGCATGAA | CTATCGAA | | GTATAT | NIISSTQTSHSSTQTE |
| | | | | ATGGAGCAGAGAAGTAAACCTATTTATCATGCGCACCTACTACTACGTC | ACTTACGC | | ACAGTT | SITLLLENEVDILNTNP |
| | | | | ACTAAATTAGAAACTGATCTGACTATTTACAGAAAAAAAGCTGCATGAAC | GGCTAACA | | CCACCG | TEGATQTQEVKDKFE |
| | | | | ATTTTTCACTAAAATACCCTAATGTAATAATTTCACAGCAAAGAATATCT | AGGTAAA | | GCTCTG | TNLTMYSGMDPKAR |
| | | | | GACCAAAAAAGGGCAATAGAAAGAAATAAACTACTATCTCAAGAAACC | GGTAACCC | | TCGACA | PPLPKLKYSSKLNELIR |
| | | | | CTAGACCGACTAAAAGAGGAAGTGAGAAAACAGCTCGAAGACGAACAA | ATTAATAT | | GACTGA | LFNNDILVDYISPDTQ |
| | | | | ACCAATAATGTTGAAAATGAAAAATTAAATTCAGAAACATATTCACACG | GGAGACA | | ACTGAG | LSDVHTLTYCTAVTIS |
| | | | | AATATACGACACTTACTCCACAAACTATTTTAACGAAGAAAACACAGCA | AGACTAAA | | AAAGGG | EQLKYKIIAIEGNARH |
| | | | | ACATACAAATATTATCTCTAGCACACAGACCTCTCATTCATCCACTCAAAC | AAGAAAAT | | GAAACA | KKNFKPPWQQRLEK |
| | | | | CGAATCAATAACTCTGCTGTTAGAAAATGAAGTCGATATTTTAAATACTA | TGAGAGG | | TATGGA | DIAKLRADIGKLTQYI |
| | | | | ACCCCACAGAAGGGGCAACACAGACACAAGAAGTTAAAGATAAATTTG | GCCGCTTC | | AATAAT | NNNRSKKVVQSVEQI |
| | | | | AAACGAATTTAACAATGTACTCAGGCATGGACCCTAAAGCAAGGCCGCC | CCGGGGG | | AATAA | FKNTKIHTSHENGNK |
| | | | | ATTGCCTAAACTTAAATATAGCTCTAAGCTAAATGAACTGATACGTCTAT | CGATCGCC | | (SEQ ID | KSQEFLDTLKQKLALK |
| | | | | TTAATAATGACATACTTGTAGATTACATCTCACCAGACACACAACTATCA | GGGGCAC | | NO: | AHRLKRYNNSQKRK |
| | | | | GATGTACATACATTAACATATTGCACCGCCGTAACTATTTCGGAACAACT | ACCTGGAG | | 1373) | NENTIFLTNEKLFYRN |
| | | | | AAAATATAAAATTATAGCAATAGAAGGAAACGCGAGACATAAAAAAAA | CTGGCGCT | | | LIKPKTDRDNSNIDIP |
| | | | | CTTCAAACCACCGTGGCAACAAAGATTGGAGAAGGATATAGCAAAATT | GGGTGTTC | | | TAEQLEMYWARLWE |
| | | | | GAGAGCAGATATTGGTAAACTGACCCAATACATAAATAATAATAGATCT | CAGCATAG | | | NSAKHNDKANWITE |
| | | | | AAAAAAGTAGTCCAAAGCGTTGAACAAATATTTAAAAACACTAAAATAC | GATCCGTC | | | EKERWDTIEEMQFD |

| | | | | | |
|---|---|---|---|---|---|
| | | ACACATCACACGAAAATGGCAATAAAAAATCTCAAGAATTTTTAGACAC | AGCGGCG | | DVTEEEITTITARLHN |
| | | ACTGAAACAAAAATTAGCTCTAAAAGCCCACAGACTAAAAAGATATAAC | 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 |
| | | CAGAGTACAAAAAACGTATCAATGCCATTTGCAAAACGAAATTATCTGG | | | FFYIKSDLSEIHRAIAQ |
| | | AAAACATCTTATAAAAGCACTGAACACCTATGCAATACCCATTCTAACCT | | | NDNNYTPLNLKQKEL |
| | | ACTCATTTGGAATAATAAAATGGAGCAAAACTGACATAGAACAAATAGA | | | IDNTENLRNRNPQKD |
| | | ACGCATAACAAGGACTACATTAACAAAACACAATAATCTTCATCCAAAAT | | | MEENWKKKALHGR |
| | | CTGCAATAGAAAGATTGACAATTAAAAGACAAGACGGGGGTAGAGGCA | | | HPHDLSQSHIDSKAS |
| | | TGATAGATATTTGGCATCTATGGCGTAAACAAATACACAGCTTAAAAAC | | | NMWLKTGSLFPETE |
| | | ATTTTTCTACATAAAATCAGATTTAAGTGAAATTCACAGAGCCATAGCAC | | | GFLIAIQDQVINTKNY |
| | | AAAATGATAACAACTACACACCGCTAAATCTCAAACAAAAAGAACTAAT | | | RKYIIKDPTIRDDKCR |
| | | AGATAATACAGAAAACCTAAGAAATAGAAACCCACAAAAAGACATGGA | | | KCNTQPETIQHITGA |
| | | AGAAAACTGGAAGAAAAAAGCGCTACATGGACGACACCCTCATGACCT | | | CSTLTQTDYTHRHNQ |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | AAGCCAATCTCACATAGACAGCAAGGCATCAAACATGTGGCTCAAAACA<br>GGAAGTCTGTTCCCCGAAACGGAAGGATTTTTAATTGCCATACAGGACC<br>AAGTAATAAACACAAAAAATTACAGAAAATATATTATTAAAGATCCCAC<br>TATTAGAGACGATAAATGCCGCAAATGCAACACCCAGCCAGAAACCATA<br>CAGCACATAACTGGAGCATGTTCAACCCTTACACAGACAGATTACACTC<br>ACAGACACAACCAACTTGCTAATATTATCCATCAACAACTGGCTCTCAAA<br>CATAAATTGATACAAAATACAAACACACCGTACTACAATTATAAACCACA<br>AACCGTTCTTGAAAATGACTCCTGTAAACTTTATTATGATCGCGCTATTCT<br>TACCGATAGGACGATTCACTACAATAGACCGGATATCACTTTACAAGAT<br>AAAAACAATAAAGTCACCTACATTATTGACATTGCAGTCCCGAATACCCA<br>CAACATTCAAAAAACGTTTACAGAAAAGATGACACAAAATACACAGAACTT<br>AAAGAAGAAATAGTTAGAATTTGGAAACAGAAAAAAGCATACATAGTC<br>CCAATAATAATCTCAACCACTGGAGTTGTCCCAAACCACATCCACAACAG<br>CTTAAAGCTTCTAGATTTAAAAGATAACATATTTATTTCACTACAAAAGG<br>CAGCTATCCTAAATACATGCAGAATAGTGAGAAAGTTCATGCAGCTTGA<br>AGAAAACCAAACTTACTACACGCAATAAAAACTAGCATAATTATTAACTC<br>ATAACTAATGTATATTACTTGGCCAAAAGCCCGTATATACAGTTCCACCG<br>GCTCTGTCGACAGACTGAACTGAGAAAGGGGAAACATATGGAAATAAT<br>AATAA (SEQ ID NO: 1127) | GAAGGCTA<br>ACCACAAT<br>A (SEQ ID<br>NO: 1251) | TAACATC<br>CGGTGC<br>AAACTC<br>ATTAAC<br>ATTAAG<br>AAAAGA<br>GAGAGG<br>AGAAAT<br>GAGAAT<br>GAGATT<br>CATTCAC<br>CTTTGG<br>CATTTGA<br>ATAGCC<br>CGGGGT<br>AGGTGA<br>AAAGTT<br>CCCAGC<br>ATATTGC | LANIIHQQLALKHKLI<br>QNTNTPYYNYKPQT<br>VLENDSCKLYYDRAIL<br>TDRTIHYNRPDITLQD<br>KNNKVTYIIDIAVPNT<br>HNIQKTFTEKMTKYT<br>ELKEEIVRIWKQKKAY<br>IVPIIISTTGVVPNHIH<br>NSLKLLDLKDNIFISLQ<br>KAAILNTCRIVRKFM<br>QLEENQTYYTQ (SEQ<br>ID NO: 1495) |
| R4 | DongAG | AB097127 | Anopheles gambiae | GAAGGCTAACCACAATAATGGAGACACGATCTATGAGAAAAAGGACCA<br>CACGATTGCCCGAGGAGGGAGCCCCTACTGGAGCTGGGCCTGGGACGGG<br>GAGACAGAGCAAGCATCCAGCGGCTGGAAGATGAAATGGTGCAAGAG<br>CGTTCTTTCAGTCAACGGGCTCTGCCCGTACCACGAACGCAAAACAGAA<br>ACGGCAGCCCTATAAATCACCAAGGCAATGCTGCATCTGCTAATGTTGC<br>TGTGGCTGATAGACAACAGTCACTCATTTTGGCAGGAGGCCGACGGCA<br>GAGGATTATGTGGACGAGAGAGATGAACCATTACGTGATCCGTTGCTA<br>CTATGTTTACACGAGGATGGAGACGGACATGCCCGGCAGAGTGAAGAT<br>GCTGGGTATGTTCAATGACCGTTTCCCTCGGTTTGCGCATCAGCTTGACC<br>TGAGCAAGTTGTATATACGGCAGCGAGCTATTATTTTTGCCTGAGGAACT<br>GGAGTTCATCAAGCTGGAGGTGCGGAGGGAATTTGGAGAGGAAGAGG<br>CAGGCTGGCGCGAGTCGAGTAGGATTTCTGCTAGGCTTAATACAATAGA<br>CCAAAATACATCAAGGGCGAGTGAGGATCGTGATTTGGATGAACCCAC<br>CGCTCCAGGATTGTCAGTGGATATCCAACACCAGATGGCCACAGCTGTT<br>ACGCAGTTCCATGGGACAGACCCCTTGTCTCGTCACCGACTACCAAAAC<br>TGCATTATTCTTACCGCCTGAAAACAGCAGTAAGCATCATAAACCAAGAT<br>GTTCTACCTCAGTATTTGGATAGCGTAGGGAGCATTGAGGATCTGCAGT<br>TAATTGTGTATTCGGCTGCGGTGGCTGTTGTACGAACGCTATGGTTGCG<br>GACCTATCCGCAAGGAGACAGCGAAGGTCGACCATGCTCCAAGGCTGA | | | |

| | | | | | |
|---|---|---|---|---|---|
| | | AAAGCCCGCCTGGATGCGACGTCTAGAAAACCGGATCAACGCAACACG | TGAGAA | | LRTYPQGDSEGRPCS |
| | | GACAAAGATTGGTCGAATGCAGGAATATCAACGGGGGAAATTCATCTAT | GTGACA | | KAEKPAWMRRLENR |
| | | GAAGGTGGTACGTCAGATTGCTGAAATGGTTAAACCTAAAGAACTACG | AAATTC | | INATRTKIGRMQEYQ |
| | | AGACCTCACTGATGCCAACATAACGGAGGTACTCGACATCCATTTACAA | GGATAA | | RGNSSMKVVRQIAE |
| | | CGGTTGAGTGCCCTTGCAAAACGATTACGACGTTATGCTGAATGCTCGA | TAATAAT | | MVKPKELRDLTDANI |
| | | AGCGGAAAGAACAAAATCGAATGTTCAACATTAACGAGAGAGAATTTT | AATAAT | | TEVLDIHLQRLSALAK |
| | | ACAACTGGATCCGAAATGATAAGCCCAATTTTAGAGAAGGGCTCCCGGA | AATAAT | | RLRRYAECSKRKEQN |
| | | TATTGGCGACTTTACACAGTTTTGGGCCAATCTATGTGAGAAACCTGTCC | AATAAT | | RMFNINEREFYNWIR |
| | | AACACAACAGCGAAGGAATGAGGTTAGCAGAAGATGAGCGCTTCAGTG | AATAAT | | NDKPNFREGLPDIGD |
| | | ATGGTATCGAAGACATGCCCGTGCTAGTTGTGAATGCTCAAGACATACG | AATAAT | | FTQFWANLCEKPVQ |
| | | TGAGGCAACGCAGTACACCAGGAATGGAGCTGCACCAGGACCCGATTT | AATAAT | | HNSEGMRLAEDERF |
| | | TGTATACAATTTTTGGTATAAAAAGCTAATCACAATCCATGAGCAGATAG | AATAAT | | SDGIEDMPVLVVNA |
| | | CGGCATGCTTCAATACGGTGTTGGAAGATTCGAGAAAACTACCAAAATT | ATGCAT | | QDIREATQYTRNGAA |
| | | TATCACCGGGGGAGTTACTTACTTTCTACCAAAAGATCAAAACACAAAA | AATAAT | | PGPDFVYNFWYKKLI |
| | | AATCCTGCGAAGTATAGACCACTTACCTGTCTTTCTAACTTAAACAAAGT | A (SEQ | | TIHEQIAACFNTVLED |
| | | GCTGTCGTCAGTGATAACGCAGAAAGTGAAAGATCATTGCGATACCAAC | ID NO: | | SRKLPKFITGGVTYFL |
| | | AACGTAATGACCGAAGAACAGACAGGACGTCGAAAAAACACGCAAGGC | 1374) | | PKDQNTKNPAKYRPL |
| | | TGTAAAGACCAGGTCATTATTGATGCAGTCATTGTTGGTCAAGCAGCCA | | | TCLSNLNKVLSSVITQ |
| | | AGAAACAAAGAAATCTGGATATGGCATACATCGATTACAAGAAGGCGT | | | KVKDHCDTNNVMTE |
| | | ATGATTCAGTACCCCATTCATACCTTCTTAAGGTACTCCAGTTGTACAAA | | | EQTGRRKNTQGCKD |
| | | GTAGACGGGAATGTCATCAAGCTGATGCAGCACGCGATGGGTATGTGG | | | QVIIDAVIVGQAAKK |
| | | AGTACATCTCTACACGTTACCGACGGAAAAGTTGTACTACGGTCAAGAT | | | QRNLDMAYIDYKKAY |
| | | CACTCAATATCAGGAGGGGGTATTTTCCAAGGTGACACCTTTAGTACGCT | | | DSVPHSYLLKVLQLYK |
| | | GTGGTTTTGTCTAGCTATGAACCCGCTTAGCAGAACACTCAACCAGCAA | | | VDGNVIKLMQHAM |
| | | TGCAACTTTGGGTATTTACTCAAAAGTGAAGAAATAAGCACGAGAATCA | | | GMWSTSLHVTDGKV |
| | | CCCACACCTTCTTTATGGATGACTTGAAGCTGTTCGCAGAAACAGTACA | | | VLRSRSLNIRRGIFQG |
| | | GAAGATGCACCACCTGTTGAAGAACGTGCAGGGATTCAGCAACGACAT | | | DTFSTLWFCLAMNPL |
| | | TAAAATGGAATTTGGTATCGGTAAATGTCGATCAATTCATCTACACCGA | | | SRTLNQQCNFGYLLK |
| | | GGTCAAGTATTGGATGCCGATAGCTTCCGTGCCAACGAACAAGAGGAA | | | SEEISTRITHTFFMDD |
| | | ATCCGCCACATGGTTCAAGGTGAAACTTACAAGTTCCTCGGTTTCCTGCA | | | LKLFAETVQKMHHLL |
| | | GCTGAGGGGTATTCACTATGCAGTGATCAAGAAAGAGCTACAGGACAA | | | KNVQGFSNDIKMEF |
| | | GTTCTTACATCGTGTTAGCTGTATCCTGAAGAGCTTTTTGTCAGTCGGCA | | | GIGKCRSIHLHRGQV |
| | | ACAAGGTGAAAGCAATAAACACATTTGCGGTGGCTCTGTTGACCTACAG | | | LDADSFRANEQEEIR |
| | | CTTTGGAGTAATGAAATGGTCTAATACTGACTTGGAAGCGTTGGAGCGA | | | HMVQGETYKFLGFL |
| | | ACAATTCGTGTGGTTTCCACTAAGCACCAAATGCGTCACCCAAAAGCGT | | | QLRGIHYAVIKKELQD |
| | | CCGTCGAGAGAGTAATCCTGCCACGAAAAATAGGAGGGGGTAGGAATCA | | | KFLHRVSCILKSFLSV |
| | | TTGATATTCAGGCACTTTGTATTTCTCAGATCCATCAGCTGCGAAGTTAC | | | GNKVKAINTFAVALL |
| | | TTCGTGGAAAGCCAAAACCGACATGAATTATACCGCACTGTGTATAAAG | | | TYSFGVMKWSNTDL |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | CAGATCACGGATTAAGCGCCCTGCATCTAGCGCAGCAAGATTACCAGCT | | | | EALERTIRVVSTKHQ |
| | | | GAATTGCAACATAAAAACCGTCGATGGAAAAGGCGCAACGTGGAAACA | | | | MRHPKASVERVILPR |
| | | | GAAGGAGTTACATGGGACGCACACCCATCAACTGAATCTGGAACATATC | | | | KIGGVGIIDIQALCISQ |
| | | | GACAAAGTGTCATCTAGCACTTGGCTTGTGAGGTGTGACCTTTTCTGTG | | | | IHQLRSYFVESQNRH |
| | | | AGACAGAAGGTTTCATGGTAGCCATCCAAGACCGGGTAATTGCGACGT | | | | ELYRTVYKADHGLSA |
| | | | GGAACTATCGGCGGTGTATATTGCGTGAAGACGTGGAGGACCGATGCA | | | | LHLAQQDYQLNCNIK |
| | | | GAAAGTGCAACTCAGGAGGAGAATCGATTGAGCATGTCATTGCCGGCT | | | | TVDGKGATWKQKEL |
| | | | GTCCAGTGCTAGCTGGGTCAGCGTATCTCGATCGCCACAACGACGTTGC | | | | HGTHTHQLNLEHIDK |
| | | | CAAAATTGTTCACCAGCAGCTTGCACTGAGGCACAAGTTGGTAGAGCGA | | | | VSSSTWLVRCDLFCE |
| | | | TTTTTACCCTGCTACCGATACCTCCCAGATCCGGTCCAGGAAAATGATTG | | | | TEGFMVAIQDRVIAT |
| | | | CATAAAGCTGTATTGGGATCGCGAAATTATAACGGACATCCTCATCCGT | | | | WNYRRCILREDVEDR |
| | | | GCCAATAGGCCAGACATCTTAGTCTACGAGAAAAGAAAGAAACGAGCG | | | | CRKCNSGGESIEHVIA |
| | | | ACCATCGACATCGACATTGCTGTAACGTTAGACCATAATGTTCAGACAAC | | | | GCPVLAGSAYLDRHN |
| | | | ATTTTCCACCAAGGTGATGAAGTATCATGATCTGGCAGAGGAGTTGAAG | | | | DVAKIVHQQLALRHK |
| | | | CAGACGTGGTATCTGGAGGATATCCGCATTGTTCCGGTAATCATCTCGG | | | | LVERFLPCYRYLPDPV |
| | | | CGACCGGAATTGTACCTATGGCCCTCTTACGTTCCCTGGACGAGCTCGA | | | | QENDCIKLYWDREIIT |
| | | | ACTGCAGAGAGAACTACCCAGGATTCAGAAGGCGGTGATTCTTCGAAC | | | | DILIRANRPDILVYEKR |
| | | | ATGTAGCACTTTAAGAAGGTTCCTGAATCCCTATAACTAACATCCGGTGC | | | | KKRATIDIDIAVTLDH |
| | | | AAACTCATTAACATTAAGAAAAGAGAGGAGAAATGAGAATGAGATT | | | | NVQTTFSTKVMKYH |
| | | | CATTCACCTTTGGCATTTGAATAGCCCGGGGTAGGTGAAAAGTTCCCAG | | | | DLAEELKQTWYLEDI |
| | | | CATATTGCTGAGAAGTGACAAAATTCGGATAATAATAATAATAATAATA | | | | RIVPVIISATGIVPMA |
| | | | ATAATAATAATAATAATAATAATAATAATATGCATAATAATA (SEQ ID NO: 1128) | | | | LLRSLDELELQRELPRI QKAVILRTCSTLRRFL NPYN (SEQ ID NO: 1496) |
| R4 | R4-1_BX | CADV010081 75 | Bursaphelenchus xylophilus | GGGATCCTGGGTTCCTACTACCTCGCTCCACCTCCTCGCGATGGATCCTG | GGGATCCT | TAAGAA | MTCNNAVVFPPADG |
| | | | | GGGAAGTCTCCGGACTGAGCTAAGAGAGCGTTAAAGTAGAGGGTGAC | GGGTTCCT | AAGCAT | NPAGTADRNFAIRFP |
| | | | | GGCGTAAGTACCTCCAAGTTGCGGTGGAGCGGAACATCTACTCTTCGGA | ACTACCTC | GAAATA | SSEPPGPSGIRPSEPL |
| | | | | GAGAGGGGAAGCTCTATGGCGGCGTTAGAAAGGTTGGACTACGGCAA | GCTCCACC | ATAAGA | DGRTGIGDVEHAQA |
| | | | | CGCCAGGGAGATGGGGAAGGTTCATCAGGTGATACTAGTTCGCTACTG | TCCTCGCG | AATCAG | GNGGFLVDVLEYKEA |
| | | | | TCATTCGATGTATCCGGAACCATACTCGCCAAGTTGTGAACTATGTGAA | ATGGATCC | ATAAGA | HRYGSKCEFCYVQTK |
| | | | | AGTCTGGATCCAATCCAAGACCACGGGGCGCAATTAAAAGGTGTGAAG | TGGGGAA | ATAACA | GTVCSKPRTDAWLK |
| | | | | CAGCTTGCTGGTGATCACCACGGTGGTACCTACACCGGCGGGGGAACA | GTCTCCGG | AGAATA | CEILFLLHHAYTANQ |
| | | | | TCTTTAATGCCGAGATGACCGCGCGAATAATAGGGAGCCGGCAAAAGC | ACTGAGCT | CTAATA | NKSIELAESAFRRAGI |
| | | | | CGAGCGTAAGTGGAAAGGATACAGAAATTTGCTCAAAAGTATACCAAC | AAGAGAG | AGTATA | TRRSKATIAKRWSLIQ |
| | | | | CCGCCACTTATAAAAGACCGAGAAAAGGGTACCACTAGAGCTCATTATA | CGTTAAAG | TCATGTA | RGKGTDYKEYWDEY |
| | | | | AAACATCTATGACATGTAATAATGCTGTGGTCTTTCCGCCAGCCGATGGT | TAGAGGGT | ACTATG | FEKFRYECNPTPIVRR |
| | | | | AACCCCGCTGGGACTGCGGATAGGAATTTTGCTATTAGGTTCCCCTCCA | GACGGCGT | ACAAAA | KRNRLAAGLQSPSSV |

| | | | | | |
|---|---|---|---|---|---|
| | | GTGAACCTCCAGGACCCTCTGGAATACGACCTTCTGAGCCTCTGGACGG | AAGTACCT | AGAACG | PNGYEFERKRTCETPL |
| | | TAGAACCGGGATTGGCGACGTTGAGCACGCCCAGGCGGGAAATGGAG | 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 |
| | | AACCTATCCCCTGACGAAAGGATGTGGATTGGAACAACAGTCGAGCGC | 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 |
| | ACTGTTTGTCGTGGGTGACCAGTATAATAGTACGTCATATCAAGACGATC | 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 |
| | TAAGATCGCGCCGTTGGTGAGGTTCGTTGTGCAGAGTATGCTGGTCACG | 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 | EVAVSWRGRLQEMV |
| | GCGGCATGACTGTGTTGCAAGAAATGTTTACGCATACCTCGCCATAAGG | | NO: | DRKVYKYTVNGEYEA |
| | TATGGAATCCCTGTGCCACACTATACGCAGCGAGTCAAGACAATAGAGA | | 1375) | DGSSRGWNIVRELN |
| | AAAACGGCGATCAGTCTGTAGAGCTATACTGGAATTACAAGTTCCCATG | | | DQYGFPVEVYTLVIG |
| | CACCCGTGCACTGGAAGCTTGCCGTCCGGATATTGTTCTCATAGATAAG | | | AGGEILPCTVKDVERL |
| | GTTAGTAAGCGAACTCATATAATTGAGGTGGCTGTCTCATGGCGGGGAC | | | TGGAATDNLIERMER |
| | GGTTACAGGAAATGGTCGATAGGAAAGTGTACAAGTATACGGTAAACG | | | SAVLGSCRIIKRHLAL |
| | GAGAATACGAAGCCGATGGCTCTTCCCGTGGGTGGAATATCGTCCGTG | | | (SEQ ID NO: 1497) |
| | AACTAAACGATCAGTACGGCTTTCCTGTCGAGGTATATACGCTAGTGAT | | | |
| | TGGTGCTGGAGGCGAAATCCTACCATGCACGGTTAAAGATGTAGAACG | | | |
| | GCTCACCGGTGGAGCGGCTACTGATAACCTCATTGAGCGTATGGAGCG | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | GAGCGCAGTCCTTGGCTCATGTCGTATCATTAAAAGGCATCTGGCACTC<br>TAAGAAAAGCATGAAATAATAAGAAATCAGATAAGAATAACAAGAATA<br>CTAATAAGTATATCATGTAACTATGACAAAAAGAACGCACCAATAAGAA<br>CATGCTTGAGTGGCCAGCTCTGCAGGCAAAAGTCGAATTTGGAACAGCC<br>GGTAATGGAAGACCTGCAACAAACGTGGGGTAGCAGGCAATATGTAAC<br>TATGACAGACCAAAACTCCGAAACTCTGGTAATGAGCCCGTGCCCCCCA<br>AGCATGTGGTCTCGTTCGATGTAGTTAGGAACAGTTCTCTTTAACCCGTG<br>ATGATTACGCCCTGTCTTAAATGGCAGGTGCCACCAAATACCGAACACT<br>CGTTGAGGTATGGTGGTCCGAATGTGAAGCTGGGAGTACAATTTGGTA<br>CGAGAGCACCAGCGCCCCCGATCTAAGTGATGACGCATGCGTCGGAAC<br>AATGAAGACGGCTGGCAAACATTCAGGAGTCGCAAA (SEQ ID NO: 1129) | | | | MISCDLERETLTQMA<br>LFRARSDKTPTHAGIP<br>APDEVREGGCGQNR<br>TNPAAPRGKAAAIQR<br>QNGITIPIXACAQSGL<br>VRTQRVQQWSAVEE<br>SALKDVVVRNTDDR<br>GLINWAKGVLPEWQ<br>RLCQLNPTMYMARS<br>SPSLSNKWASLRRTH<br>VGPGCPSKEGSGPSQ<br>DLSDVKIQPARLAHD<br>TVAELPQRTVPCGTD<br>GHGVIDSDETETALA<br>EVSRSSPFGEREPLDL<br>GATERITRKRLRNAV<br>RDVVPPRKRRVPSTP<br>SRKEQDLVPEVDGPA<br>PTDVLTHPPTESEPEP<br>MLDPLSLVQLVRPQL<br>GRAMGWAAEEMEL<br>GNVVMDVELKREFN<br>REVRRVGRTPPDQM<br>YKRGAGPPLPQKREP<br>ERVALLEQLIAARVER<br>GINRGLDWFLELNVA |
| R4 | R4-1_HG | . | Heterodera glycines | TGGCGATACTCGGAACCTCCGGGGAGCCTGGTAGGAGTTGGCCTACAG<br>GTCGCGAAAGTCCCTAGGTGCTGCACGGGTTGCGCTAATCCGAGGGCG<br>CTGGGTTACCTTCCCATCGGCCAAAAACGTCTGGGCCTTCTTAGCCGCG<br>GGTCCAGTATTTCTGTTGAGCCTGACAGTTCTTCCCGGATATGGCGAAA<br>GATTACAGGGCGGTATTTCGTGAAACCTAAAAAWGGTCGGGCCGAATG<br>GCACGGACAGACTCACTTCGGAGTGAGCTCGGGGGCATCCGTGTGTTA<br>CCCCGCTGCACCACGCCGAAGCTGTCATAGCGAGCCCGAAGGGGAATG<br>GCCATGGAGACTCCAGCCTCACCCTGTAACTCGAACCTAAGTCCAGGCC<br>CCTTTCTGGTGTTGGCTCGCACTGGTTAGGAACACGACAGTCTCGTGTA<br>ATCCCACACGCACCGAAAGCCCTAGCCCTCGTAGGCGAAGGCTGCGTTG<br>CTGGTTTCAGAACTGTGAGCWAGTGGGGTTCGGATGGCCGAGTGTACC<br>AACCTGCTTGCTAACAGGTAGCATAGAGTAATATGCTAGTAAGCAGGGC<br>ACGAGAAAGGGCCGTAAAGGCTCCGGCTATGAAGGACCTTGCGACCAC<br>GCGTGTGTCTCCCACGTGCGGATTCTTGAAGCCAGAGTCTTGCACTGCG<br>CGCAGGATGGAGCCTGTGCAACTCCTCCCTCGCTGATCGCAGGAGTGG<br>AGGATCACCACTCTTTTTACCTTGCTAGCTTGGGGTACCACCTTGAGCTG<br>GGGCCGGCCTTGCTAGCTTGGGGTACGACCCTTTGAGCTGGGGCTGGC<br>CTTACTAGCTTGGCGCGCCACCCTTGGAGCTGGGGTGGCGCAAGATCAC<br>TTGTATACGGTCTAACCAATACATTGAAAAGCGATCAAAGCGAAATGA<br>TATCCTGTGATTTGGAAAGAGAAACCTTAACCCAAATGGCTCTCTTTAGA<br>GCCCGCAGTGACAAGACGCCGACCCACGCCGGGATACCAGCCCCTGAT<br>GAGGTCCGAGAGGGAGGGTGCGGCCAAAACCGAACCAATCCCGCGGC<br>CCCCCGTGGTAAGGCGGCAGCGATTCAAAGGCAAAACGGCATAAACTAT<br>TCCTATAAAWGCCTGCGCTCAATCCGGCCTAGTCCGGACCCAGAGGGTT<br>CAGCAGTGGTCAGCCGTGGAGGAGTCCGCCTTAAAGGATGTGGTGGTG<br>CGAAACACAGACGACCGGGGGCTCATTAATTGGGCCAAAGGTGTGCTC | TGGCGATA<br>CTCGGAAC<br>CTCCGGGG<br>AGCCTGGT<br>AGGAGTTG<br>GCCTACAG<br>GTCGCGAA<br>AGTCCCTA<br>GGTGCTGC<br>ACGGGTTG<br>CGCTAATC<br>CGAGGGC<br>GCTGGGTT<br>ACCTTCCC<br>ATCGGCCA<br>AAAACGTC<br>TGGGCCTT<br>CTTAGCCG<br>CGGGTCCA<br>GTATTTCT<br>GTTGAGCC<br>TGACAGTT<br>CTTCCCGG<br>ATATGGCG<br>AAAGATTA<br>CAGGGCG | TGAGGA<br>CTCANA<br>ATTGAC<br>AATACA<br>CCTCAG<br>A (SEQ<br>ID NO:<br>1376) | |

| | | | | |
|---|---|---|---|---|
| | | 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 |
| | | CCGCTAGGGTTCTGAGCCGCCGGGAGAGAGTGGGAAACTCTTGCTGATC | 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 |
| | | CTGGGGAGTTGGGGTAGAGTTACAGGGCCCGCACATTGCTGCTGAACGG | GCGGATTC | VVLFEWNGVRRLSVI |
| | | GAACACAGTGGGCCCATCTCAGGCAACCAGAAAACTCACTGGAGAACT | TTGAAGCC | EVAVSDVANMQTQ |
| | | GGTCCTGAGGAGGGAAGCGGAACGGCTCAGTAGATGGAGAAGTAAAG | AGAGTCTT | HIRKSHRYGTNSTEPF |
| | | TGAAACCTGGCTGTGGCATGACTGGAGGGGCATGGCGGGATGTGCCCG | GCACTGCG | VAGVTPTYRNDCLAA |
| | | GAATCGATGTCCATCTTAGCAATAGGTGGCTCGTCAAAGGTGCACTGTC | CGCAGGAT | QLRAKFKAQQVDVIP |
| | | TCCGACCGTTGTCTCAAACAGCCTTGCCATACGAGCAAACACGGTGATT | GGAGCCTG | IIVGTTGETLDGEFGR |
| | | CTGAGAGGGAGTGGCGGGGGGCTACACCAAAGGTACCCTCCTCCGGTGC | TGCAACTC | IRKGLPMLTKLQMPR |
| | | AGAGGATGTGGTAACACGGGAGAAAACACGCAGACACATTGTGTCAGCA | 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 |
| | | | | GGGAGCAGGCTCCCCTTTCTCCGCCGGGGCGGAGGAGGCTACTGTCC | GAAATGCG | CCTCGAT | EIDSADPNRSPSRGP |
| | | | | CGACGCAACGCCGCACGCCACAAGTGGTCCAGGGAAGACCGAGTTG | 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 |
| | | GGAACTTAGTGACACCCTACTGTCGAACTACGTCCCATCTGAGATCGAC | CTGGTCAT | PQRTQRKKDPPWKH |
| | | TCGGACGACCCCAACCAGTCGCCGCGACGTGGGCCTCGGCACTTACCG | GCTGAATC | RLEKNIAEIRKHLSWI |
| | | GCCCACGATATGAGCCTGAGCGACTCGATGGACGAAGAGACACGAACCT | 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 | WMTTGKTVLIQKDK |
| | | 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 |
| | | | CCAATTCAAAGAGCGACGGCGCCTCGATCGGACTAACAAATGGGAAGAG | | | 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 |
| | | | 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 |
| | | TCCGGAAAGCCCTACCCAATTAAAAGGGGAGTGCTCCAAGGAGACAGT | 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<br>ATTGAGACCGCTAGTCACATCGTGACTGAATGTAGTACACACCGGGTTC<br>ACATGTATATCGAACGTCACGACTCTGTGGCGCGAAATATATATGCTGT<br>CCTGGCAAAAAATTGCGGTTTTTGGATTCCTCATTACTCACAGAAAATCC<br>CAACAGTAAAGATTACGAAGTCATATGAGTTGTATTGGAATTACAAGTT<br>CCCGTGCACGCAAGCACTGGAGGCTTGTCGTCCCGATATCGTTCTCATT<br>GACAGAGCGAAGAAAAGAATCCTCGTGGTGGAAGTGGCGGTCTCGTAT<br>GTCACACGGCTAGAGCAGATGACGCAGAGGAAACTCTATAAATACGGG<br>GTGAACGGCGAGTACCAGGCTGACGGGGAAACCCGTGGCTGGAACATT<br>TGTCGTGAACTTGTACAAAAGTACAATATGAGAATTGATCTCTGCATAG<br>TCGTAATTGGGGCATGTGGAGAAATACTCCCTTGCATGGTTAAAGAAAT<br>TGAAAAAATATCCAAGGTCTCTGGTCGGCAGCTCCTAGAACGGTGCCAG<br>CGTAGCGCTGTACTTGGAACTGTGCGGACCGTAAGGAGACACTTGGCA<br>AACTAACAAGTGTAATAAAAACCACCCATGCGTAGTAAACCGATCAATT<br>ATCTAGCAAAATCGCAGGTCAGAAGACCAAAGAACCGACCCAGAGGAA<br>TAGGACCAGAGCTGAAACTCTCAGATACGCCAACGGTCCTAATAAAACG<br>TCGTTAAGTAAAGCATCGTTAAGTACAAAACAAAAGCACTGTAAACGCG<br>AGGCCCCCTCTTTGCCAAACTCCGGTAATCCTCGTAGTACGGTGCTTTTT<br>CCCGCTTCACGAATCAGAACGCTGCCAGATCTTGTCCGACATGGGCCGT<br>AGGGTTTGGGAGTACCAGCGTGGGGCGGAGAGCGTACCTGGGTACACT<br>GCATAATCGGGTCTCAGAAAACCACCTATGGTTTATTATTCTGTCTCCCA<br>TCTGCAGGGTAGCTTTTCGTTGGGCCATAGGAGCCTAGGGGCAAGAGT<br>GCATGTAGTCTTCAACGGCCATGCCAGGGAAACTTGTGAGAGGTGAGG<br>GATAACTAGCATCAGATAATATCAGTCATGAAATTAGTAACAACCAACG<br>TTCACCGTCGTTGGCAAAACACCGACTAACGATGCTAGTTAGAAAGAGT<br>CGGGTCTTCCCAAAGTTAGGTGCTTGCACCGAAGCCGATCCGCTCTACC<br>CACAGCTCTGCCCAGCGTT (SEQ ID NO: 1132) | CAATGAAA<br>GCCAACCG<br>GGTTTACA<br>TTATCGGT<br>CTACCAAG<br>GGCAATGA<br>CCAGAGAA<br>ACTGCGAC<br>TATGCCGT<br>AACCCTTC<br>GCAGATTG<br>CCGATGAG<br>AACCATCG<br>ATCGTAAG<br>TCGAAGCC<br>AAGCGGAT<br>TGACCAGT<br>GGAGGGT<br>TATCCCGA<br>CAACAGC<br>(SEQ ID<br>NO: 1255) | ATAATAT<br>CAGTCA<br>TGAAAT<br>TAGTAA<br>CAACCA<br>ACGTTC<br>ACCGTC<br>GTTGGC<br>AAAACA<br>CCGACT<br>AACGAT<br>GCTAGT<br>TAGAAA<br>GAGTCG<br>GGTCTT<br>CCCAAA<br>GTTAGG<br>TGCTTG<br>CACCGA<br>AGCCGA<br>TCCGCTC<br>TACCCAC<br>AGCTCT<br>GCCCAG<br>CGTT<br>(SEQ ID<br>NO:<br>1378) | VVIGACGEILPCMVK<br>EIEKISKVSGRQLLERC<br>QRSAVLGTVRTVRRH<br>LAN (SEQ ID NO:<br>1500) |
| R4 | R4-2_HG | ABLA010003389 | Heterodera glycines | TGTGGCGATACTCGGAACCTCCGGGGAGCCTGGTAGGAGTTGGCCTAC<br>AGGTCGCGAAAGTCCCTAGGTGCTGCACGGGTTGCGCTAATCCGAGCC<br>CCTTCCGGTGTTGGCTAGCACTGGAGGGGGGGGGGTCCAAAGAATACG<br>TGAAGCGGACCCTGTGTTTTCTCTGACCTCATACGGGTATATCCCGATGA<br>TATCTATACAGTGTTCATTTTCATTTCTTAACCTGCTTTTTCCTCAAGAGA<br>AATATGACCACATTCGTTTGCCATTTCGAAGGTGCGACTTTTGAGGGAG<br>TTCTCAGGGACATCAAGCTGTTCATGGACGGCTTGACGGCTCGTGGCGG<br>CACGCAGCCGCGCGGGCCAGTTAGGGCTGGTGCCAATGAGGCAGCGA<br>AGAAGTCTCTGAAACGCCAAAAGCAGCGGGAGAAGAGAGAAAGCG<br>GGGACGGGTTAGAGATAATAGCATGATTGACGGCGAAGCAATACTTTG | TGTGGCGA<br>TACTCGGA<br>ACCTCCGG<br>GGAGCCTG<br>GTAGGAGT<br>TGGCCTAC<br>AGGTCGCG<br>AAAGTCCC<br>TAGGTGCT<br>GCACGGGT | TAGTCG<br>TAGCCC<br>AGATGT<br>CGACAG<br>CCCCTAC<br>ACCGGT<br>AGATGG<br>AAGGAG<br>GTATGT<br>ATCTAA | MRKSFLQHIPELSSHI<br>AMSVPARNYPKMCS<br>LAQGSSGTLSHNGKG<br>VAMHRCPSDDCAGK<br>DPPQRGSQKGNLRS<br>VRWTPSEEKAVFEY<br>WSRLEQHAMLNGSE<br>ARGTCAITRSQFLIH<br>WDGERESRSLSDGV<br>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 |
| | TGGTAGCGAAGCCCGGGGGGACCTGTGCCATAACTAGAAGTCAGTTTCTC | AGAGAAAT | GGTAAC | SMDREERVRKRMLP |
| | ATTCATTGGGATGGCGAGAGAGAGTCAAGATCCTTGTCGGATGGTGTG | 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 |
| | AAATCAACAACATCCGTAGACACATCGGGTGGATTACCGATGAGTTAAC | 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 |
| | TGGGTATCGGAACACCTGACCACATACTCTATTTGCCGATTGAACAAC | 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 |
| | GTTACAAGGAGACACGTTGAGCCCGCTGCTCTTTGTCTCTCAGTAGCC | CATTGCCA | 1379) | | FRVGEMIFMEPTRG |
| | GTTGTTTCTTCTGCAGTAGGCAGCCTCTTTGATTTTGAGGTCACCATACC | AAGGGGC | | | ARAIVKILRKENDSRR |
| | TGGACGAGGAGTGATGCAGCAGCAGAACCACTTGTTCTATATGGATGA | CGAAACCA | | | LAAWKGRPMGGRV |
| | CTTTAAAGGGTTTGCTCCTAGTGAAGCGTCCTTGACTAGGATGTTAGTA | TTGAGCTG | | | VSLPELDQVHSYHWL |
| | ACATTGGAGCGAACCGCTTCAGCCCTCGGGTTGAAAATCAATAAGAGG | GACGCTGA | | | IRARIGRRSFRDCIAA |
| | AAATGCGCTCTGGTACACCCCCGTGAAAGAGAGAACGAGGAAACTGGA | TGGGCCGT | | | QEGQLKARELMCPHI |
| | AGTGATATCCCGGTCCTTGGTCTACGAGACACTTATAAATACCTTGGAAT | CTGCCGCG | | | NAKAKWCRRCGDG |
| | CGAAGAGAGATTTGGTATCGTTTTCGAGGATGCGTGGGACCGAGTCCG | CAAAAGG | | | RVETEQHILSGCAWS |
| | TACCAAGATGTTTGAGAGAATGCGCACCTTGCTATGCACGGAACATACC | 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 |

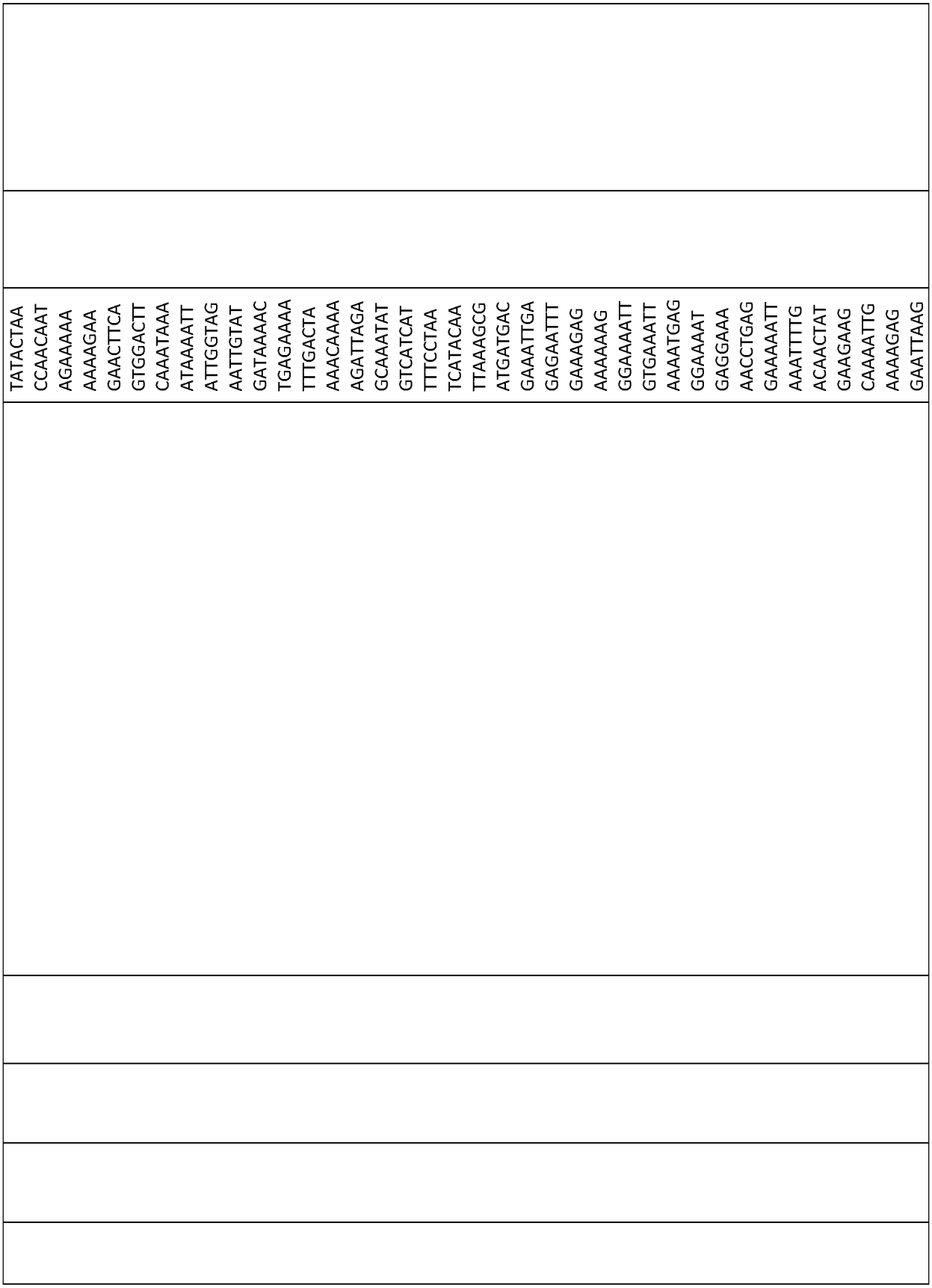
| | | | | | |
|---|---|---|---|---|---|
| | | AAAGTGTTCCGAAGTTATGCCCTTCTCTCGAAAAATTCCGCCCAGCGAGT | GATCTCAG | | GLAYGTSFPTVKVVPI |
| | | ATCTGACCTTGATGGGGATGCCTCCCCAGGATTCCGAGTAGGGGAAAT | CAATTGAT | | VVGSCGEVLPNITKRL |
| | | GATATTCATGGAACCAACTCGGGGTGCCAGAGCAATCGTAAAAATCCTC | CATATTGA | | SELGIPKRGIPSLLESI |
| | | CGTAAGGAGAATGATTCGCGACGCTTGGCTGCCTGGAAAGGTAGACCT | TCTCTTTTT | | QRAAIIGSGHVIRAHL |
| | | ATGGGAGGTCGGGTCGTGTCATTGCCAGAGCTCGACCAGGTTCATAGCT | CCAGTTGT | | SVPRSESET (SEQ ID |
| | | ACCACTGGTTGATCCGTGCGAGGATTGGCAGGCGTTCGTTCCGAGACTG | TGCGATAA | | NO: 1501) |
| | | CATTGCAGCACAGGAGGGACAGTTGAAAGCGCGGGAGCTTATGTGTCC | ATTATCGT | | |
| | | ACATATAAATGCGAAAGCAAAGTGGTGCCGACGATGTGGGGACGGTCG | GCATTATT | | |
| | | TGTTGAAACCGAGCAGCATATCCTTAGTGGTTGTGCTTGGTCACGGACT | TCTTCGAT | | |
| | | GGAACTATGTTAGATCGACATAACGGTGTTGTACGGCAAGTACACACG | TTCTCCAA | | |
| | | GCTTTATGCAGGAAATATGGACTGCCCGTTTCTTCTCATGTCGTTCCTTT | AGCTTAAC | | |
| | | GCATGCGGTAATCGAGAATGAACATGCCAAAATCCTGTACGATGTCGCG | TCCTTTCCT | | |
| | | TTGCATACTTCGCCTGCTGGCGTCCTCCCTAGGGAGGATGGGAGTACGA | AATACCTA | | |
| | | GTTATACGGGTCTGCGTAGCACCAGGCCCGATATGGTAATATTCGATAA | 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 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| R4 | R4-2_SRa | CACX010020 06 | Strongyloides ratti | CGTCAGAAGAGCAGGTGTTTTTCAAAGCAAAGACTTATTCTACGAAGGG | GGGGTGG | CGTCAGAA | TAATTA | MQKFSVPKDSSQIFL |
| | | | | GAAAGATGATCAAACATGCAGATTTGGCTGCAATGAAATAGAATCAAAC | TACCTCTG | GAGCAGG | GATTTAT | VDSILNKHICSTKNKV |
| | | | | TACCATGTGGTGACATCTTGCAAATACCATTCATACACAAGAAGACACG | GAATGATG | TGTTTTTC | ATGTGC | KDVIKRRSIIKTLICAA |
| | | | | ATATGGTTGTGTACAATATATTAAAAAGACTAAAAGAAAAGTATAAGTT | TGTCGGAT | AAAGCAAA | CACTTTA | GLTLRKLVCGKLGNN |
| | | | | AGATGGTGATTTACAATATGGTAGATCAGTATTTAATGGGAAGAAGGA | AAGCACTC | GACTTATT | CCTCAA | KYNSKINQLWKKERK |
| | | | | AAAAATTAGGATCAGATCTGGCCATAAGTTTTTGACTGAACAACCGTTA | TGCCTTAT | CTACGAAG | AGATAT | IINCIEDLKHLIETNKR |
| | | | | GTAAACAATAAACCGGATATTGTTATAACGATGAAAGAAGGGAAAAAA | AAAGCTGT | GGGAAAG | ATTAATA | RHNFGKRLRNAKVSP |
| | | | | AAGATAACGTACATACTACTAGAGATGTCGATACCACATATCTAGAACTTGA | CGTTCTGC | ATGATCAA | CTAACTT | SEMLKDYYNKLRYIK |
| | | | | AGATGCAAGAAAGAATAAAGTATAAAAAATATTGTGTGAGCTCCATGGT | CCAGGTTT | ACATGCAG | AATTATA | NEITACLDEHKKAILR |
| | | | | AAAAATAACGAATGATAATGTTGATTCGATAGCTCGTGATTTTAATCTAC | CTCCTAAC | ATTTGGCT | ATTTATT | TKFKLTPSIKIISNIQN |
| | | | | TAAATTTAATGGAAAGAAAAGAAAGGTGTAAGATAAAGTTTGGATCATT | CAGGTTGA | GCAATGAA | ATGAAT | HNDEEAELPKEEEFV |
| | | | | | ACTTGTAA | | | |
| | | | | | TGAGCCTT | | | |
| | | | | | TGGGTATC | | | |
| | | | | | TGGTCGG | | | |
| | | | | | GGTCTGCG | | | |
| | | | | | CGGAGATA | | | |
| | | | | | GCTCGTGA | | | |
| | | | | | GGAGCTGT | | | |
| | | | | | TATTACAT | | | |
| | | | | | TATTGGTT | | | |
| | | | | | GAACGCTT | | | |
| | | | | | TTGGGTCG | | | |
| | | | | | TACTGCAC | | | |
| | | | | | AGATAAAT (SEQ ID NO: 1256) | | | |

| | | | | | |
|---|---|---|---|---|---|
| | TGTATTTGGCTGTTATGGAGAATATGTCTTGACTGAATCTGCTCTAAGAA | ATAGAATC | TATAATA | KYYKELFTNKDGDDK |
| | CACAGAAAATCCTGATTGAATTGGGATTCTCCAAAAAAGAGATGGATAG | AAACTACC | TAATAA | ETPHLDNWLKKFSKT |
| | CTTGATCAAAGAGTGTTCTTACAGTTAAATGAATGAGACAGCAAGAATT | ATGTGGTG | GTTTTAA | LIVDWTINDKEILEAL |
| | ATTATGAAACATCTAGAAGGCGAAAAGAATCAAGATGAAATCAGCCATT | ACATCTTG | AATAAA | KYCGNFKAPGSDMV |
| | 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 |
| | GTTAGTTTGTGGGAAACTTGGAAATATAAATATAATTCCAAGATAAAT | GGATCAGA | | KCGIYGTDDLGKRLM |
| | CAACTTTGGAAAAAAGAAAGGAAAATTATTAATTGTATAGAAGATTTAA | TCTGGCCA | | LKESSLNFPYTSEYKYL |
| | AGCATTTGATAGAAACGAATAAACGTAGGCACAATTTTGGGAAAAGATT | TAAGTTTT | | GLVENSLDLKDINIQL |
| | AAGGAACGCCAAAGTGAGCCCTTCTGAGATGTTGAAGGATTACTACAAC | TGACTGAA | | FKDKILSKYSTIFESRL |
| | AAGTTAAGATATATTAAAAATGAAATTACGGCATGCTTGGATGAACACA | CAACCGTT | | TTHQKRKVFNSTISPC |
| | AAAAAGCAATTTTAAGAACAAAATTTAAACTAACACCAAGCATTAAGAT | AGTAAACA | | AAYYLGNLITNKCSIQ |
| | CATCTCAAATATTCAAAATCATAATGATGAAGAGGCTGAACTCCCAAAG | ATAAACCG | | ELLNECKKFDQMVR |
| | GAAGAAGAATTCGTCAAATACTATAAAGAACTATTTACAAACAAAGATG | 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 |
| | | CTGAATCCACACAAAAATGTGGCATATATGGCACAGATGATCTGGGAAAG | GATAGCTC | | STKKLENFAALNNLK |
| | | AGACTAATGCTGAAAGAAAGTTCATTAAACTTTCCATATACTTCCGAGTA | GTGATTTT | | KGENFVEILSHKANFK |
| | | CAAGTATCTTGGTCTGGTAGAAAATTCATTGGACTTAAAGGATATCAAT | AATCTACT | | RVHFMPLVFCTFGEI |
| | | ATCCAACTGTTTAAAGATAAAATACTATCTAAATACTCAACTATTTTTGAA | AAATTTAA | | PKETMKYLEKLNFSN |
| | | TCAAGACTCACTACACACCAGAAAAGAAAAGTATTTAATAGTACACAATAT | 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 | | |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| | | | | | |
| | | | | | |


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) | | | |
| R4 | R4-3_BX | CADV010088 32 | Bursaphelenchus xylophilus | GGGTCCTCGGTTCTTACTACCGTGCTCCACCTCCTCGCGATGGACCCTGG | GGGTCCTC | TGAGAC | | MISRSQADRPVEGQ |
| | | | | GGTAGGCCTCCGGGCTGAGCTAAGCAGAGCATTAAAGTAAAGAGTGAC | GGTTCTTA | CACCCAT | | PVTAMSFHNLEPNN |
| | | | | GGCGCAGTTGCTTCCAAGTTGCGGTGGGCGGAACATCTACTCTTCTGA | CTACCGTG | GCGCAG | | LYPENLRPTGSQDAN |
| | | | | GAGAGGGGAAGCCCTATGGCGGCGTTAGAAAGGTTGGACTGCGGCAA | CTCCACCT | AGTATC | | RGVADIAEEVTGPSG |
| | | | | CACTAGCCATGATCAGATTCGATCAAAATAGCCTCTGGGGCTGGCGACC | CCTCGCGA | CGAATC | | LVTNEEAARAPPLFV |
| | | | | CTATAACGGATTGTAAACTGAACTATGCTAACCTGTCAGTAAAGACAGA | TGGACCCT | AGTGAA | | EGEYKRAHCGGGKC |
| | | | | ACGGGGGCTTTGCAATCTAAGGTGTTGGCAGACCACTAAAACTGCCCTT | GGGGTAG | AGTCCA | | HYCRVLWIGARSSKA |
| | | | | TGATGAGGGTGAGATAACATCCTACGGAACAGCCCCTGCTGATCCAAAC | GCCTCCGG | AGTTTCA | | RTDSWNLCEILFLINK |
| | | | | AAAATCATCCACAAATCGGAGGGTTTTCTTGGTAGTTTCCCTTGGCACGT | GCTGAGCT | GGACAG | | CMELGNVRRIYSPLE |
| | | | | CTTGTTTCAAAAGCCAGAATATAAAACAATACCGAAGTACAGAATGGCT | AAGCAGA | AAACGT | | SSLKEAGINRTRHAIV |
| | | | | GGAACCTAACAGCCATGGCACAGAAAACCGGTGGGTGCACACCGGAGA | GCATTAAA | CAGATA | | KCRLAVMRDRFVDN |
| | | | | CAGCAGGGGAAAGATTGCAAAGTGCCGTAAATAAAGATGGTAGCTCTC | GTAAAGA | AGTCCA | | APYSEHWRLYNACA |
| | | | | TGACCTGAGGGAATGGGACACTGTCCGATTAATGCCTTATCCCGAAGAC | GTGACGGC | AGAGAA | | ENRAVVPMDSATT |
| | | | | GGTCAAGGTTCTATCTTATCGGTCTCCAAAGGGCTCTGGCCATAGAAGC | GCAGTTGC | ACGAGA | | VKKRTARQAGLESPS |
| | | | | TGCGAGGAAGCCGTAACCCTACGCAGATTGCCGCCGTGCATGATAAGTC | TTCCAAGT | AAACAA | | QIGVAGKRVHEAETG |
| | | | | GTAGCCAAGCGGATCGACCAGTGGAGGGCCAACCCGTAACAGCCATGA | TGCGGTG | GTTCAA | | TDRINAVIETNTTPLE |
| | | | | GTTTCCATAATTTAGAACCAAACAACCTGTACCCAGAAAATCTTCGTCCT | GGGCGGA | GTATGC | | DIDLSPETPEGLAELP |
| | | | | ACTGGGTCACAGGACGCGAATCGCGGAGTTGCGGACATTGCTGAGGAG | ACATCTAC | AAGAGT | | STVEIMELTEDGSRSR |
| | | | | GTGACTGGTCCTTCGGGGCTGGTTACTAACGAGGAGGCAGCGCGAGCA | TCTTCTGA | TAATCA | | GTANDADGGVSISDP |
| | | | | CCGCCGCTTTTCGTAGAGGGCGAATATAAAAGGGCACACTGCGGTGGG | GAGAGGG | ATAAGA | | LRNRPSSSQESRNVP |

| | | | | | |
|---|---|---|---|---|---|
| | 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 |
| | TCCACGAAGCTGAAACTGGAACTGATCGGATTAACGCCGGTGATTGAAAC | 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 |

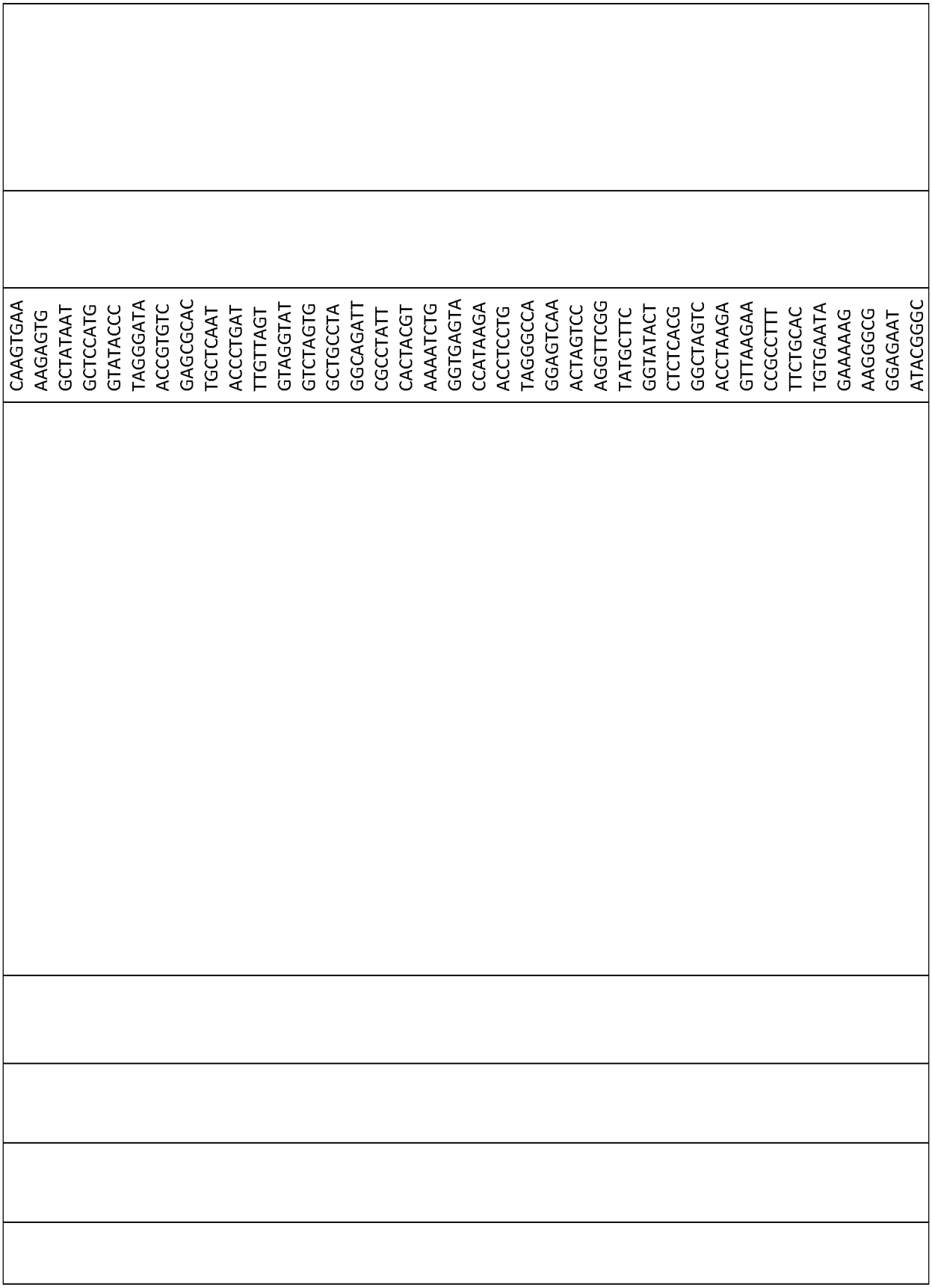
| | | | | |
|---|---|---|---|---|
| | AAACACTTGCTTTACCGGACGGAATGGAGAAATATCTCCTGAAGTCGCT | GAATATAA | GTAGGG | TMRQVVTRLNSVIGP |
| | AGAGAGTTGGAAGACAAAACTGGTGCTTAGTAATCGAGGTAAAGTTGC | AACAATAC | TGGGCT | LVRFVAQSVLTSQAK |
| | CACTGGGCGACCTTACAAGATTAAGAGAGGTGTTCTGCAAGGTGACAG | CGAAGTAC | AAAGGT | VSWIYNKISDLDSKIR |
| | 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 |
| | GAGCCGAGCCATCTATTGTCTGAAGAGCCCCGAGTCTCGTTGAGATTCGG | 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 | WNMIRELSEVYNMIK |
| | 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 | |
| | TCGTATGTCACACGGCTAAAGCAGATGGTGTCACGAAAAAGTGTATAAAT | GCCGTAAC | TTGCACC | |
| | ATGGAGTAAACGGTGAAAAGGGAGCCGACGGCGAAAGCCGCGGCTGG | CCTACGCA | GAAGCC | |
| | AATATGATCCGTGAACTATCTGAAGTTTATAATATGAAAGTGAACCTTTG | GATTGCCG | GATCCG | |
| | CGCAGTCGTGATTGGGGCGAGTGGGGAAGTTCTCCCTTGTACGGTCAA | CCGTGC | CTCTACC | |
| | AGCCATTCAGAGCATATCTAGCAAGACCTCCAGCCGACAGCTCCTAGAG | (SEQ ID | CACAGC | |
| | CGGTGCCAGCGTAGCGCTGTCCTAGGGTCGACGAGGGTTGTCAAACGA | NO: 1258) | TCTGCCC | |
| | CATCTTGCAGAATTCCATTGAGACCACCCATGCGCAGAGTATCCGAATC | | AGCGTT | |
| | AGTGAAAGTCCAAGTTTCAGGACAGAAACGTCAGATAAGTCCAAGAGA | | (SEQ ID | |

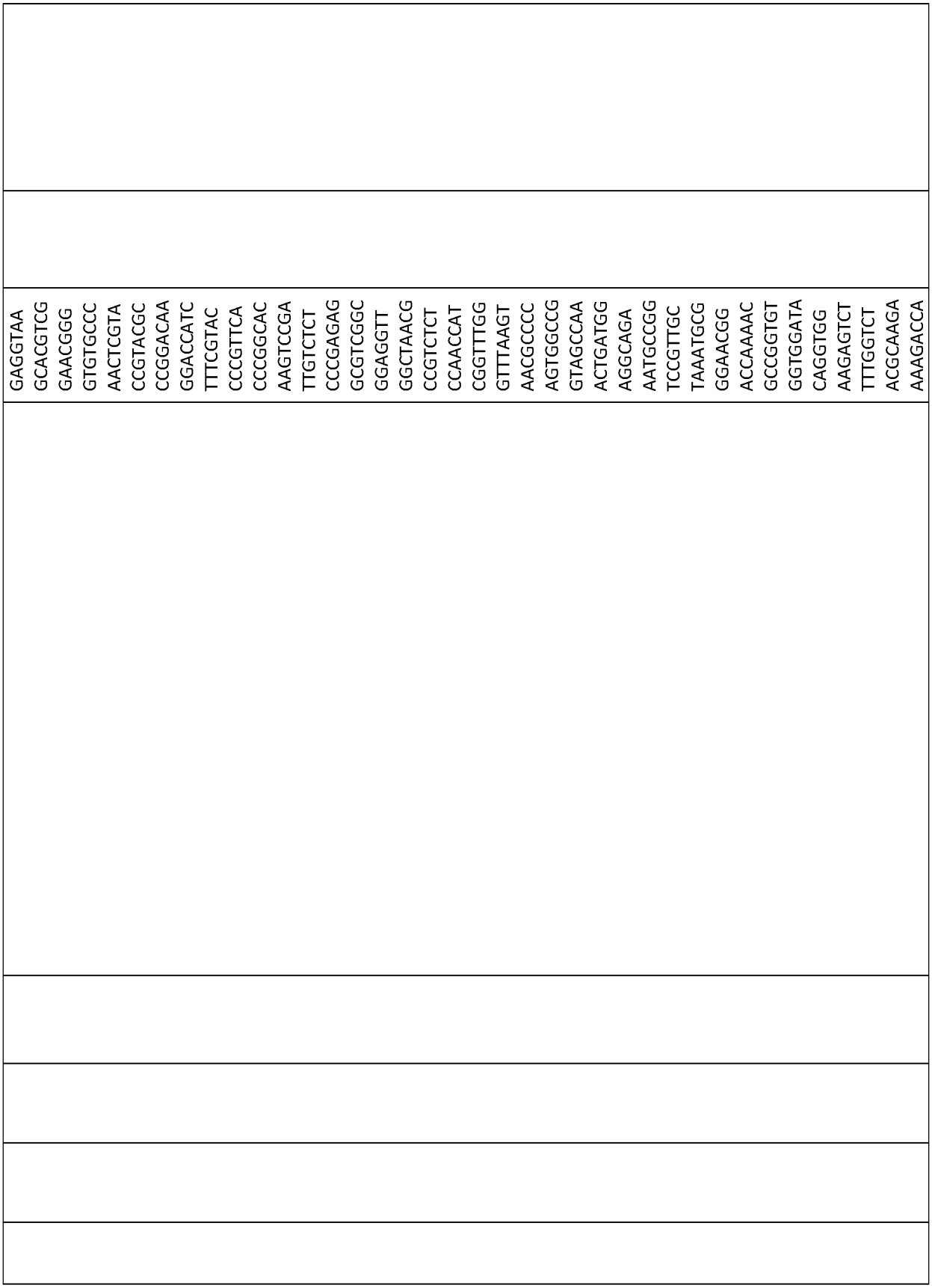
| | | | | | | |
|---|---|---|---|---|---|---|
| | | | AACGAGAAAACAAGTTCAAGTATGCAAGAGTTAATCAATAAGAGAGTA<br>CCGTAAATGTATGACCCCCCCCTTTGCCAAGTCGACAACTGTCATGCAG<br>GTGTCTCTCTTTTCACCCGCCATATGGACCAAACGCTATCCAGCCTCGCT<br>CAGAAGAGCCTTAGGGCTGGGGAGTACCACATGTGGCGGAAACTGAAT<br>CTGGATGCGATGCATACCGGGAGCGCAGCGAAATCACTTAACGCTGGT<br>GCACCTCCTGCTATCGTAGTACTTCCTAGATAGATGAGTAGGGTGGGCT<br>AAAGGTAGTCGTCTTCAATGGCAATACCGAGGGGTCTCGTGACAGGTG<br>AGGGATAACTAGTATCAGATAATATCAGTCATGAAATTAGCAACAACCA<br>ACGTCACCGTCGTTGGCAAAACACCGACTAACGATGCTAGTTAGAAAGA<br>GTCGGGTCTTCCCAAAGTTAGGTGCTTGCACCGAAGCCGATCCGCTCTA<br>CCCACAGCTCTGCCCAGCGTT (SEQ ID NO: 1135) | | | |
| R4 | R4-4_BX | . | Bursaphelenchus xylophilus | TGCCAGCGGTGTTGATTAGGTCCAAGTTCTTTGGCCAAAGATCCGCCCT<br>CGGTTTAGCAGTACCGAACGAGTATACCTTCAAGTGGTGGCACTGAATT<br>AGACTGAATACTCTGAACTGTAGACTTTTGTGCAACTGTGTATGGTGTG<br>GAAGACTTGTTTGTACCACTATCAGCTTTATTGGGGCTCGTTACTGTTTC<br>ATACAGGTAGATGTCCCCTTTAGAGATTTCCCTGCAGTTTGCGCTCCAAG<br>TCGCTAGCCTCTTGCGTGTAGTCAAAGGAATACATTCGCCGTCGGTGAC<br>AGGGCTATACCCGGCGACTACGGACTTGTTTATTACGTAGTGCAGCCTC<br>GTTTAAGACGAATGTGAAAAGAAGGTGTGATTACTAAGCGTTATGAGTC<br>GGGTTATCTGGAAACTCCGCCCCGCCGCAATGGCTTCAAGGGTCAAACT<br>GCTAACATTTAAACCAATATTTGGCAGCAGCAGACTATGATGTCAGAG<br>TGCCGGGGGTCTATCGATATAAGACTGAAAGCGATAGACGTGGAGTGA<br>AAGGATTCCCTTTGTTAAGAGAATCGGTAGAATTCACTTTTTACTTATTC<br>AAGAACTTAACAGCAACAAGCACTCGCGAGGATTACCGCCCCAGATTCG<br>GTCGGCGGTACTTCACCTGCTTTCTTCCACTTTCGGAATCTGGCATCCTG<br>GCTTTCAGTGTGGTGATGGCCGGCTTGAGTTTTCTTGAGTCGTGCGAGT<br>GCCTCATCTGGGACGTCCGGACCGATTGATGGAGTCTGCAGTGGACGA<br>GGACTTGATGGACCGTAACCATAGTATATCCCTCATGCGTCTTCTCGACT<br>CGAGGGGGTAGCTTGCACTACCCACCCTTCTCTTCCGATTGGGATTTA<br>GACCTAGCCCTCTGGTGTGTCTCGACCCGCGATATCAGATTCCTGAATCG<br>ACTGTGAGAAATGTCTACGCGCAAAGATCGACCCATTGCCACCCGGCTA<br>TGTGGATCGGGCTCTTGACTGCTTATCTCCGGCTTTAATCGCTTGAGGAA<br>AGGGGGGTGTCGCCCGAAAGGGTTACGCGATCATCCGATTCCTACCGT<br>AAAAACGTAGTTGAAGTAGGATTGAACCTGAGTACCAAGTGAAAAGAG<br>TGGCTATAATGCTCCATGGTATACCCTAGGGATAACCGTGTCGAGCGCA<br>CTGCTCAATACCCTGATTTGTTAGTGTAGGTATGTCTAGTGGCTGCCTAG<br>GCAGATTCGCCTATTCACTACGTAAAATCTGGGTGAGTACCATAAGAAC<br>CTCCTGTAGGGCCAGGAGTCAAACTAGTCCAGGTTCGGTATGCTTCGGT | TGCCAGCG<br>GTGTTGAT<br>TAGGTCCA<br>AGTTCTTT<br>GGCCAAA<br>GATCCGCC<br>CTCGGTTT<br>AGCAGTAC<br>CGAACGA<br>GTATACCT<br>TCAAGTGG<br>TGGCACTG<br>AATTAGAC<br>TGAATACT<br>CTGAACTG<br>TAGACTTT<br>TGTGCAAC<br>TGTGTATG<br>GTGTGGAA<br>GACTTGTT<br>TGTACCAC<br>TATCAGCT<br>TTATTGGG<br>GCTCGTTA<br>CTGTTTCA<br>TACAGGTA<br>GATGTCCC | TAATTA<br>GGCAGT<br>GCTCCTA<br>GCGGAG<br>TGCCGT<br>GAAGTG<br>GTGTCA<br>GTACCC<br>GTCGTG<br>TGGAAA<br>GCCCAG<br>GAGGGT<br>TAGTAC<br>CGACAG<br>TGGGAA<br>ACCCGC<br>TGCACG<br>CAACCT<br>AAAGAC<br>ATTTGCC<br>CTTCGG<br>GGGAGA<br>GGTATG<br>AGACCA<br>CACTAG<br>TCATGG<br>TTGCTTG | MEILWEDLRLKIEDR<br>YGVTLPQRSASSLKN<br>QYPKVILRGLPDSGLP<br>WAGVQVNDTGQVV<br>VVDHAEAATLRGSSP<br>AAVDGEAEEPVVPPL<br>PAAEVVESAADAAVP<br>DPQSEIVADQGVETR<br>PVENPPANSRETETE<br>PVEVEPYLEGQYKFF<br>VSKILGKSMWRKPIK<br>YPRRVPETLWRQAN<br>ELIERSIRQGEVSIQSL<br>NCMVYAAGCAVKSS<br>LDKKDQEAKRRESE<br>WYACRKAEIKALERY<br>LNFIDLELKRRSASRP<br>LTSRQRQNLGVLITKY<br>GRARVRSGVRLSELQ<br>AMLRDALVGIRKCM<br>AKRSADKKRKQGKFV<br>PIQRYLEPSSAEPRLS<br>PDTVRAYWNDIVGS<br>SQQSTSDSTIQDWSS<br>NLSVPSQELNASKIM<br>GWWRAAVSKSKPN<br>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 | | KTLNPGPTLRPNMG |
| | 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 |
| | CTCGGAATTGCAGGCAATGTTACGAGACGCCTTAGTGGGAATCCGAAA | 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 |
| | | GGCGACACTCTCTCCTCACATGGTTTTGCCTATGCATGGCTCCAATATC | TTTTTACTT | | VNSNHEFSESIPYPPG |
| | | AGCCTCTATTAAAACCCTTAACCCTGGCCCAACCCTTCGGCCGAACATGG | ATTCAAGA | | VNLANEIRVLYPQFT |
| | | GGAATGGGCGAAATAGAGGGCAGGTGGCGATCCAAGTTAGTCATGTGT | ACTTAACA | | GGVKVFPMIISPTGE |
| | | TCTACATGGATGACTTGAAAGTGTATTGCCCCCCGCGTAGCCGATCAACG | GCAACAAG | | VHMQFVPHLAELLE |
| | | CAGAATGGAACAGAAATATTCCACAGTTGTTCGGTGAAATTGGACTAAGC | CACTCGCG | | NPNIPRILEKIQRSVVL |
| | | ATAAATGCGAGTAAGAGCGCAGCAGCCGCAGCCGTGGGAAGATACGTT | AGGATTAC | | GTDYIIRSYFAM (SEQ |
| | | GAATCCGAGTTGCCAGTTCTTGGGACGAAGGATGAGTACAAGTACCTG | CGCCCCAG | | ID NO: 1504) |
| | | GGAATAGAAAGCGGTTTTGTTGTAAATGAGGTGGCCGCCCTTGACAGA | ATTCGGTC | | |
| | | ATGCAGGCCGTTCTGCTAAACCGTGTAGAAGCGATTCTATCGGTTAAAG | GGCGGTAC | | |
| | | AACATACTGTTGGTCAGAGGCGCGACGCTATACGAGCGAAAGCGATCC | TTCACCTG | | |
| | | CTGGGGGGCCTATATCCTGGGTCACATCATTTTGAGTGACTTGGACCC | 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 | | |
| | | ATCATTACTGGCTGCGCCGAATTGCGCAAAAGCCATATGAAGGTGAGG | 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 |
| | | | | CGTTAGCCACGAAGTGGGGGGCTCTTAAACGCCCGAGAGCTCCGATGG | 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 | GCCCTA | GWYTKTKATREDLQ |
| | TGTTGACTTTGAAGGGGAACCGAGGTTTCGTGCCTTTCGGAAAGCCTTT | AGCTGGG | CTAGCC | CTAGCC | MLISMMESELARRKE |
| | TACGGGTATTTCAGATGGGCTGTCAACTCGTTTGATAGGGAACCTGTCA | AATATGGG | AAACGC | AAACGC | KRKPGAKELENIHKLV |
| | AGCGAGTTAGGCGGGACTGTCCCAAGGTATTTTACGCTTATGCGGATTA | TTAGGCGT | (SEQ ID | (SEQ ID | ARLGTRSTSGIVRRLE |
| | CCTTATCGCAACCGGAAGTTCTAAGGCGTTGGGACCGAATCAATCCAGG | CCTCTGAC | NO: | NO: | MTRQRLKLLEDRISLH |
| | ATTGGTCGCTTGAACGGATTGGTCTATGCGGCAGCTAGAACAATCCACC | ATATAAGA | 1383) | 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 |
| | GTGATAGGGGCGACCCTTCTAATTACCGACCTATAACATGTTTAAATACA | 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) |
| | | | TGATCGGAACGGCTAGAGTAGTCCGAGCACACACCTCGCCTGCTCCTAGTC | | | | |
| | | | GCTAAGGGGTCCGGAAATGGTCCGGTCCTGCGCTACCCGGTTCTGGTA | | | | |
| | | | GCACGTTCAAGCGCTCAATCGCCTGCCTTGTAGGCAGTCCATCTGTGGA | | | | |
| | | | AGTCGCGCTCTTGATACAGATGTGGACGGATGGAAGCAGATGATAGAG | | | | |
| | | | CCGGTGACGGCCCTACTAGCCAAACGC (SEQ ID NO: 1137) | ATAAATAA | TAACTTA | | ITYTANMALVTLFME |
| R4 | R4_Hmel | . | ATAAATAATAATAATAATAATAAGCCCCCTAAAATCCAACCATA | ATAAATAA | TAACTTA | | ITYTANMALVTLFME |
| | | Heliconius melpomene | 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 |
| | ATCTTATTATAACAAAATGCGAAACCACGTTATTGAATAATTATTTAGAA | 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 |
| | AAGTAATTTACAACCCGTTGAACCACTAGACCAACATTGTATAAAATAGAC | | 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 |
| | | GTATTACTGACATAATTTATAAACATATCGATCAAAATAACATATTAGCT | | | 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<br>TAGTTCTTTCAACAACCGGAGTAGTACCTCATACGCTACACACGAGTCTG<br>GAAACTTTGGGTATTCACCGGTTAAGTTATATTCTTTTACAGAAAGCCGC<br>AATCTTAAATACATGCCGTATAGTTAGGAAGTTTTTATCATCAAATAATT<br>AACTTATTGTCAGAATTCCTTACTAGTAATAATAATTATCGCTGAAAATC<br>TCCACCCAAATATTGCTTGGCTATATGCTCGCAATTTTTTGGTTAACGTAC<br>CCCAATGATTTGGGAGAACAAAAATGGTAAAACTATAATAATAATAATT<br>ATATTAATAT (SEQ ID NO: 1138) | TTCTATGC<br>GCCTTATG<br>CGACTGGA<br>TAGGCCAG<br>TGGTTTAC<br>GCCGCTGA<br>CTTTGGTG<br>CGGAAGG<br>TTGTCGGT<br>TCGAATCC<br>AGGCGAG<br>CCCTTAGG<br>CAAGGCTC<br>CTTACGCA<br>TATATGCC<br>TACACCTC<br>GGT (SEQ<br>ID NO:<br>1262) | TAGAGG<br>ACCCGA<br>GTCTGA<br>AGGAAG<br>GAGGCA<br>CCGCCC<br>AGGAGG<br>GCGAGG<br>AAGAGA<br>TTTTTTT<br>TTATATA<br>TATATAT<br>ATATATA<br>TA (SEQ<br>ID NO:<br>1385) | DKTVHYNRPDILLHD<br>KVKXSVYLIDIAIPNTH<br>NLASTFSNKIDKYTDL<br>TIELKSQWKVQSVTT<br>VPIVLSTTGVVPHTLH<br>TSLETLGIHRLSYILLQ<br>KAAILNTCRIVRKFLSS<br>NN (SEQ ID NO:<br>1506) |
| R4 | Rex6 | Takifugu rubripes | MSGTXTDRVIPARTS<br>PGSTRSASGVGEPGP<br>PDVKLATGTRHSWS<br>RAENVVLMECYYGS<br>NPSERGYMQRMWE<br>KWVLRNPTSSLTKKQ<br>LLAQCSNIRNKKLLSQ<br>LEIDEARRCASPTVQI<br>CYGKGEPGRQVSXG<br>VISSSPPNIEIGYKAP<br>MTDGLGTRAADLRE<br>RIMKSWGNSTTSLPR<br>LTHKVPDQSLLEDM<br>NTALSTIPTTTITETN<br>QLMYAAATVILQML<br>GYKMKSMNSQKEQ<br>MAPWRRRLEAKIMA<br>TRREVSLLTELSRGVN<br>LRTEXPKKYNKLSTTE<br>ALETAKQRLTALATRL<br>KRYTREVEARRINKVF<br>STNPAKVYSQWQGN<br>KMTTDPPRAETEQY<br>WKSIWEKEATHNTX<br>AQWLQDLQTEHSQL<br>PEQDPVVITLADIQTR<br>VSKMKSWTAPGPDK<br>IHAYWLKKLTALHERL<br>AAQMNQLLTSGNHP | TTCTATGCCTTATGCGACTGGATAGGCCAGTGGTTTACGCCGCTGAC<br>TTTGGTGCGGAAGGTTGTCGGTTCGAATCCAGGCGAGCCCTTAGGCAA<br>GGCTCCTTACGCATATATGCCTACACCTCGGTATGAGCGGAACMWTAA<br>CTGATAGAGTCATACCGGCTCGGACGTCGCCCGGGTCAACAAGGTCCG<br>CGTCAGGGGTCGGGGAACCAGGGCCACCTGATGTCAAATTGGCTACTG<br>GAACAAGGCATTCATGGAGCAGGGCAGAGAACGTGGTACTGATGGAG<br>TGTTACTACGGGAGTAACCCCAGCGAGAGAGGGTACATGCAGAGGATG<br>TGGGAGAAATGGGGTGCTTCGAAACCCCACATCCTCACTGACTAAGAAGC<br>AGCTCTTAGCTCAGTGTTCGAATATCCGCAATAAGAAGCTGCTATCACA<br>GCTAGAGATTGACGAGGCACGGCGNTGTGCCTCACCCACGGTACAAAT<br>ATGCTACGGCAAGGGGGAGCCAGGACGCCAGGTCAGCNTGGGGGTGA<br>TTTCATCATCACCCCCCAATATCGAGATTGGGTACAAAGCCCCAATGACA<br>GACGGACTCGGCACAAGGGCAGCTGACCTGAGAGAGAGAATCATGAA<br>GTCCTGGGGGAACTCAACTACCTCCCTGCCAAGGCTAACACACAAAGTA<br>CCAGACCAATCTCTACTGGAGGACATGAACACGGCACTGTCAACCATCC<br>CTACCACTACCATCACTGAGACCAATCAGCTGATGTATGCAGCGGCAAC<br>AGTAATCCTACAGATGCTTGGCTATAAGATGAAGAGCATGAACAGCCAG<br>AAGGAGCAAATGGCCCCATGGAGGAGAAGGCTAGAGGCAAAAATCAT<br>GGCGACACGGAGAGAAGTCAGCCTTCTAACAGAGCTGAGTAGAGGCGT<br>GAATCTAAGGACAGAGCNGCCCAAGAAATATAACAAACTGTCCACAACT<br>GAGGCACTGGAGACTGCTAAGCAAAGGCTCACAGCCCTGGCTACCCGA<br>CTAAAGAGGTACACAAGAGAAGTAGAGGCAAGGAGAATAAACAAGGT<br>GTTCTCCACCAATCCAGCNAAGGTCTACTCTCAATGGCAGGGCAACAAG<br>ATGACAACAGACCCCCCCAGGGCTGAGACTGAGCAATACTGGAAGAGT<br>ATCTGGGAGAAAGAGGCAACGCACACAACTANTGCCCAATGGCTGCAA<br>GACCTACAGACAGAGCACAGCCAACTTCCAGAACAAGACCCAGTAGTCA<br>TCACCTTAGCAGACATCCAAACAAGAGTGTCCAAAATGAAGAGCTGGAC<br>AGCACCAGGGCCCGATAAGATCCACGCCTACTGGCTTAAGAAGCTGACT<br>GCACTCCATGAACGCCTGGCAGCACAGATGAACCAGCTGCTAACATCAG | | | |

| | | | | |
|---|---|---|---|---|
| | | 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 |
| | | GGCTATCATGGAAGGACAAGCCCCTGCATGGNATGTACCACCGACAAA | | RLLSECLRQQKPTKEE |
| | | TTGAGGAAGTGGCTGATATCGAGAAAACATACCAGTGGCTGGAAAAGG | | EPEGLSWKDKPLHG |
| | | CCGGACTGAAAGACAGCACAGAGGCACTACTCATGGCTGCACAAGAAC | | MYHRQIEEVADIEKT |
| | | AGGCCCTGAGCACCAGAGCAATAGAGGCCAGGGTCTACCATACCAGAC | | YQWLEKAGLKDSTEA |
| | | AAGACCCCAGGTGCAGGCTGTGTGGAGACGCCCCTGAGACAGTCCAGC | | LLMAAQEQALSTRAI |
| | | ACATCACAGCAGGGTGCAAGATGTTGGCAGGCAAGGCATACATGGAGC | | EARVYHTRQDPRCRL |
| | | GGCATAACCAGGTGGCTGGCATAGTGTACAGGAACATCTGCACTGAGT | | CGDAPETVQHITAGC |
| | | ATGGACTGGAGGTCCCAGGGTCCAGGTGGGAGACACCCCCGAAAGTGC | | KMLAGKAYMERHN |
| | | TGGAGAACAAGCAGGCCAAGATCCTGTGGGACTTCCAGATCCAGACTG | | QVAGIVYRNICTEYGL |
| | | ACAAGATGGTGGTGGCCAACCAGCCTGACATAGTGGTGGTGGATAAAC | | EVPGSRWETPPKVLE |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | NKQAKILWDFQIQTD<br>KMVVANQPDIVVVD<br>KHQKTVVVIDVAIPS<br>DSNIRKKEHEKLEKYQ<br>GLKEEMERMWGMK<br>ATVVPVVIGTLGAVT<br>PKLSRWLQQIPGTTS<br>EISVQKSAVLGTAKIL<br>RRTLRLPGLW (SEQ<br>ID NO: 1507) |
| | | | ACCAGAAGACAGTGGTGGTGATAGATGTAGCAATCCCAAGTGATAGCA<br>ACATCAGGAAGAAGGAACACGAGAAGCTGGAGAAGTACCAAGGGCTG<br>AAGGAGGAGATGGAGAGAATGTGGGGCATGAAGGCAACAGTGGTCCC<br>AGTAGTGATTGGGACACTAGGGGCAGTAACACCCAAGCTGAGTAGATG<br>GCTCCAACAGATACCAGGAACCACATCAGAGATCTCTGTCCAGAAGAGC<br>GCAGTCCTAGGAACAGCTAAGATCCTGCGCAGAACCCTCAGACTCCCAG<br>GCCTCTGGTAGAGGACCCGAGTCTGAAGGAAGGAGGCACCGCCCAGG<br>AGGGCGAGGAAGAGATTTTTTTTTATATATATATATATATATA (SEQ<br>ID NO: 1139) | | |
| | | | | | |

*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.

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, 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).

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.

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 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. 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 GeneWriter polypeptide, wherein expression of the GeneWriter 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, the object sequence may contain a non-coding sequence. For example, the template RNA may comprise a promoter or enhancer sequence. 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 GeneWriter 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 GeneWriter 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 GeneWriter polypeptide, driven by a tissue-specific promoter, e.g., to achieve higher levels of GeneWriter protein in target cells than in non-target 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, or ROSA26. In some embodiment 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 4 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 4. 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 4. 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 4. 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 4. In some embodiments, the heterologous object sequence is inserted within a gene indicated in Column 5 of Table 4, 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 4.** 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.

| 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 | CGCATGAATGGATGAACGAG ATTCCCACTGTCCCTACCTACT ATCCAGCGAAACCACAGCCA AGGGAACGGGCTTGGCGGA ATCAGCGGGGAAAGAAGACC CTGTTGAGCTTGACTCTAGTC TGGCACGGTGAAGAGACATG AGAGGTGTAGAATAAGTGGG AGGCCCCCGGCGCCCCCCCG GTGTCCCCGCGAGGGGCCCG | RNA28SN1 | 106632264 |

| | | | | | |
|---|---|---|---|---|---|
| | | TATGACTCTCTTAAGGTAGCC AAATGCCTCGTCATCTAATTA GTGACG (SEQ ID NO: 1509) | GGGCGGGGTCCGCCGGCCCT GCGGGCCGCCGGTGAAATAC CACTACTC (SEQ ID NO: 1514) | | |
| 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 | TGAAACTTAAAGGAATTGAC GGAAGGGCACCACCAGGAGT GGAGCCTGCGGCTTAATTTG ACTCAACACGGGAAACCTCA CCCGGCCCGGACACGGACAG GATTGACAGATTGATAGCTCT TTCTCGATTCCGTGGGTGGTG GTGCATGGCCGTTCTTAGTTG GTGGAGCGATTTGTCTGGTT AATTCCGATAACGAACGAGA CTCTGGCATGCTAACTAGTTA | RNA18SN1 | 106631781 |

| | | | | | |
|---|---|---|---|---|---|
| | | TCCGGGGGGAGTATGGTTGC AAAGC (SEQ ID NO: 1512) | CGCGACCCCCGAGCGGTCGG CGTCCC (SEQ ID NO: 1517) | | |
| 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 GGATGAACGAGATTCCCACT GTCCCT (SEQ ID NO: 1518) | ACCTACTATCCAGCGAAACCA CAGCCAAGGGAACGGGCTTG GCGGAATCAGCGGGGAAAG AAGACCCTGTTGAGCTTGACT CTAGTCTGGCACGGTGAAGA GACATGAGAGGTGTAGAATA AGTGGGAGGCCCCCGGCGCC CCCCCGGTGTCCCCGCGAGG GGCCCGGGGCGGGGTCCGCC GGCCCTGCGGGCCGCCGGTG AAATACCACTACTCTGATCGT TTTTTCACTGACCCGGTGAGG CGGGGGG (SEQ ID NO: 1524) | RNA28SN1 | 106632264 |
| 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 | GGAGCCTGCGGCTTAATTTG ACTCAACACGGGAAACCTCA CCCGGCCCGGACACGGACAG GATTGACAGATTGATAGCTCT TTCTCGATTCCGTGGGTGGTG GTGCATGGCCGTTCTTAGTTG GTGGAGCGATTTGTCTGGTT AATTCCGATAACGAACGAGA CTCTGGCATGCTAACTAGTTA | RNA18SN1 | 106631781 |

| | | | | | |
|---|---|---|---|---|---|
| | | GTTCCGGGGGGAGTATGGTT GCAAAGCTGAAACTTAAAGG AATTGACGGAAGGGCACCAC CAGGAGT (SEQ ID NO: 1520) | CGCGACCCCCGAGCGGTCGG CGTCCCCCAACTTCTTAGAGG GACAAGTGGCGTTCAGCCAC CCGAG (SEQ ID NO: 1526) | | |
| 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%.

*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).

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).

*Applications*

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 still other embodiments, a promotor can be operably linked to a coding sequence.

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.

*Administration*

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 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. In some embodiments, the cell is an HSC and p53 is not upregulated or is upregulated by less than 10%, 5%, 2%, or 1%, e.g., as determined according to the method described in Example 30. 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 August 27, 2018. 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: Delivery of a Gene Writer™ system to mammalian cells

This example describes a Gene Writer™ genome editing system delivered to a mammalian cell for site-specific insertion of exogenous DNA into a mammalian cell genome.

In this example, the polypeptide component of the Gene Writer™ system is the R2Bm protein from Bombyx mori and the template RNA component is RNA for the R2Bm retrotransposase from Bombyx mori containing a mutation in the reverse transcriptase domain that renders the retrotransposase inactive.

HEK293T cells are transfected with the following test agents:

1. Scrambled RNA control
2. RNA coding for the polypeptide described above
3. Template RNA described above
4. Combination of 2 and 3

After transfection, HEK293T cells are cultured for at least 4 days and then assayed for site-specific genome editing. Genomic DNA is isolated from each group of HEK293 cells. PCR is conducted with primers that flank the R2Bm integration site in 28s rRNA genes. The PCR product is run on an agarose gel to measure the length of the amplified DNA.

A PCR product of the expected length, indicative of a successful Gene Writing™ genome editing event that inserts the sequence for the mutated R2Bm retrotransposase into the target genome, is observed only in cells that were transfected with the complete Gene Writer™ system of group 4 above.

**Example 2: Site-specific targeted delivery of a Gene Writer™ system into insect cells**

This example describes a Gene Writer™ genome editing system delivered to an insect cell at a specific target site of the genome.

In this example, the polypeptide component of the Gene Writer™ system is derived from R2Bm of Bombyx mori, which is modified by replacing its DNA binding domain in the amino

terminus of the polypeptide with a heterologous zinc-finger DNA binding domain. The zinc finger DNA binding domain is known to bind to DNA in the BmBLOS2 loci of B. mori cells (Takasu et al., insect Biochemistry and Molecular Biology 40(10): 759-765, 2010). The template RNA is RNA for the R2Bm retrotransposase from Bombyx mori containing a mutation in the reverse transcriptase domain that renders the retrotransposase inactive. Furthermore, the template RNA is modified at the 5' end to have 180 bases of homology to the target DNA site.

B. mori insect cell lines are transfected with the following test agents:

1. Scrambled RNA control
2. RNA coding for polypeptide component described above
3. Template RNA described above
4. Combination of 2 and 3

After transfection, the cells are cultured for at least 4 days and assayed for site-specific Gene Writing™ genome editing. Genomic DNA is isolated from the cells and PCR is conducted with primers that flank the target integration site in the genome. The PCR product is run on an agarose gel to measure the length of DNA. A PCR product of the expected length, indicative of a successful Gene Writing™ genome editing event that inserts the sequence for the mutated R2Bm retrotransposase into the target insect cell genome, is observed only in cells that were transfected with the complete Gene Writer™ system of group 4 above.

**Example 3: Site-specific targeted delivery of a Gene Writer™ system into mammalian cells**

This example describes a Gene Writer™ genome editor system used to insert a heterologous sequence into a specific site of the mammalian genome.

In this example, the polypeptide of the system is the R2Bm protein from Bombyx mori and the template RNA component is RNA coding for the GFP protein and flanked at the 5' end by the 5' UTR and at the 3' end by the 3' UTR of the R2Bm retrotransposase from Bombyx mori. The GFP gene has an internal ribosomal entry site upstream of its start codon and a polyA tail downstream of its stop codon.

HEK293 cells are transfected with the following test agents:

1. Scrambled RNA control
2. RNA coding for the polypeptide described above
3. Template RNA coding for GFP described above
4. Combination of 2 and 3

After transfection, HEK293 cells are cultured for at least 4 days and then assayed for a site-specific Gene Writing™ genome editing event. Genomic DNA is isolated from the HEK293 cells and PCR is conducted with primers that flank the R2Bm integration site in 28s rRNA genes. The PCR product is run on an agarose gel to measure the length of DNA. A PCR product of the expected length, indicative of a successful Gene Writing™ genome editing event, is detected in cells transfected with the test agent of group 4 (complete Gene Writer™ system). This result demonstrates that a Gene Writing genome editing system can insert a novel transgene into the mammalian cell genome.

The transfected cells are cultured for a further 10 days, and after multiple cell culture passages are assayed for GFP expression via flow cytometry. The percent of cells that are GFP positive from each cell population are calculated. GFP positive cells are detected in the population of HEK293 cells that were transfected with the test agent of group 4 (complete Gene

Writer™ system). This result demonstrates that the novel transgene written into the mammalian cell genome is expressed.

**Example 4: Targeted delivery of a gene expression unit into mammalian cells using a Gene Writer™ system**

This example describes the making and using of a Gene Writer genome editor to insert a heterologous gene expression unit into the mammalian genome.

In this example, the polypeptide of the Gene Writer system is derived from the R2Bm polypeptide of Bombyx mori as modified by replacing its DNA binding domain in the amino terminus of the polypeptide with a heterologous zinc-finger DNA binding domain. The zinc finger DNA binding domain is known to bind to DNA in the AAVS1 locus of human cells (Hockemeyer et al., Nature Biotechnology 27(9): 851-857, 2009). The template RNA comprises a gene expression unit. A gene expression unit comprises at least one regulatory sequence operably linked to at least one coding sequence. In this example, the regulatory sequences include the CMV promoter and enhancer, an enhanced translation element, and a WPRE. The coding sequence is the GFP open reading frame. The gene expression unit is flanked at the 5' end by 180 bases of homology to the target DNA site and at the 3' end by the 3' UTR of the R2Bm retrotransposase from Bombyx mori.

HEK293 cells are transfected with the following test agents:

1. Scrambled control RNA
2. RNA coding for the polypeptide component described above
3. Template RNA comprising the gene expression unit (as described above)
4. The complete Gene Writer system comprising both (2) and (3)

After transfection, HEK293 cells are cultured for at least 4 days and assayed for site-specific Gene Writing genome editing. Genomic DNA is isolated from the HEK293 cells and PCR is conducted with primers that flank the target integration site in the genome. The PCR product is run on an agarose gel to measure the length of DNA. A PCR product of the expected length, indicative of a successful Gene Writing™ genome editing event, is detected in cells transfected with the test agent of group 4 (complete Gene Writer™ system).

The transfected cells are cultured for a further 10 days, and after multiple cell culture passages are assayed for GFP expression via flow cytometry. The percent of cells that are GFP positive from each cell population are calculated. GFP positive cells are detected in the population of HEK293 cells that were transfected with group 4 test agent, demonstrating that a gene expression unit added into the mammalian cell genome via Gene Writing genome editing is expressed.

**Example 5: Targeted delivery of a gene expression unit into an intronic region of mammalian cells using a Gene Writer™ system**

This example describes the making and use of a Gene Writing genome editing system to add a heterologous sequence into an intronic region to act as a splice acceptor for an upstream exon.

The target integration site is the first intron of the albumin locus. Splicing into the first intron a new exon containing a splice acceptor site at the 5' end and a polyA tail at the 3' end will result in a mature mRNA containing the first natural exon of the albumin locus spliced to the new exon. Because the first exon of albumin is removed upon protein processing, the cell

expressing the newly formed gene unit will secrete a mature protein comprising only the new exon.

In this example, the Gene Writer genome editor polypeptide is derived from the R2Bm Gene Writer genome editor of Bombyx mori as modified by replacing the DNA binding domain in the amino terminus of the polypeptide with a heterologous zinc-finger DNA binding domain. The zinc finger DNA binding domain is known to bind tightly to the albumin locus in the first intron as described in Sarma et al., Blood 126, 15: 1777-1784, 2015. The template RNA is RNA coding for EPO with a splice acceptor site immediately 5' to the first amino acid of mature EPO (the start codon and signal peptide is removed) and a 3' polyA tail downstream of the stop codon. The EPO RNA is further flanked at the 5' end by 180 bases of homology to target DNA site and at the 3' end by the 3' UTR of the R2Bm retrotransposase from *Bombyx mori.*

HEK293 cells are transfected with the following test agents:

1. Scrambled control RNA
2. RNA coding for the polypeptide described above
3. Template RNA comprising the EPO splice acceptor described above
4. The complete Gene Writer system comprising both (2) and (3)

After transfection, HEK293 cells are cultured for at least 4 days and assayed for site-specific Gene Writing genome editing and appropriate mRNA processing. Genomic DNA is isolated from the HEK293 cells. Reverse transcription-PCR is conducted to measure the mature mRNA containing the first natural exon of the albumin locus and the new exon. The RT-PCR reaction is conducted with forward primers that bind to the first natural exon of the albumin locus and with reverse primers that bind to EPO. The RT-PCR product is run on an agarose gel to measure the length of DNA. A PCR product of the expected length is detected in cells

transfected with the test agent of group 4, indicative of a successful Gene Writing genome editing event and a successful splice event. This result demonstrates that a Gene Writing genome editing system can add a heterologous sequence encoding a gene into an intronic region to act as a splice acceptor for the upstream exon.

The transfected cells are cultured for a further 10 days, and after multiple cell culture passages are assayed for EPO secretion in the cell supernatant. The amount of EPO in the supernatant is measured via an EPO ELISA kit. EPO is detected in HEK293 cells that were transfected with the test agent of group 4, demonstrating that a heterologous sequence can be added into an intronic region via Gene Writing genome editing, to act as a splice acceptor for the upstream exon and is actively expressed.

**Example 6: Targeted delivery of R2Tg retrotransposon to mammalian cells**

This example describes targeted integration of the R2Tg retrotransposon element (see first row of Table 3 herein) to mammalian cells via DNA or RNA delivery.

R2Tg is an endogenous retrotransposon from the zebra finch (*Taenopygia guttata*). Because non-LTR R2 elements are not present in the human genome and are thought to be highly site-specific, the ability of R2Tg to accurately and efficiently integrate itself into the human genome would demonstrate the capability to perform genomic targeted integration and possibly enable human gene therapy.

In the DNA delivery method, plasmid harboring R2Tg (PLV014) was designed and synthesized such that the R2Tg element was codon optimized and flanked by its native un-translated regions (UTRs), with or without further flanking by 100bp homology to the rDNA target locus. The R2Tg element expression was driven by the mammalian CMV promoter.

Further, a 1bp deletion mutant (678*) having a frameshift in the coding sequence of the retrotransposase was constructed as an inactivated control ("frameshift mutant"). Each plasmid was introduced into HEK393T cells via FuGENE® HD transfection reagent. HEK293T cells were seeded in 96-well plate, 10,000 cells/ well 24 hr before transfection. On the transfection day, 0.5µl transfection reagent and 80ng DNA was mixed in 10µl Opti-MEM and incubated for 15min at room temperature. Then the transfection mixture was added to the medium of the seeded cells. 3 days after transfection, genomic DNA was extracted for retrotransposition assays.

Next, integration of the R2Tg transposase into the human genome was assessed. Based on homology to the finch genome, a putative integration site in human rDNA was tested. Advanced Miseq and ddPCR assays were used to assess integration.

Bias in Miseq library construction was eliminated by introducing random unique molecular indices (UMIs) into initial PCRs (Figure 7). Nested PCR was performed by first amplifying the expected 3' junction of R2Tg and the rDNA locus for 30 cycles. One Miseq adapter, a multiplexing barcode, and an 8bp UMI were introduced at this step. A second PCR was used to further enrich for expected products and add the second Miseq adapter. Samples were sequenced on the Miseq for 300 cycles. After demultiplexing, the samples were analyzed via Matlab. First, the UMI on each sequence was located by searching for neighboring sequence. A database of UMIs was created and next collapsed by uniqueness. For each unique read, for a search was performed for the sequence of the expected rDNA integration site and isolated sequences of aligned human genomic DNA and exogenous DNA. Exogenous DNA was then aligned to the expected integration sequence. Results of the Miseq analysis pipeline are shown in Figures 8A-8B. Extensive unique integrations into the predicted integration site were found in cells treated with the wildtype R2Tg construct flanked by 100bp homology to the target rDNA

locus, but not with the frameshift mutant controls. Most integration events have a complete template RNA sequence integrated in the 565 bp most proximal to the integration site as demonstrated by sequencing reads that align perfectly to the expected sequence. A subset of integration events with the experimental R2Tg have either a ~300 bp or ~450 truncation as based upon sequencing reads that align to the expected sequence after a gap directly adjacent to the target site (Figure 8A). More specifically, 86.17% of integrants observed were non-truncated in the 565 bp most proximal to the integration site. In contrast, Fig. 8B shows no integration events detected. Constructs without flanking rDNA homology showed insignificant integration signals near noise.

ddPCR was next performed to confirm integration and assess integration efficiency. A Taqman probe was designed to the 3'UTR portion of the R2Tg element. A forward primer was synthesized to bind directly upstream of the probe, and a reverse primer was synthesized to bind the rDNA. Therefore, amplification of the expected product across the integration junction degrades the probe and creates a fluorescent signal. ddPCR was performed on several replicate experiments of the above plasmids to determine the average copy number of the R2Tg integration event. The results of ddPCR copy number analysis (in comparison to reference gene RPP30) are shown in Figure 9. Across several plasmid transfection conditions, average integration of 5 or more copies of R2Tg per genome at the target site when delivered with homology was noted with significant increase above control constructs. In contrast, the average copy number per genome in the frameshift mutant negative control was typically lower than 1. Insignificant signal was seen when constructs without homology were delivered to cells. The experiments collectively suggest efficient integration of the R2Tg retrotransposon into human cells at the target site.

In the RNA delivery method, R2Tg RNA (RNAV019) was designed such that the R2Tg element was codon optimized and flanked by its native untranslated regions (UTRs). More specifically, the construct includes, in order: a T7 promoter, a 5' 28S target homology region 100 nucleotides in length, a R2Tg wild-type 5' UTR, the R2Tg codon-optimized coding sequence, a R2Tg wild type 3' UTR, and a 3' 28S target homology region 100 nucleotides in length. The 100bp 28S homology sequences were added outside of the UTRs to enhance the integration. R2Tg RNA was synthesized, and cap and polyA tail were added. The R2Tg element transcription was driven by the T7 promoter. The RNA was introduced into HEK393T cells via Lipofectamine™ RNAiMAX or TransIT®-mRNA transfection reagent with series of RNA dosages. HEK293T cells were seeded in 96-well plate 24hr before transfection. On the transfection day, transfection reagent and RNA were mixed in 10µl Opti-MEM, and the transfection mixture was added to the medium of the seeded cells. 3 days after transfection, genomic DNA was extracted to measure retrotransposition efficiency using ddPCR with the same design as the DNA delivery.

The results of ddPCR copy number analysis (normalized to reference gene RPP30) are shown in Figure 12. Across several transfection conditions, the average integration was measured to be 0.01 of R2Tg copies per genome, significantly above the limit of detection. The results indicate successful integration of the R2Tg retrotransposon into human cells using an RNA delivery method.

**Example 7: Targeted delivery of a heterologous object sequence using R2Tg retrotransposon to mammalian cells**

This example describes the delivery of a transgene to human cells by utilizing the R2Tg retrotransposon system with multiple delivery machineries, including RNA-mediated delivery of a heterologous object sequence to human cells by utilizing the R2Tg retrotransposon system.

R2 proteins recognize their template RNA structure in untranslated regions (UTRs) of each element to form ribonucleoprotein particles, which serve as the intermediates of downstream integration into a host genome. Therefore, the decoupling of UTRs from their native context and the introduction of UTRs into alternate exogenous sequence was engineered to deliver into the genome a desired nucleic acid using R2Tg machinery.

Trans-transgene integration was tested by constructing 1) R2Tg coding sequence and 2) transgene cassette flanked by R2Tg UTR sequences and 100bp homology to 28S rDNA into separate driver and transgene plasmids, respectively. Figure 13 illustrates the dual plasmid system. The dual plasmids were introduced into HEK293T cells via FuGENE® HD transfection reagent at multiple driver to transgene molar ratios. In addition to the WT R2Tg driver, backbone plasmid was used as a control. HEK293T cells were seeded in 96-well plates at 10,000 cells/ well 24 hr before transfection. On the transfection day, transfection reagent and plasmids were mixed in 10µl Opti-MEM and incubated for 15 minutes at room temperature, then added to the medium of the seeded cells. 3 days after transfection, genomic DNA was extracted for ddPCR assays to investigate the trans-retrotransposition efficiency. Figure 14 demonstrates the ddPCR results for conditions with excess of transgene relative to driver.

Similar to the trans-transgene delivery with plasmids, RNA delivery was performed by constructing an amplicon of the coding sequence of R2Tg preceded by the T7 promoter

sequence. The constructed amplicons contained the experimental R2Tg element as well as the 1bp deletion frameshift mutant control. Separately, an amplicon was constructed that contained exogenous sequence coding for GFP and an EGF1-alpha reporter that was flanked regions sufficient to drive integration into the genome by R2Tg. More specifically, the construct included: a T7 promoter driving transcription of the RNA, wherein the RNA comprises, from 5' to 3', (a) a 5' 28S homology region of 10 nt in length, (b) a 5' untranslated region, (c) an anti-sense TKpA polyA sequence, (d) an anti-sense heterologous object sequence that encodes GFP, (e) an anti-sense Kozak sequence, (f) an anti-sense EF1 alpha promoter, (g) a 3' untranslated region that binds the GeneWriter protein, and (h) a 3' 28S homology region of 10 nt in length. Each RNA was transcribed via the New England Biolabs HiScribe T7 ARCA kit and purified via Zymo RNA clean and concentrator.

The resulting heterologous object RNA and R2Tg RNA (either the experiment R2Tg element or frameshift mutant) were introduced into human HEK293T cells via TransIT®-mRNA Transfection Kit at 1:1 molar ratio. HEK293T cells were seeded in 96-well plate, 40,000 cells/ well 24 hr before transfection. On the transfection day, 1µl transfection reagent and 500ng total RNA was mixed in 10µl Opti-MEM and incubated for 5 min at room temperature. Then the transfection mixture was added to the medium of the seeded cells. 3 days after transfection, genomic DNA was extracted for PCR assays.

Nested PCR was performed by with a first 30 rounds of PCR across the 3' end of the expected transgene-rDNA junction, followed by 20 additional rounds of PCR amplification using an inner primer set. One of three replicates of nested PCR performed on genomic DNA extracted from cells treated with the wild-type transposase reaction produced a PCR product of the expected size (approximately 596 bp). In contrast, no PCR product was observed in genomic

DNA extracted from cells treated with the frameshift-inactivated R2Tg mutant control, or no-transfection control. The PCR product was gel-purified via Zero Blunt® TOPO® PCR Cloning Kit, and the resulting colonies were Sanger sequenced. Each individual PCR product sequence was then aligned to the expected integration sequence. The fraction of PCR product sequences that align to the expected integrated heterologous object sequence is shown in Figure 10. The majority of PCR products had the expected integrant as demonstrated by the sequencing alignment directly adjacent to the expected integration site at the right side of the alignment figure. This demonstrates RNA-mediated integration of the exogenous sequence via R2Tg machinery into human cells.

**Example 8: Targeted delivery of R2Tg retrotransposon to mammalian cells**

This example describes targeted integration of the R2Tg retrotransposon element to mammalian cells via DNA delivery.

Plasmid harboring R2Tg (PLV014) and control plasmid were designed and synthesized as described above in Example 6. Each plasmid was introduced into HEK393T cells via FuGENE® HD transfection reagent. HEK293T cells were seeded in 96-well plate, 10,000 cells/ well 24 hr before transfection. On the transfection day, 0.5µl transfection reagent and 80ng DNA was mixed in 10µl Opti-MEM and incubated for 15 min at room temperature. Then the transfection mixture was added to the medium of the seeded cells. 3 days after transfection, genomic DNA was extracted for retrotransposition assays or cells were frozen and underwent targeted locus amplification.

Target locus amplification was performed against hg38 reference human genome and the rDNA locus sequence hsu13369 (GenBank: U13369.1). Two independent primer sets were used

to perform targeted locus amplification. Analysis with both primer sets showed that the 28S rDNA locus sequence is the only integration site detected above a 1% threshold. Thus, integration of the R2Tg transposon in mammalian cells is specific to this target site.

**Example 9: Improved trans RNA-templated integration into mammalian cells by RNA refolding or ratio of driver to template RNA**

RNA templates are designed as in previous examples. Two RNAs consisting of a driver and a transgene payload are delivered to mammalian cells. To better promote folding, denaturing the payload RNA by heating to 95C and cooling to room temperature are performed to encourage proper secondary structure formation. In some embodiments, cooling the RNA to room temperate will increase integration efficiency.

The molar ratio of transgene to driver is also varied to evaluate suitable stoichiometry of components. Integration is analyzed via ddPCR and sequencing. In some embodiments, a higher ratio of driver to transgene is used. In some embodiments, a higher ratio of transgene to driver is used.

Previous examples with cis transgene integration are similarly assayed for stoichiometry of driver to payload. Integration is analyzed via ddPCR and sequencing. In some embodiments, a higher ratio of driver transcription or translation to transgene transcription will result in higher integration efficiency. In some embodiments, a higher ratio of transgene transcription to driver transcription and translation will result in higher integration efficiency.

**Example 10: Hybrid Capture Assay**

A hybrid capture experiment was performed to obtain an unbiased view of the specificity of retrotransposon integration into a target site. Retrotransposon experiments were performed as

in previous examples by integrating R2Tg flanked by its native UTRs and 100bp of homology to either side of the expected R2 rDNA target. The rDNA target site had two flanking sets of 100 nucleotides identity to the corresponding native target site. The retrotransposon was delivered to human 293T cells via plasmid or mRNA. Genomic DNA was extracted after 72 hours. After extraction, each genomic DNA sample was subjected to hybrid capture according to protocol with a custom probe set (Twist). Biotinylated probes were designed such that ~120bp probes spanned both strands of the R2Tg coding sequence and UTRs. First, a next-generation library was created by fragmentation of the genomic DNA and ligation of sequencing adapters according to a protocol from Twist ( available on the world wide web at: twistbioscience.com/ngs_protocol_custompanel_hybridcap) . Next, probes were hybridized to genomic DNA libraries and the enriched samples were amplified. Final libraries were seqenced on the Miseq using 300bp paired-end reads. Custom Matlab scripts were used to analyze reads. The resulting analysis is shown in Figures 15A and 15B for RNA delivery. Hybrid capture indicated on-target integration of R2Tg to the expected locus. With RNA delivery, 1 possible off-target with a single read was identified at an unexpected 3' junction in the data, compared to more than 100 reads at the expected locus, indicating a specificity of greater than 100:1. At the 5' junction, all 50 reads were at the expected locus, indicating a specificity of greater than 50:1. This experiment indicates a high specificity of integration.

**Example 11: Long-read PacBio Analysis**

Long-range PCR amplification can be performed to measure integration of the desired full-length sequence into the target site in the human genome and to measure whether mutations are introduced during insertion. Retrotransposon integration experiments are performed as

described in previous examples. In one example, PCR amplification is used to generate amplicons by designing one primer targeting the genomic integration site and one primer targeting the integrant sequence. In this example, these primers are designed to maximize the length of the amplified genomic locus fused with the integrant sequence. By pooling amplicons spanning both ends of the integrant and performing long-read next-generation sequencing, the fidelity of each integration is be evaluated.

In another example, hybrid capture is performed as described in a previous example but with a larger target library length during initial library generation. The resulting library is then subjected to long-read next-generation sequencing.

In some embodiments, long-read next generation sequencing will show that there are less than 10%, 5%, 2%, 1%, 0.5%, 0.2%, or 0.1% SNPs in the integrated DNA across samples. In some embodiments, long-read next generation sequencing will show that less than 10%, 5%, 2%, or 1% of integrated DNA has a SNP. In some embodiments, long-read next generation sequencing will show that less than 10%, 5%, 2%, or 1% of integrated DNA has an internal deletion. In some embodiments, long-read next generation sequencing will show that less than 10%, 5%, 2%, 1%, 0.5%, 0.2%, or 0.1% of total integrated DNA across the population is deleted. In some embodiments, long-read next generation sequencing will show that less than 10%, 5%, 2%, or 1% of integrated DNA is truncated.

**Example 12: Experiment with different homology lengths and point mutations in homology**

In this example, experiments are designed to characterize suitable lengths and starting positions of homology to the target site for efficient retrotransposon integration. Also, the homology is used to support the mechanism of integration being reverse transcription-driven.

A series of SNPs were introduced within the 100bp downstream homology of R2Tg plasmids by modifying plasmid PLV014. The design of the SNPs is listed in Figure 16. After the transfection, nested PCR was applied to recover the 3' integration junction site, producing a PCR product with an expected amplicon size of about 738 bp, and the PCR product was Sanger sequenced to check whether any SNPs were incorporated. In this experiment, a lack of SNP genetic markers being incorporated into the junction sequences indicates that the integration was driven by reverse transcription. The SNP design and the sequencing result are illustrated in Figure 16. No SNP introduction was observed for the 18 genetic markers designed, consistent with the integration of R2Tg being directed by reverse transcription.

This example also describes the evaluation different homology regions to the target site to identify shorter regions that promote efficient integration into the genome. This example describes two approaches. First, different windows of 100 bp of homology to the target site are tested, starting from bp 1-100 3' of the target site, then testing 2-101 3' of the target site, 3-103 3' of the target site, and so on, through bp 30 – 131 3' of the target site. Second, shorter lengths of homology to the target site sufficient for DNA integration are tested, starting with bp 0-100 3' of the target site, then testing 0-95 3' of the target site, 0 – 90 3' of the target site, etc. through bp 0-10 3' of the target site. After the transfection of each plasmid into 293T cells, ddPCR is used to measure the retrotransposition efficiency.

In this example, different UTR regions with different lengths are evaluated to identify shorter sequences for efficient integration into the genome. The 3'UTR is tested by dividing this 325 bp sequence into 3 regions, 1-100 bp, 101-200 bp, and 201-325 bp. Constructs of R2Tg containing each truncated 3'UTR are generated to test the integration efficiency respectively.

**Example 13: Assess whether p53 or other repair pathways are upregulated**

This example describes an evaluation of the effect of exogenous R2Tg retrotranspositon on gene expression, especially tumor suppressor and DNA repair genes. An R2Tg expressing plasmid is delivered to multiple cancer cell lines, including 293T, MCF-7, and T47D. After confirmation of integration in each cell line, RNA-seq is conducted to assess the effect on gene expression profile. Gene set enrichment analysis is then applied to evaluate whether any DNA repair pathways are upregulated after retrotransposition. MCF-7 and T47D are breast cancer cell lines with wild type and mutant p53 respectively, which are be used to evaluate the relationship between p53 and retrotranspositon specifically. In some embodiments, p53 is not upregulated when a retrotransposon Gene Writer integrates into the genome. In some embodiments, no DNA repair genes are upregulated when a retrotransposon Gene Writer integrates into the genome. In some embodiments, no tumor suppressor genes are upregulated when a retrotransposon Gene Writer integrates into the genome.

**Example 14: Retrotransposition in presence of DNA repair inhibitors**

In this example, experiments will test the effect of different DNA repair pathways on R2Tg retrotransposition via the application of DNA repair pathway inhibitors or DNA repair pathway deficient cell lines. When applying DNA repair pathway inhibitors, PrestoBlue cell viability assay is performed first to determine the toxicity of the inhibitors and whether any normalization should be applied for following assays. SCR7 is an inhibitor for NHEJ, which is applied at a series of dilutions during R2Tg delivery. PARP protein is a nuclear enzyme that binds as homodimers to both single- and double-strand breaks. Thus, its inhibitors are be used in the test of relevant DNA repair pathways, including homologous recombination repair pathway and base excision repair

pathway. The experiment procedure is the same with that of SCR7. Cell lines with deficient core proteins of nucleotide excision repair (NER) pathway are used to test the effect of NER on R2Tg retrotransposition. After the delivery of R2Tg element into the cell, ddPCR is be used to evaluate the retrotransposition in the context of inhibition of DNA repair pathways. Sequencing analysis is also be performed to evaluate whether certain DNA repair pathway plays a role in the alteration of integration junction. In some embodiments, R2Tg integration into the genome will not be decreased by the knockdown of any DNA repair pathways, suggesting that R2Tg does not rely on the host cell pathways for DNA integration.

**Example 15: Retrotransposition in fibroblasts and in T cells**

In this example, the previously performed R2Tg retrotransposition analysis of 293T cells is repeated in non-dividing cells, including fibroblast and T cells. Compared to 293T cells, non-dividing cells are sometimes more difficult to transfect with lipid reagent. Thus, nucleofection is used for the delivery of R2Tg element. The subsequent retrotransposition assay for integrating efficiency and sequencing analysis will be performed as described herein for 293T cells. In some embodiments, R2Tg integrates into the genome of fibroblasts and T cells.

**Example 16: Single cell ddPCR**

In this example, a quantitative assay is used to determine the frequency of targeted genome integration at single cell level, and that information can be compared to the copy number of targeted genome integration per genome quantified from genomic DNA.

Approximately 5000 transfected cells will be collected and mixed with ddPCR reaction mixture before distributing into about 20,000 droplets, with the aim of each droplet containing

one cell or no cells. ddPCR assays including 5'UTR and 3'UTR assays will be performed as described above to determine the frequency of R2 or transgene integration at single cell level. A control experiment will be performed in parallel using genomic DNA harvested from the same number of cells to determine the targeted genome integration efficiency per genome. In some embodiments, the frequency of targeted genome integration at the single cell level is calculated to be 1-80%, e.g., 25%, wherein the indicated percentage of cells have one or more copies of the transgene integrated into the desired locus.

**Example 17: Single cell analysis via colony isolation**

In this example, a quantitative assay is used to determine genome integration copy number in cell colonies derived from single cell.

Single cell colonies will be isolated by colony picking up or by limited dilution and cultured in a 96 well format. When the cells reach >80% confluency, half of the cells will be frozen for backup and genomic DNA from the other half of the cells will be harvested for ddPCR. Optimized ddPCR assays including 5'UTR and 3'UTR assays will be performed as described previously to determine the frequency of R2 or transgene integration. At least 96 colonies will be screened for each R2 element with appropriate controls. The total number of colonies to be screened will be determined by single cell ddPCR data if applicable or the first set of single cell colony screen data. In some embodiments, the frequency of targeted genome integration at the single cell level will be calculated to be 1-80%, e.g., 25%, wherein the indicated percentage of cells have a single copy of the transgene integrated into the desired locus. The assay can also be used to determine the percentage of colonies that have more than one copy of the transgene integrated into the desired locus.

**Example 18: DNA binding affinity and/or re-targeting**

The DNA targeting module of wild-type R2 is made of a cysteine-histidine zinc finger and c-Myb transcription factor binding motifs. This N-terminal module can be substituted with different DNA binding modules such as DNA binding protein(s) (e.g., transcription factors), zinc finger(s) (e.g., natural or designed motifs), and/or nucleic acid guided, catalytically inactive endonucleases (e.g., Cas9 bound with a guide RNA (e.g., sgRNA) to form a Cas9-RNP). This DNA binding module is swapped for the naturally occurring module and, in some cases placed with a flexible linker attaching it to the RNA binding/RT module. Additionally, in some constructions, this new DNA binding module is placed in tandem with the same and/or different DNA binding modules. Furthermore, some constructions may split the GeneWriter protein where one protein molecule contains the RNA binding module and the other protein contains the RT and endonuclease modules. In some embodiments, swapping of the DNA module increases specificity and/or affinity to a genomic location and in some cases allows for the specific targeting of new genomic locations.

**Example 19: Assays to measure DNA binding affinity**

DNA binding activity of GeneWriters described herein (and DNA binding domains for the same) can be tested, e.g., as described in this example. DNA binding modules are purified by recombinantly expressing them in cells (e.g., *E. coli*) or they are expressed in a cell-free reactions of transcription and translation (e.g., T7 RNA polymerase + wheat germ extract). The purified DNA binding module(s) is tested for binding affinity by measuring the Kd in a binding assay (e.g., EMSA, Fluorescence anisotropy, dual-filter binding, FRET, SPR, or thermophoresis

(temperature related intensity change). The protein (DNA binding module) is labeled and/or the DNA molecule is labeled with a molecule that is compatible with the above binding assays (e.g., dye, radioisotope (for example, Protein: $^{35}$S-methionine, maleimide dye, DNA: $^{32}$P end or internal label, DNA with a linked amine reacted with NHS-ester dye). The molecules are measured by changing their concentrations and fitting to a binding curve which calculates the binding affinity. In some assays, the nucleic acid sequence specificity is tested by mutational analysis of the DNA sequence or mutation to the DNA binding module by amino acid changes or alterations to protein-nucleic acid complex (e.g., Cas9-RNP DNA binding module). In some embodiments, increasing the Kd of the DNA binding module will decrease off-target insertions and, in some cases, will increase the activity of on-target sites by increasing the dwell time of the R2-RNA complex at the specific genomic location.

**Example 20: Assays to determine global specificity *de novo***

The DNA binding module is expressed in cells (e.g., animal cells, e.g., human cells) as the DNA binding module alone, in the context of the full-length retrotransposon R2, or a control without retrotransposase. The expression of the module or retrotransposon is delivered to cells using conventional methods of delivering DNA, RNA, or protein. The complex is crosslinked (e.g., using chemical or UV light) or is not crosslinked. The cells are lysed and treated with DNase I so that only the bound DNA is protected from degradation. DNA is extracted, NGS library preparation of DNA fragments and de novo binding sites are identified, analogous to ChIP-seq or DIG-seq. In some embodiments, potential off-target sites are identified that can be followed-up to remove false-positives. In other embodiments this assay confirms the in vitro assay on the specificity of the DNA binding module to bind at its intended site and not at others.

An orthogonal assay to identify DNA binding sites in high-througput uses the method described by Boyle et al, PNAS 2017 where the DNA binding domain is tested in a cell-free setting to determine specificity along with systematic analysis of sequence mutants related to the new DNA binding module.

**Example 21: Modularity of RNA molecule**

The RNA molecule binds to the R2 protein via interactions found in the reverse transcriptase module, designated as a sub-module "RNA binding". The protein recognizes specific structures in the 5' and/or 3' UTRs to interact with the RNA. In some embodiments, swapping of the UTR modules increases protein interactions, changes the protein specificity to bind the UTR, stabilizes against nucleases, and/or improves cellular tolerance (e.g., leads to a reduced innate immune response). In other embodiments, addition and/or swapping of the RNA binding module of the R2 protein is compatible with the use of different sequence or ligands that are linked to the transgene and/or element module of the RNA. In some embodiments, combinations of new ligands in place of the UTRs will have better affinity to the RNA binding domain of R2 and lead to better insertion efficiency. In some embodiments, the changes to the sequence of the UTRs or changes to the base modifications of the UTRs will increase the secondary structure stability that leads to better interaction with the RNA binding module.

**Example 22: Assays to measure RNA binding affinity to new sequences**

New UTR modules are tested in a binding assay. In the case of new RNAs, they are synthesized either by cell-free in vitro transcription using a synthetic DNA template or by chemical synthesis of the RNA in full-length or chemical synthesis of pieces that are ligated

together to form a single RNA molecule. The binding affinity of the purified UTRs are measured in a binding assay (e.g., EMSA, Fluorescence anisotropy, dual-filter binding, FRET, SPR, or thermophoresis (temperature related intensity change)). The UTR module and/or RNA binding module/RT module is detected with or without a label which is described above for labeling RNAs. Measurement of the molecules at different concentrations is performed to determine a binding affinity. In some embodiments, alterations to/swapping of the 5' and/or 3' UTR binding module and/or changes to the RNA binding/RT module will lead to better interactions than the wild-type R2 protein or UTR. In some embodiments, the increased interaction will lead to an increase in the efficiency of retro-transposition and in some cases increases specificity of the R2 protein to interact with the RNA.

**Example 23: Alternative UTRs**

While not wishing to be bound by theory, in some embodiments the UTRs act as a handle for the R2 protein to interact with the RNA which it uses as a template for RT in concert with it binding a genomic location, nicking the DNA with its endonuclease module, and then using the bound RNA as a template for RT insertion at the cleavage site in the DNA. For the UTR to keep the template in close proximity to the RT module, then the UTR modules can be substituted with different ligands that would bind to a specific RNA binding module engineered into the R2 protein. Thus, in some embodimetns, the alternative non-RNA UTR is either a protein, small molecule, or other chemical entity that is attached covalently, through protein-protein interaction, small molecule-protein interaction, or through hybridization. In some embodiments the RNA binding module binds specifically to a ligand that is not RNA that is attached to the transgene module RNA that increases the efficiency, stability, and/or rate of retro-transposition.

**Example 24: Assays to measure the activity of UTR constructs**

Binding assays to measure affinity of R2 protein with engineered UTRs are performed as described above, e.g., for a protein-nucleic acid interaction. In cases of protein-protein or protein-small molecule interactions the assay uses a label on the RNA transgene module where the UTR module is attached.

**Example 25: Targeted genomic integration**

In this example, Gene Writing technology is delivered to target cells and to non-target cells, and new DNA is integrated into the genome in target cells at a higher frequency than in non-target cells. As described in more detail below, this approach takes advantage of the non-target cell having an endogenous miRNA that the target cell does not have (or has at a lower level). The endogenous miRNA is used to reduce DNA integration in the non-target cell.

The polypeptide used is the R2Tg protein and the template RNA component is RNA coding for the GFP protein and flanked at the 5' end by the 5' UTR and at the 3' end by the 3' UTR of the R2Tg retrotransposase. The 5' UTR is flanked by 100 bp of homology to the 5' of R2Tg 28s rDNA target site and the 3' UTR is flanked by 100 bp of homology to the 3' of R2Tg 28s rDNA target site. The GFP gene is facing in the antisense direction with regard to the 5' and 3' UTRs and has its own promoter and polyadenylation signal.

The template RNA further comprises a microRNA recognition sequence. This microRNA recognition sequence is bound by microRNAs in the non-target cells, leading to the inhibition (e.g., degradation) of the template RNA prior to genomic integration.

In this example the target cells are hepatocytes and the non-target cells are macrophages from the hematopoietic lineage. The target cells and non-target cells are cultured separately. The template RNA and retrotransposase protein can be delivered to cells as described herein, e.g., as RNA or using viral vectors (e.g. adeno-associated viral vectors), wherein the template RNA is transcribed from viral vector DNA.

Three days after treating the cells, GFP expression and genomic integration are assayed.

GFP expression is assayed via flow cytometry. In some embodiments, GFP expression will be higher in the hepatocyte population than in the macrophage population.

Genomic integration (in terms of copy number per cell normalized to a reference gene) is assayed via droplet digital PCR using methods described herein. In some embodiments, genomic integration will be higher in the hepatocyte population than in the macrophage population.

**Example 26: Testing 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 these proteins. This experiments tested flexible linkers in different locations and lengths, in order to determine if the DNA binding domain (DBD) was modular. These experiments also provide support for being able to separate the DBD from the rest of R2Tg and replacing it with any DNA targeting protein sequence. This example thus supports an understanding that the transposases described herein can withstand the tested levels of sequence divergence at a plurality of locations (e.g., in the predicted -1 RNA binding motif, in an alpha helix, and in a coil region located C-terminal to the

predicted c-myb DNA binding motif, e.g., as described below) identified by structural modeling, while maintaining function.

Briefly, the two linkers (Linker A: SGSETPGTSESATPES (SEQ ID NO: 1023), and Linker B: GGGS (SEQ ID NO: 1024)) were inserted into 3 locations, noted herein as versions v1, v2, and v3. v1 was located at the N-terminal side of an alpha helical region of R2Tg that preceded the predicted -1 RNA binding motif, v2 was located at the C-terminal side of an alpha helical region of R2Tg that preceded the predicted -1 RNA binding motif, and v3 was located C-terminal to a random coil region that came after the predicted c-myb DNA binding motif of R2Tg. For each of v1, v2, and v3, one of linkers A or B were added by PCR to a DNA plasmid that expressed R2Tg, thereby yielding sequences v1A (v1 + linker A), v1B (v1 + linker B), v1C (v1 + linker C), v2A (v2 + linker A), v2B (v2 + linker B), and v2C (v2 + linker C), as shown in Table 5 below. The insertion of the linkers was verified by Sanger sequencing and the DNA plasmids were purified for transfection.

**Table 5. Amino acid sequences of R2Tg mutants with linkers in the DNA binding domain (DBD)**

| R2Tg Mutant + Linker | Amino Acid Sequence | SEQ ID NO |
|---|---|---|
| R2Tg with DBD Linker v1A | MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNS<br>LANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDL<br>VSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVD<br>LVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYE<br>CVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLP<br>RDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQT<br>GEEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVS<br>HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETE<br>KAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQ<br>ERIVASQPKETSNRGAHKRCWTKEEEELLIRLEAQF<br>EGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEE<br>PREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNG<br>PGRRPLPGRAAAGGRTMDEIRRHPDKGNGQQRPTKQ<br>KSEEQLQAYYKKTLEERLSAGALNTFPRAFKQVMEG<br>RDIKLVINQTAQDSGSETPGTSESATPESCFGCLES | 1017 |

| | | |
|---|---|---|
| | ISQIRTATRDKKDTVTREKHPKKPFQKWMKDRAIKK<br>GNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE<br>IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELI<br>TAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFS<br>RTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRL<br>KDINNWRPITIGSILLRLFSRIVTARLSKACPLNPR<br>QRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFV<br>DIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYE<br>NISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL<br>AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLS<br>DSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKP<br>TKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQF<br>DPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDIL<br>KTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEW<br>LHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQAR<br>RLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGD<br>RENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFP<br>KPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWI<br>QYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQY<br>IKACRHCDADIESCAHIIGNCPVTQDARIKRHNYIC<br>ELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVK<br>DARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLET<br>EVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELG<br>LSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSM<br>VM | |
| R2Tg with DBD Linker v1B | MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNS<br>LANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDL<br>VSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVD<br>LVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYE<br>CVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLP<br>RDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQT<br>GEEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVS<br>HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETE<br>KAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQ<br>ERIVASQPKETSNRGAHKRCWTKEEEELLIRLEAQF<br>EGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEE<br>PREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNG<br>PGRRPLPGRAAAGGRTMDEIRRHPDKGNGQQRPTKQ<br>KSEEQLQAYYKKTLEERLSAGALNTFPRAFKQVMEG<br>RDIKLVINQTAQDGGGSCFGCLESISQIRTATRDKK<br>DTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYL<br>DRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETT<br>GSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEM<br>SKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTT<br>GKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIG<br>SILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSE<br>NLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQ<br>HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNT<br>HTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEES<br>GKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISI | 1018 |

| | | |
|---|---|---|
| | LETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAA<br>WTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLS<br>TKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIA<br>DHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAIL<br>YSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDT<br>MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPP<br>SSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKK<br>WTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLL<br>TALQLRANVYPTREFLARGRQDQYIKACRHCDADIE<br>SCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDW<br>VVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVR<br>YEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVT<br>FVGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAET<br>FSTVALFSSVDIVHMFASRARKSMVM | |
| R2Tg with DBD Linker v2A | MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNS<br>LANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDL<br>VSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVD<br>LVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYE<br>CVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLP<br>RDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQT<br>GEEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVS<br>HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETE<br>KAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQ<br>ERIVASQPKETSNRGAHKRCWTKEEEELLIRLEAQF<br>EGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEE<br>PREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNG<br>PGRRPLPGRAAAGGRTMDEIRRHPDKGNGQQRPTKQ<br>KSEEQLQAYYKKTLEERLSAGALNTFPRAFKQVMEG<br>RDIKLVINQTAQDCFGCLESISQIRSGSETPGTSES<br>ATPESTATRDKKDTVTREKHPKKPFQKWMKDRAIKK<br>GNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE<br>IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELI<br>TAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFS<br>RTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRL<br>KDINNWRPITIGSILLRLFSRIVTARLSKACPLNPR<br>QRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFV<br>DIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYE<br>NISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL<br>AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLS<br>DSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKP<br>TKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQF<br>DPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDIL<br>KTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEW<br>LHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQAR<br>RLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGD<br>RENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFP<br>KPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWI<br>QYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQY<br>IKACRHCDADIESCAHIIGNCPVTQDARIKRHNYIC<br>ELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVK<br>DARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLET | 1019 |

| | | |
|---|---|---|
| | EVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELG<br>LSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSM<br>VM | |
| R2Tg with DBD Linker v2B | MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNS<br>LANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDL<br>VSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVD<br>LVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYE<br>CVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLP<br>RDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQT<br>GEEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVS<br>HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETE<br>KAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQ<br>ERIVASQPKETSNRGAHKRCWTKEEEELLIRLEAQF<br>EGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEE<br>PREEPGTCHHTRRAAASLRTEPEMSHHAQAEDRDNG<br>PGRRPLPGRAAAGGRTMDEIRRHPDKGNGQQRPTKQ<br>KSEEQLQAYYKKTLEERLSAGALNTFPRAFKQVMEG<br>RDIKLVINQTAQDCFGCLESISQIRGGGSTATRDKK<br>DTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYL<br>DRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETT<br>GSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEM<br>SKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTT<br>GKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIG<br>SILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSE<br>NLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQ<br>HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNT<br>HTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEES<br>GKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISI<br>LETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAA<br>WTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLS<br>TKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIA<br>DHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAIL<br>YSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDT<br>MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPP<br>SSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKK<br>WTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLL<br>TALQLRANVYPTREFLARGRQDQYIKACRHCDADIE<br>SCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDW<br>VVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVR<br>YEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVT<br>FVGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAET<br>FSTVALFSSVDIVHMFASRARKSMVM | 1020 |
| R2Tg with DBD Linker v3A | MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNS<br>LANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDL<br>VSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVD<br>LVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYE<br>CVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLP<br>RDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQT<br>GEEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVS<br>HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETE | 1021 |

| | | |
|---|---|---|
| | KAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQ<br>ERIVASQPKETSNRGAHKRCWTKEEEELLIRLEAQF<br>EGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEE<br>PREEPGTCHHTRRAASGSETPGTSESATPESASLRT<br>EPEMSHHAQAEDRDNGPGRRPLPGRAAAGGRTMDEI<br>RRHPDKGNGQQRPTKQKSEEQLQAYYKKTLEERLSA<br>GALNTFPRAFKQVMEGRDIKLVINQTAQDCFGCLES<br>ISQIRTATRDKKDTVTREKHPKKPFQKWMKDRAIKK<br>GNYLRFQRLFYLDRGKLAKIILDDIECLSCDIPLSE<br>IYSVFKTRWETTGSFKSLGDFKTYGKADNTAFRELI<br>TAKEIEKNVQEMSKGSAPGPDGITLGDVVKMDPEFS<br>RTMEIFNLWLTTGKIPDMVRGCRTVLIPKSSKPDRL<br>KDINNWRPITIGSILLRLFSRIVTARLSKACPLNPR<br>QRGFIRAAGCSENLKLLQTIIWSAKREHRPLGVVFV<br>DIAKAFDTVSHQHIIHALQQREVDPHIVGLVSNMYE<br>NISTYITTKRNTHTDKIQIRVGVKQGDPMSPLLFNL<br>AMDPLLCKLEESGKGYHRGQSSITAMAFADDLVLLS<br>DSWENMNTNISILETFCNLTGLKTQGQKCHGFYIKP<br>TKDSYTINDCAAWTINGTPLNMIDPGESEKYLGLQF<br>DPWIGIARSGLSTKLDFWLQRIDQAPLKPLQKTDIL<br>KTYTIPRLIYIADHSEVKTALLETLDQKIRTAVKEW<br>LHLPPCTCDAILYSSTRDGGLGITKLAGLIPSVQAR<br>RLHRIAQSSDDTMKCFMEKEKMEQLHKKLWIQAGGD<br>RENIPSIWEAPPSSEPPNNVSTNSEWEAPTQKDKFP<br>KPCNWRKNEFKKWTKLASQGRGIVNFERDKISNHWI<br>QYYRRIPHRKLLTALQLRANVYPTREFLARGRQDQY<br>IKACRHCDADIESCAHIIGNCPVTQDARIKRHNYIC<br>ELLLEEAKKKDWVVFKEPHIRDSNKELYKPDLIFVK<br>DARALVVDVTVRYEAAKSSLEEAAAEKVRKYKHLET<br>EVRHLTNAKDVTFVGFPLGARGKWHQDNFKLLTELG<br>LSKSRQVKMAETFSTVALFSSVDIVHMFASRARKSM<br>VM | |
| R2Tg with DBD Linker v3B | MASCPKPGPPVSAGAMSLESGLTTHSVLAIERGPNS<br>LANSGSDFGGGGLGLPLRLLRVSVGTQTSRSDWVDL<br>VSWSHPGPTSKSQQVDLVSLFPKHRVDLLSKNDQVD<br>LVAQFLPSKFPPNLAENDLALLVNLEFYRSDLHVYE<br>CVHFAAHWEGLSGLPEVYEQLAPQPCVGETLHSSLP<br>RDSELFVPEEGSSEKESEDAPKTSPPTPGKHGLEQT<br>GEEKVMVTVPDKNPPCPCCGTRVNSVLNLIEHLKVS<br>HGKRGVCFRCAKCGKENSNYHSVVCHFPKCRGPETE<br>KAPAGEWICEVCNRDFTTKIGLGQHKRLAHPAVRNQ<br>ERIVASQPKETSNRGAHKRCWTKEEEELLIRLEAQF<br>EGNKNINKLIAEHITTKTAKQISDKRRLLSRKPAEE<br>PREEPGTCHHTRRAAGGGSASLRTEPEMSHHAQAED<br>RDNGPGRRPLPGRAAAGGRTMDEIRRHPDKGNGQQR<br>PTKQKSEEQLQAYYKKTLEERLSAGALNTFPRAFKQ<br>VMEGRDIKLVINQTAQDCFGCLESISQIRTATRDKK<br>DTVTREKHPKKPFQKWMKDRAIKKGNYLRFQRLFYL<br>DRGKLAKIILDDIECLSCDIPLSEIYSVFKTRWETT<br>GSFKSLGDFKTYGKADNTAFRELITAKEIEKNVQEM<br>SKGSAPGPDGITLGDVVKMDPEFSRTMEIFNLWLTT | 1022 |

| | | |
|---|---|---|
| | GKIPDMVRGCRTVLIPKSSKPDRLKDINNWRPITIG<br>SILLRLFSRIVTARLSKACPLNPRQRGFIRAAGCSE<br>NLKLLQTIIWSAKREHRPLGVVFVDIAKAFDTVSHQ<br>HIIHALQQREVDPHIVGLVSNMYENISTYITTKRNT<br>HTDKIQIRVGVKQGDPMSPLLFNLAMDPLLCKLEES<br>GKGYHRGQSSITAMAFADDLVLLSDSWENMNTNISI<br>LETFCNLTGLKTQGQKCHGFYIKPTKDSYTINDCAA<br>WTINGTPLNMIDPGESEKYLGLQFDPWIGIARSGLS<br>TKLDFWLQRIDQAPLKPLQKTDILKTYTIPRLIYIA<br>DHSEVKTALLETLDQKIRTAVKEWLHLPPCTCDAIL<br>YSSTRDGGLGITKLAGLIPSVQARRLHRIAQSSDDT<br>MKCFMEKEKMEQLHKKLWIQAGGDRENIPSIWEAPP<br>SSEPPNNVSTNSEWEAPTQKDKFPKPCNWRKNEFKK<br>WTKLASQGRGIVNFERDKISNHWIQYYRRIPHRKLL<br>TALQLRANVYPTREFLARGRQDQYIKACRHCDADIE<br>SCAHIIGNCPVTQDARIKRHNYICELLLEEAKKKDW<br>VVFKEPHIRDSNKELYKPDLIFVKDARALVVDVTVR<br>YEAAKSSLEEAAAEKVRKYKHLETEVRHLTNAKDVT<br>FVGFPLGARGKWHQDNFKLLTELGLSKSRQVKMAET<br>FSTVALFSSVDIVHMFASRARKSMVM | |

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 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 integration per genome. The 5′ and 3′ junctions were quantified by generating two different amplicons at each end.

v3 (near the c-myb binding motif in the DBD) decreased integration activity with either linker A or B. v1 (N-terminal to the alpha helix preceding the -1 RNA binding motif) had comparable activity to the wild-type when used with linker A (16 AA) versus the shorter linker B (4 AA). This could be related to amino acid selection, length, or three-dimensional structure. v2

(C-terminal to the alpha helix preceding the -1 RNA binding motif) did not tolerate linker A; however, linker B had activity that was comparable and slightly better than the wild-type. v1 and v2 may therefore be considered preferred locations to add a linker that can separate R2Tg's DNA binding domain and the rest of the protein.

**Example 27: Long-read sequencing to determine integration fidelity**

Retrotransposon integration experiments were performed as described in previous examples. In one example, PCR amplification was used to generate amplicons by designing one primer targeting the genomic integration site and one primer targeting the integrant sequence. In this example, these primers were designed to maximize the length of the amplified genomic locus fused with the integrant sequence. By pooling amplicons spanning both ends of the integrant and performing long-read next-generation sequencing, the fidelity of each integration was evaluated.

A cis construct of R2Tg was integrated into 293T cells via plasmid transfection as described herein. Amplicons spanning each end of the integrations were generated with flanking randomized UMIs to control for PCR bias. These amplicons were sequenced with PacBio next-generation sequencing. The resulting sequences were collapsed to remove reads with identical UMIs. By aligning unique reads, a coverage plot was constructed as shown in Figures 20A-20B. Sequence coverage largely shows uniform coverage across amplicons, indicating significant fidelity of integration. An associated reverse-transcriptase deficient mutant control produced no signal. Internal deletions were also analyzed in Figures 21A-21B. Internal deletions were generally low relative to overall unique read counts, with some clustering at the 5' junction of rDNA-R2Tg.

In another example, hybrid capture may be performed as described in a previous example but with a larger target library length during initial library generation. The resulting library can then be subjected to long-read next-generation sequencing.

**Example 28: Targeted delivery of R2Gfo and R4Al retrotransposon to mammalian cells**

This example describes targeted integration of the R2Gfo and R4Al retrotransposon elements to mammalian cells via DNA delivery.

In one example, we assayed the full R2 element R2-1_GFo (Repbase; Kojima et al *PLoS One* 11, e0163496 (2015)) from the medium ground finch, *Geospiza fortis* ("R2GFo"). In another example, we assayed the full R4 element R4_AL (Repbase; Burke et al *Nucleic Acids Res.* 23, 4628–34 (1995)) from the large roundworm, *Ascaris lumbricoides* ("R4Al"). Because non-LTR R2 and R4 elements are not present in the human genome and are thought to be highly site-specific, the ability of retrotransposons to accurately and efficiently integrate itself into the human genome would demonstrate the capability to perform genomic targeted integration.

Plasmids harboring R2Gfo (PLV033) or R4Al (PLV462) were designed for cis integration of the R2Gfo or R4Al elements as in previous examples. Plasmids were synthesized such that the wildtype element was flanked by its native un-translated regions (UTRs) and 100bp of homology to its rDNA target (Figure 22). The element expression was driven by the mammalian CMV promoter. We introduced each 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, 0.5µl transfection reagent and 80ng DNA was mixed in 10µl Opti-MEM and incubated for 15 minutes at room temperature. The transfection mixture was then added to the medium of the seeded cells. Three days after

transfection, genomic DNA was extracted for retrotransposition assays. R2Tg was also delivered in parallel in the same format to serve as a comparison.

ddPCR was performed to confirm integration and assess integration efficiency. A Taqman probe was designed to the 3'UTR portion of each element. A forward primer was synthesized to bind directly upstream of the probe, and a reverse primer was synthesized to bind the rDNA. Thus, amplification of the expected product across the integration junction would degrade the probe and create a fluorescent signal. The results of the ddPCR copy number analysis (in comparison to reference gene RPP30) are shown in Figure 23. R2Gfo integration achieved a mean copy number of 0.21 integrants/genome in this experiment. R4Al achieved a mean copy number of 0.085 integrants/genome.

**Example 29: Integration of retrotransposons into human fibroblasts**

This example describes the cis integration of R2Tg into human fibroblasts. Briefly, a plasmid designed to integrate R2Tg in cis was synthesized such that R2Tg was flanked by its native UTRs and homologous sequence to its rDNA target as in previous examples. 0.5µg PLV014 (wild-type) and PLV072 (EN mutant) plasmids were transfected into 100,000 human dermal fibroblasts isolated from neonatal foreskin (HDFn, C0045C, ThermoFisher Scientific) respectively using the Neon transfection system. Two programs were performed, each in duplicate. The setting for Program 1 was 1700V pulse voltage, 20ms pulse width, and 1 pulse number. The setting for Program 2 was 1400V pulse voltage, 20ms pulse width, and 2 pulse number. Both programs achieved 95% transfection efficiency measured using plasmid encoding the EGFP. Three days post transfection, genomic DNA was extracted for the ddPCR assay.

ddPCR was performed to confirm integration and assess integration efficiency. A Taqman probe was designed to the 3' UTR portion of the R2Tg element. A forward primer was synthesized to bind directly upstream of the probe, and a reverse primer was synthesized to bind the rDNA. Thus, amplification of the expected product across the integration junction would degrade the probe and create a fluorescent signal. The results of ddPCR copy number analysis (in comparison to reference gene RPP30) are shown in Figure 24. Wild-type (WT) R2Tg integration achieved a mean copy number of 0.036 integrants/genome in this experiment, significantly higher than a control R2Tg plasmid with a point mutation abolishing endonuclease activity (EN).

**Example 30: Evaluation of DNA damage response upon retrotransposon transfection**

DNA damage (e.g., resulting from DSB formation or replication fork collapse) leads to the activation of p53, which among many other transcriptional responses, leads to the upregulation of p21, resulting in cell cycle arrest or apoptosis. Genome editing using CRISRP/Cas9 has been shown to activate p53 and p21, which is a potential safety and efficacy problem for CRISPR/based therapeutics. To establish whether R2Tg delivery to the cell leads to activation of p53 and p21, U2OS cells were seeded at a density of $4x10^4$ cells/well and transfected 24 hours later using the Fugene HD and Lipofectamine reagents with either 500 ng of R2Tg-WT plasmid or 500 ng of R2Tg-EN (a variant of R2Tg with a mutation in the endonuclease (EN) domain, rendering R2Tg inactive). To control for transfection efficiency, U2OS cells were also transfected with a plasmid expressing GFP. Lastly, as a positive control for p53 and p21 activation, U2OS cells were treated with one of the DNA damage-inducing agents etoposide (20μM) or bleomycin (10 μg/ml). The U2OS cells were collected 24 hours after transfection/treatment. Protein lysates were prepared in RIPA buffer and run on an SDS-PAGE

gel, followed by transfer to nitrocellulose, followed by probing with antibodies against p53 and p21, as well as Actin and Vinculin. As shown in Figure 25, no R2Tg-induced upregulation of p53 or p21 above the GFP plasmid control was detected in either transfection condition.

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, wherein one or both of (i) or (ii) are derived from an avian retrotransposase; 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.

2. The system of claim 1, wherein the avian retrotransposase is a retrotransposase from *Taeniopygia guttata*, *Geospiza fortis*, *Zonotrichia albicollis*, or *Tinamus guttatus*.

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, wherein at least one of (i) or (ii) is heterologous, 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.

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

(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 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 10 bases of 100% identity to a target DNA strand, 5' end of the template RNA.

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

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

7. The system of any of claims 1-5, wherein (a) and (b) are separate nucleic acids.

8. The system of any of the preceding claims, 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.

9. The system of any of claims 1-8, wherein the heterologous object sequence comprises an open reading frame in a 5' to 3' orientation on the template RNA.

10. The system of any of claims 1-8, wherein the heterologous object sequence comprises an open reading frame in a 3' to 5' orientation on the template RNA.

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

12. The system of claim 11, wherein the template RNA further comprises a 5' untranslated sequence.

13. The system of any of the preceding claims, wherein the template RNA further comprises a promoter operably linked to the heterologous object sequence.

14. The system of claim 13, wherein the promoter is disposed between the 5' untranslated sequence and the heterologous object sequence.

15. The system of claim 13, wherein the promoter is disposed between the 3' untranslated sequence that binds the polypeptide and the heterologous object sequence.

16. The system of any of claims 11-15, wherein the 5' untranslated sequence is a sequence of column 6 of Table 3, or a sequence having at least 80% identity thereto.

17. The system of any of claims 11-16, wherein the 3' untranslated sequence is a sequence of column 7 of Table 3, or a sequence having at least 80% identity thereto.

18. The system of any of the preceding claims, 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, or a storage protein.

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

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

21. A method of modifying a target DNA strand in a cell, comprising administering the system of any of claims 1-20 to the cell, thereby modifying the target DNA strand.

22. The method of claim 21, which results in the addition of at least 5 base pairs of exogenous DNA sequence to the genome of the cell.

23. The method of claim 21, which results in the addition of at least 100 base pairs of exogenous DNA sequence to the genome of the cell.

24. The method of any of claims 21-23, 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.

25. The method of any of claims 21-24, which results in about 50-100% of insertions of the heterologous object sequence into the target DNA being non-truncated.

26. The method of any of claims 21-25, wherein the nucleic acid of (a) is not integrated into the genome of the cell.

27. The method of any of claims 21-26, wherein the template RNA comprises at least 10 bases of 100% identity to the target DNA strand, at the 5' end of the template RNA.

28. The method of any of claims 21-27, wherein the template RNA comprises at least 10 bases of 100% identity to the target DNA strand, at the 3' end of the template RNA.

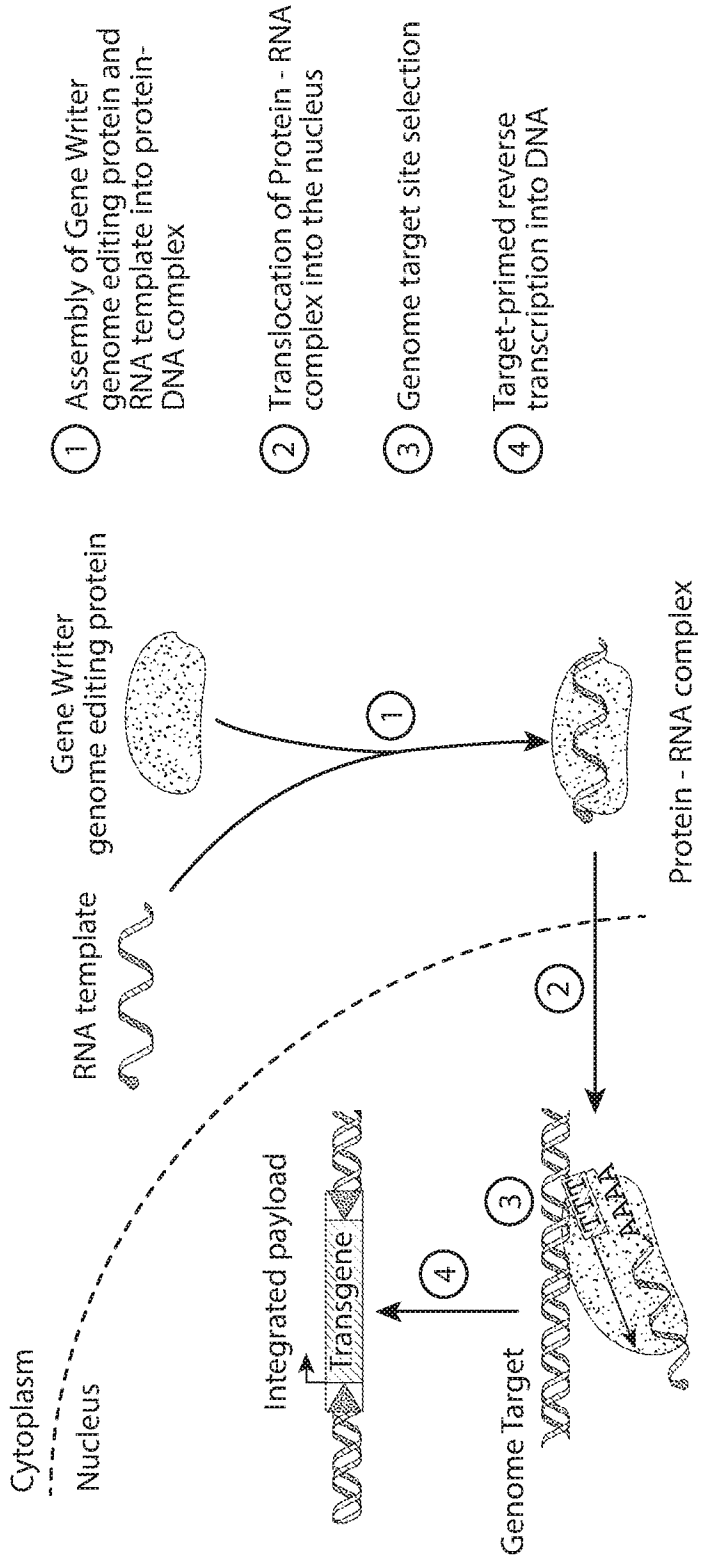

1 Assembly of Gene Writer genome editing protein and RNA template into protein-DNA complex
2 Translocation of Protein - RNA complex into the nucleus
3 Genome target site selection
4 Target-primed reverse transcription into DNA
Gene Writer genome editing protein
RNA template
Protein - RNA complex
Cytoplasm
Nucleus
Integrated payload
Transgene
Genome Target
TTTT
AAAA

Figure 1

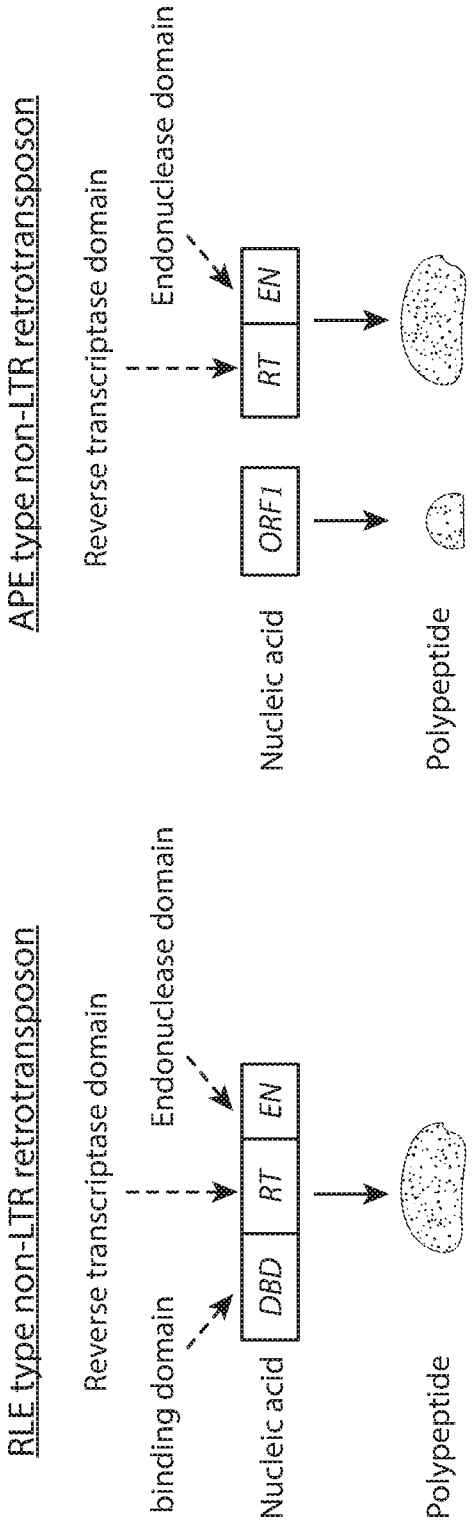

RLE type non-LTR retrotransposon
Reverse transcriptase domain
DNA binding domain
Endonuclease domain
Nucleic acid
DBD
RT
EN
Polypeptide
APE type non-LTR retrotransposon
Reverse transcriptase domain
Endonuclease domain
Nucleic acid
ORF1
RT
EN
Polypeptide

Figure 2

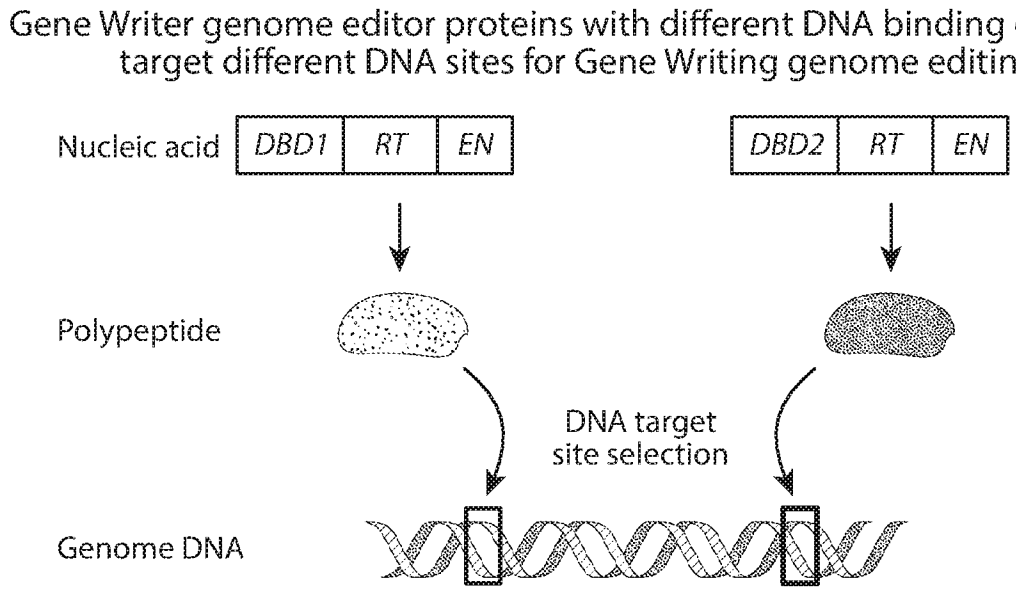

Gene Writer genome editor proteins with different DNA binding domains target different DNA sites for Gene Writing genome editing
Nucleic acid
DBD1
RT
EN
DBD2
RT
EN
Polypeptide
DNA target site selection
Genome DNA

Figure 3

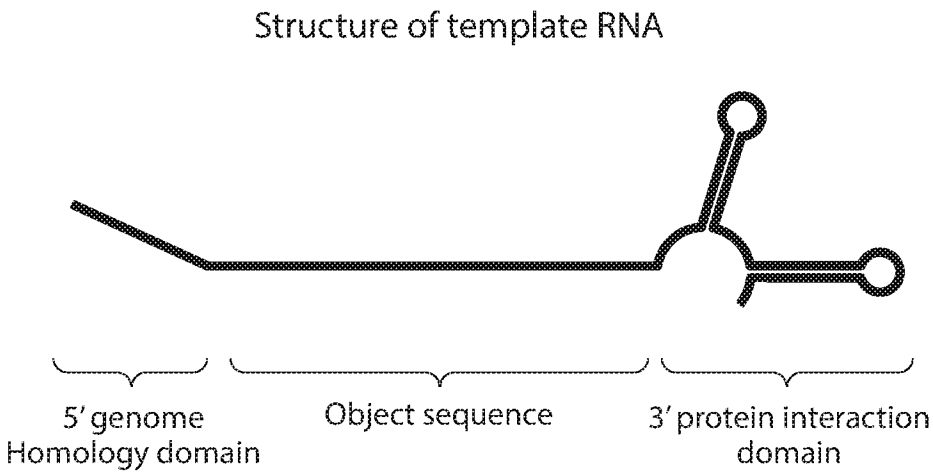

Structure of template RNA
5' genome
Homology domain
Object sequence
3' protein interaction
domain

Figure 4

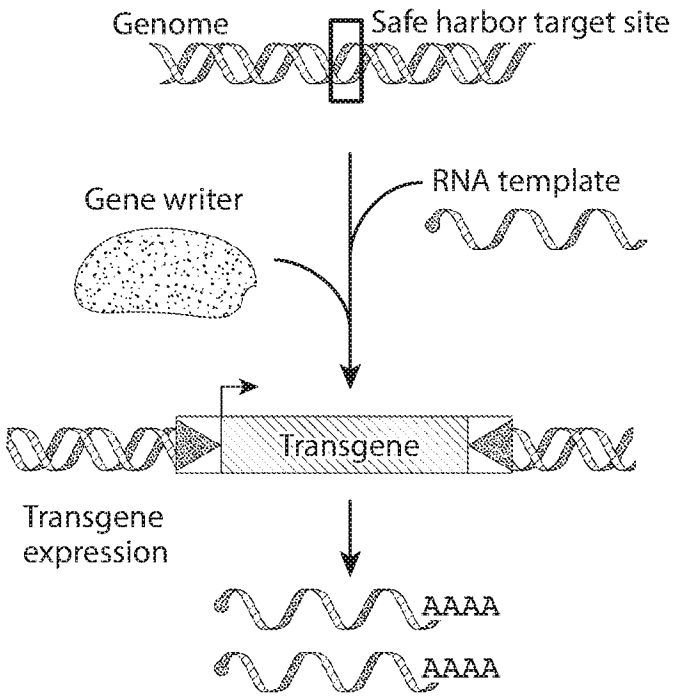



Figure 5

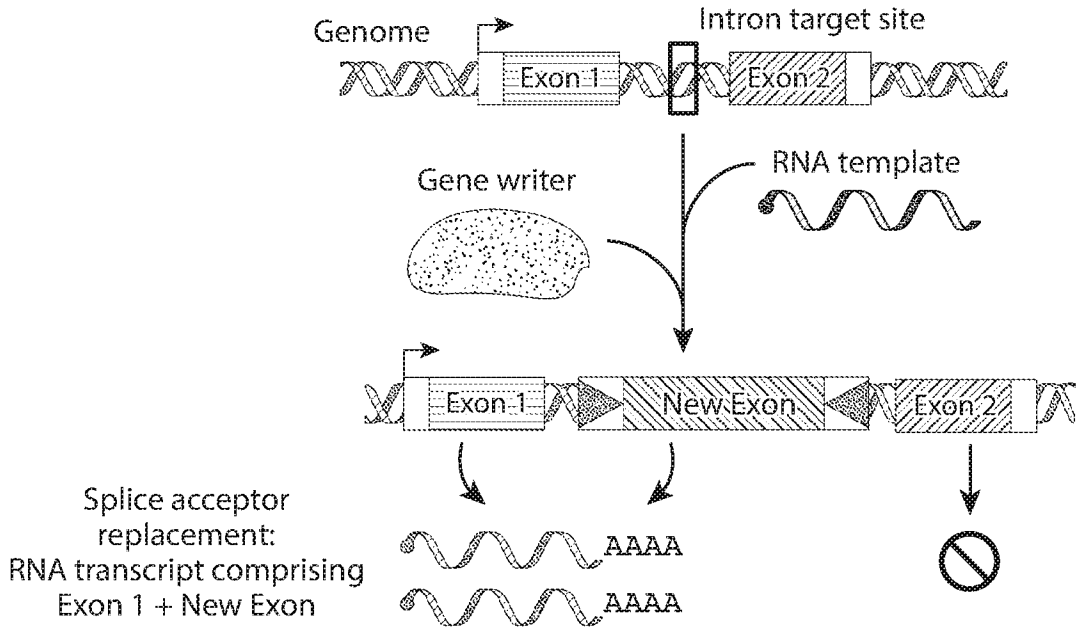



Figure 6

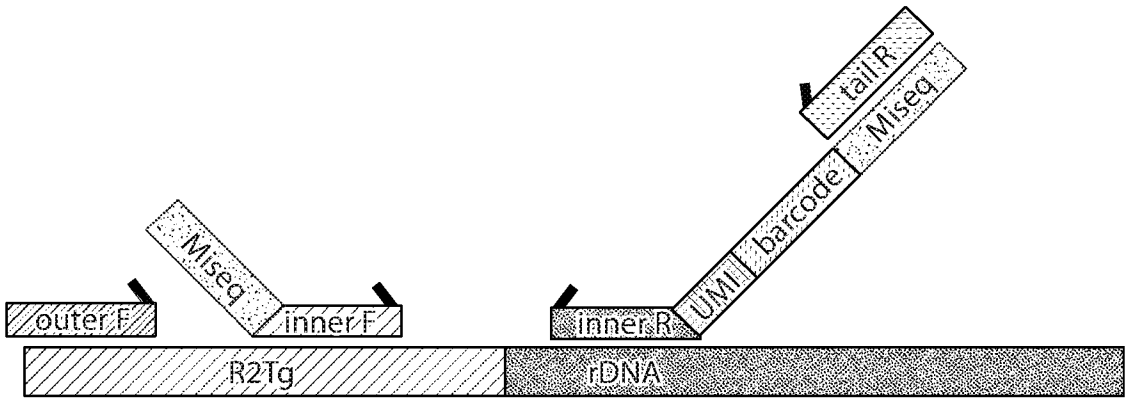

tail R
Miseq
barcode
UMI
Miseq
outer F
inner F
inner R
R2Tg
rDNA

Figure 7

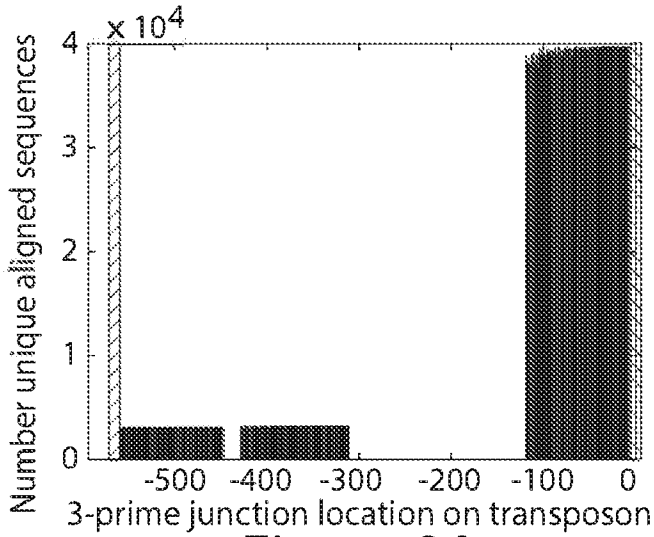



Figure 8A

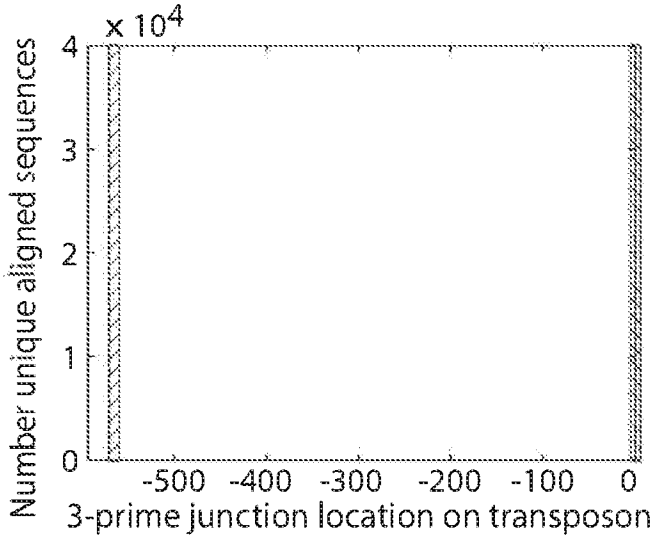



Figure 8B

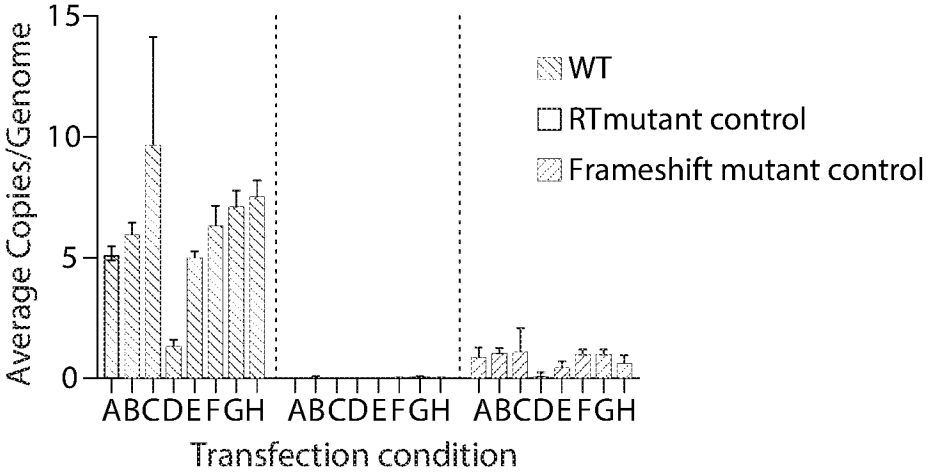



Figure 9

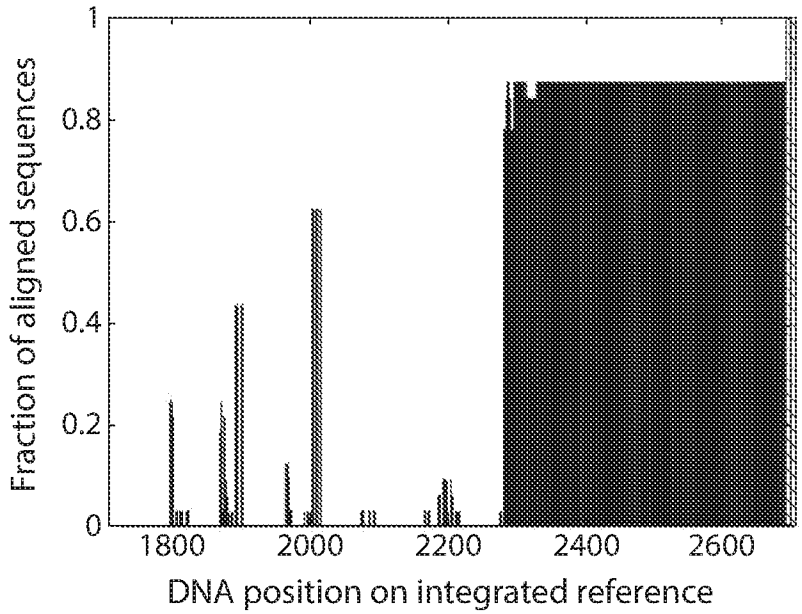



Figure 10

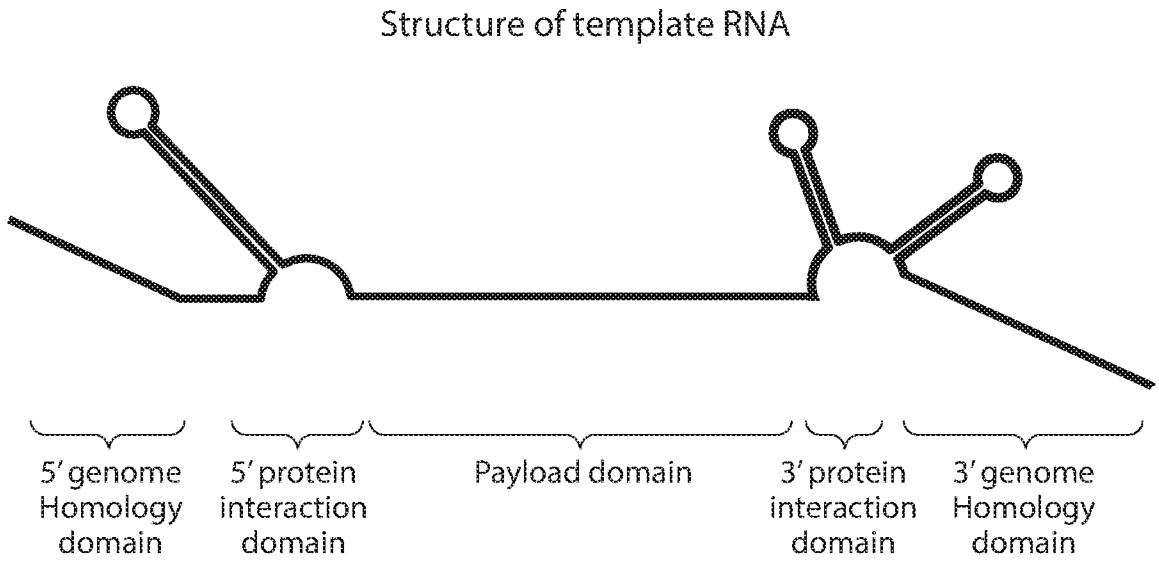



Figure 11

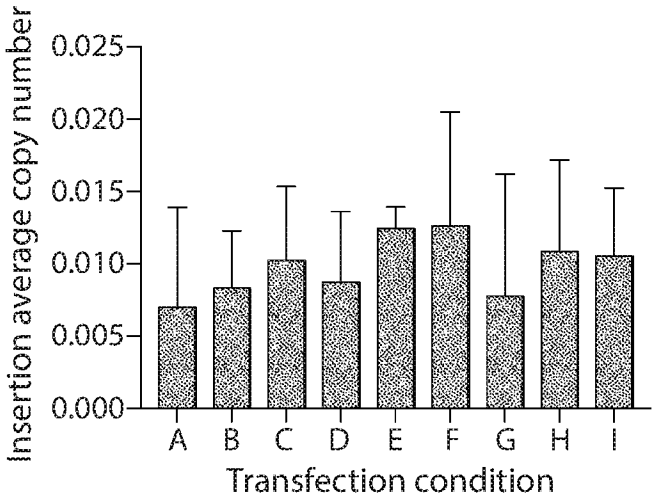



Figure 12

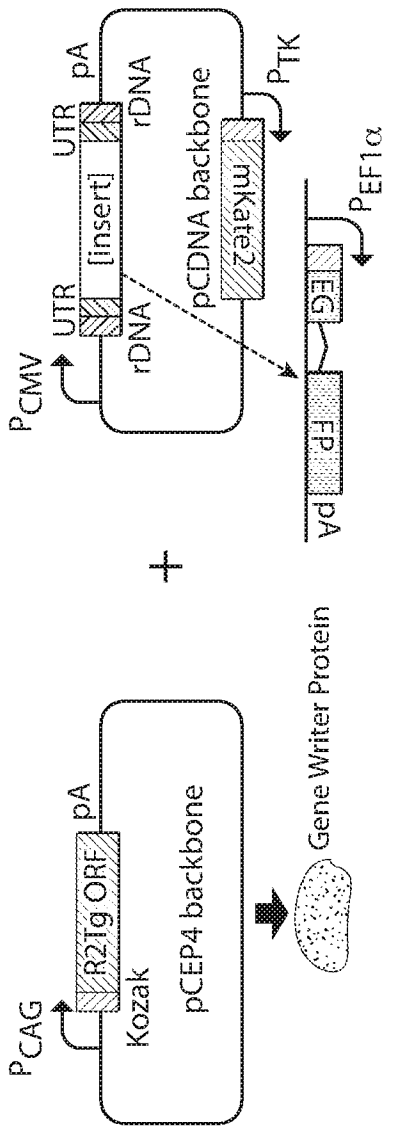

P_CAG
Kozak
R2Tg ORF
pA
pCEP4 backbone
Gene Writer Protein
+
P_CMV
UTR
rDNA
[insert]
UTR
pA
rDNA
pCDNA backbone
mKate2
P_TK
P_EF1α
EG
EP
pA

Figure 13

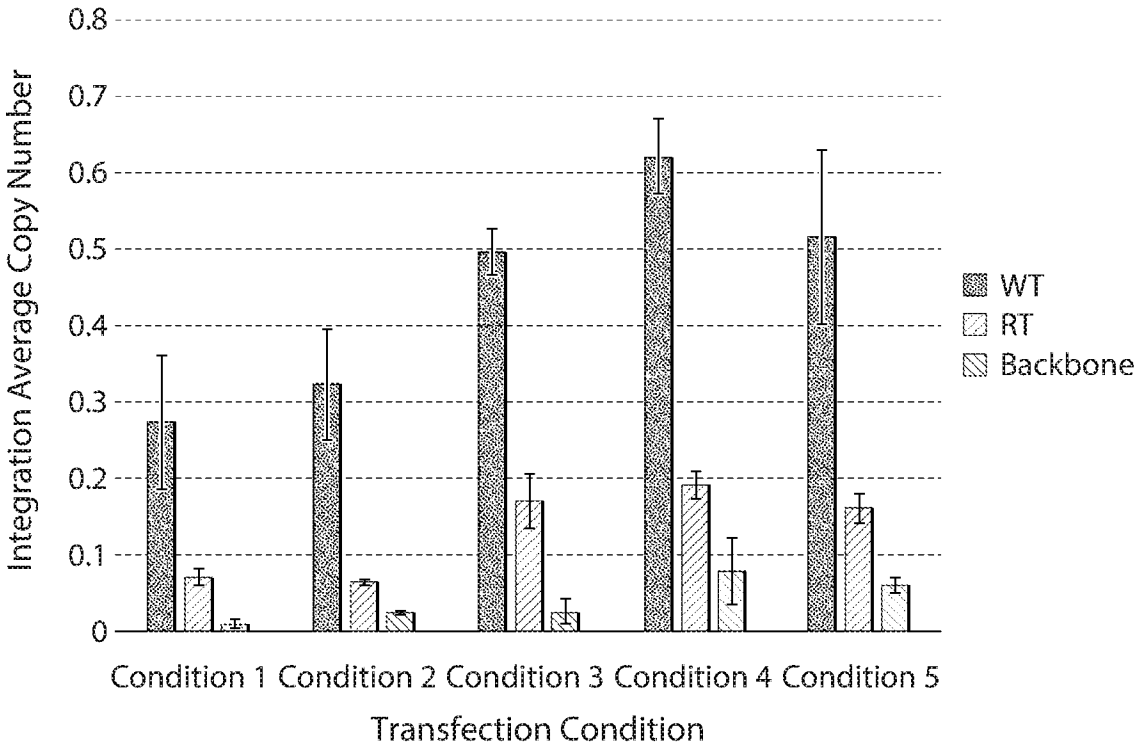

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Integration Average Copy Number
WT
RT
Backbone
Condition 1
Condition 2
Condition 3
Condition 4
Condition 5
Transfection Condition

Figure 14

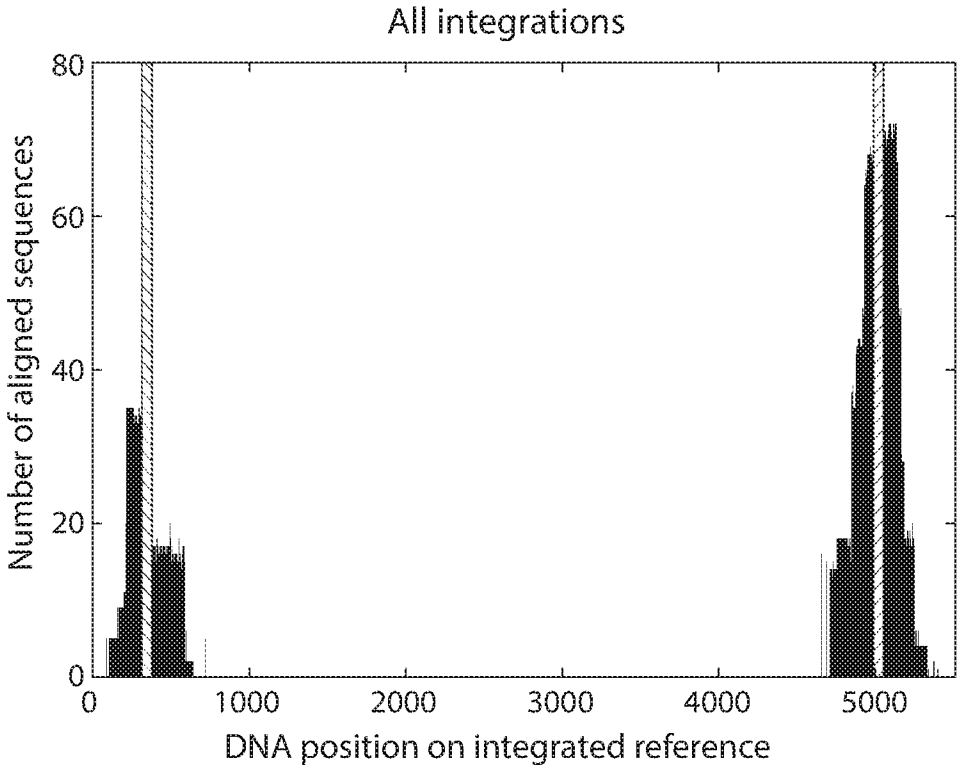



Figure 15A

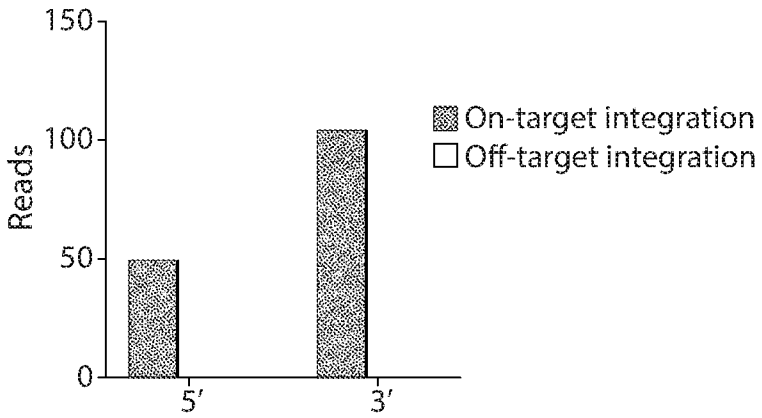



Figure 15B

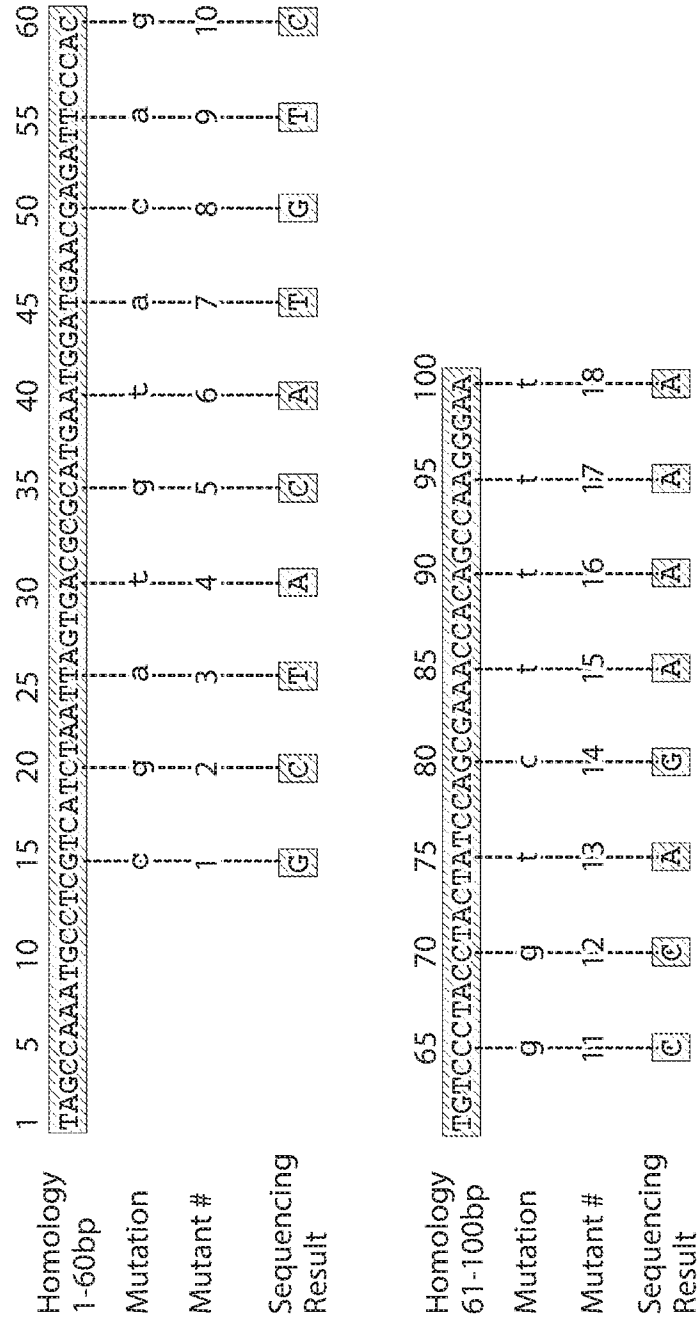

1 5 10 15 20 25 30 35 40 45 50 55 60
TAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATTCCCAC
Homology 1-60bp
Mutation
Mutant #
Sequencing Result
65 70 75 80 85 90 95 100
TGTCCCTACCTACTATCCAGCGAAACCACACCAGCCAAGGGAA
Homology 61-100bp
Mutation
Mutant #
Sequencing Result

Figure 16

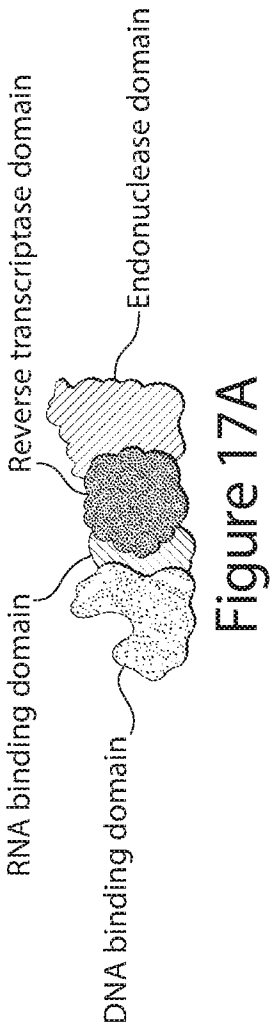



Figure 17A

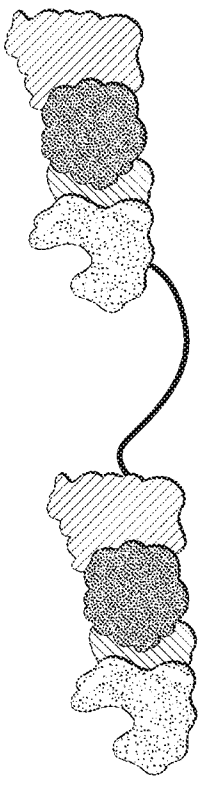


Figure 17B

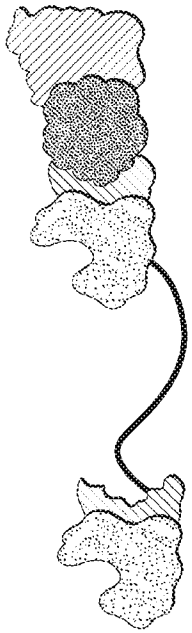


Figure 17C

Figure 17D

Figure 17E

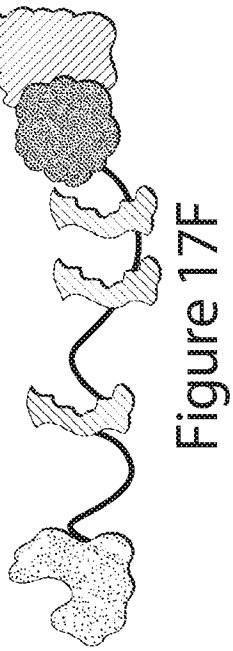


Figure 17F

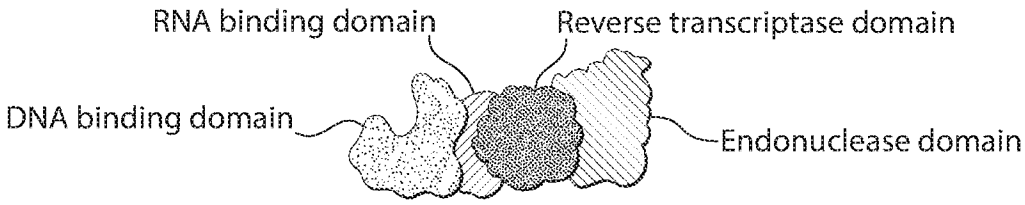
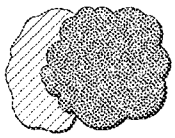



Figure 18A

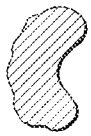
DNA binding
Zinc fingers
Cas9
Transcription factors

Figure 18B



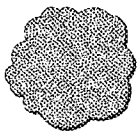
Full-length Reverse transcriptase
Different full-length RT from other R2
Different full-length RT combined with additional RNA binding domain

Figure 18C



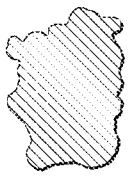
RNA binding
B-box protein
MS2 coat protein
dCas protein
UTR binding protein

Figure 18D



RT without RNA binding
Truncated RT from R2
HIV RT
AMV RT

Figure 18E



Endonuclease
Cas9 nickase
Cas ortholog
FokI
Restriction enyme

Figure 18F

FIG. 19

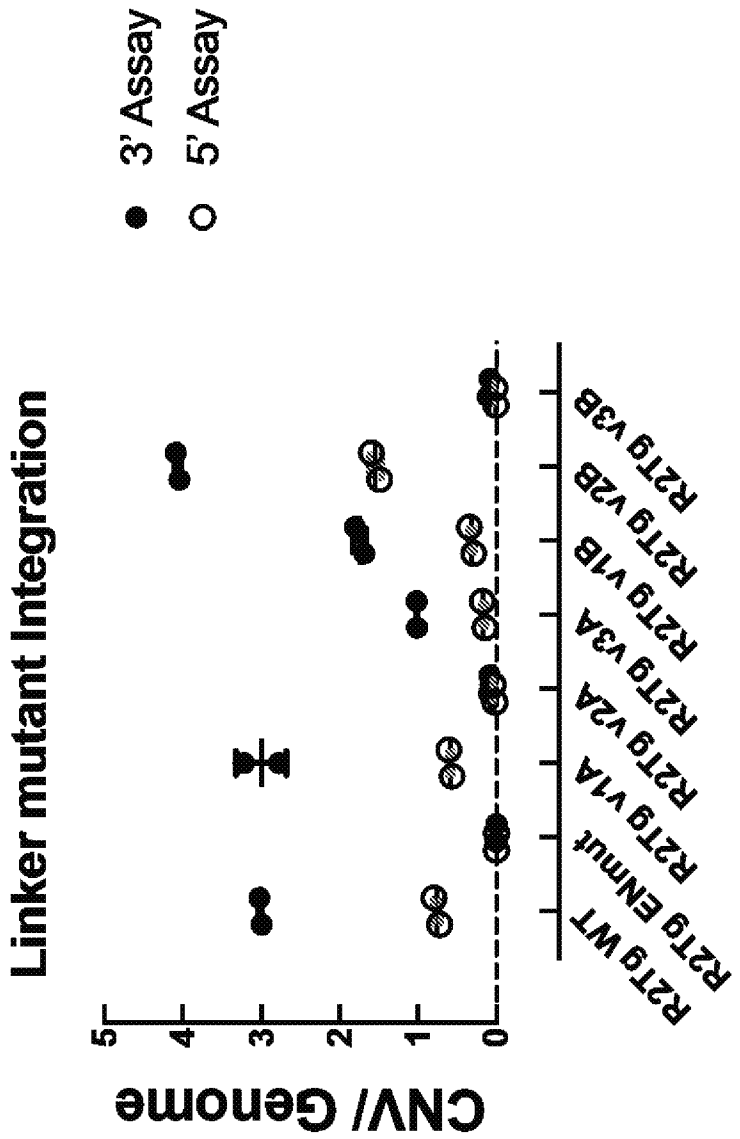
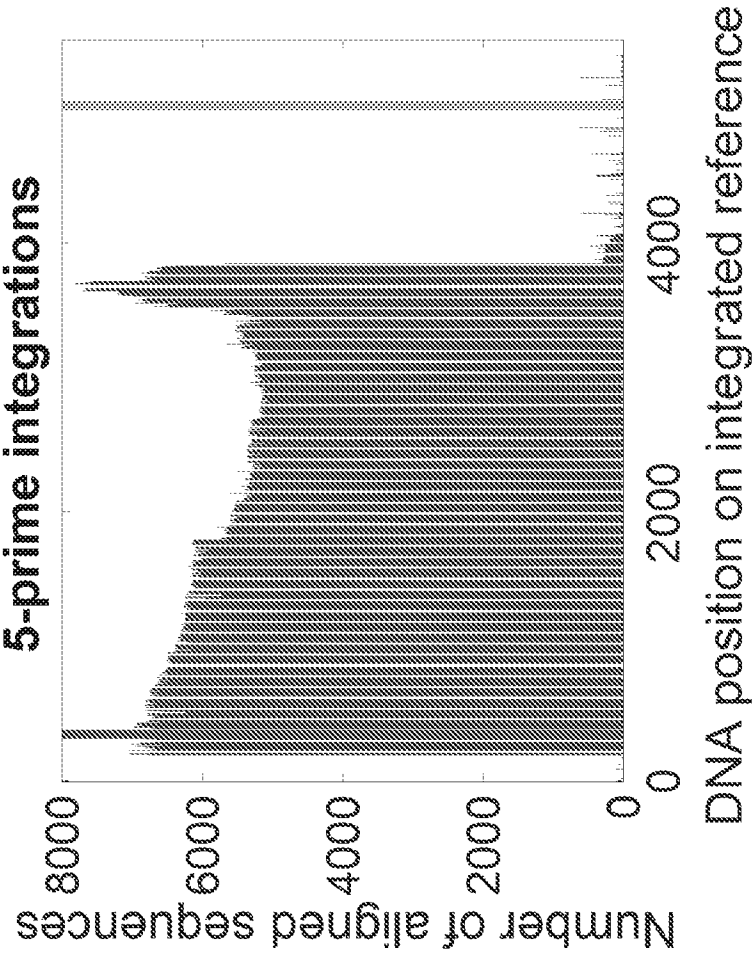

Linker mutant Integration
3' Assay
5' Assay
CNV/ Genome
0
1
2
3
4
5
R2TG WT
R2TG ENMut
R2TG V1A
R2TG V2A
R2TG V3A
R2TG V1B
R2TG V2B
R2TG V3B

FIG. 20A


5-prime integrations
Number of aligned sequences
8000
6000
4000
2000
0
0
2000
4000
DNA position on integrated reference

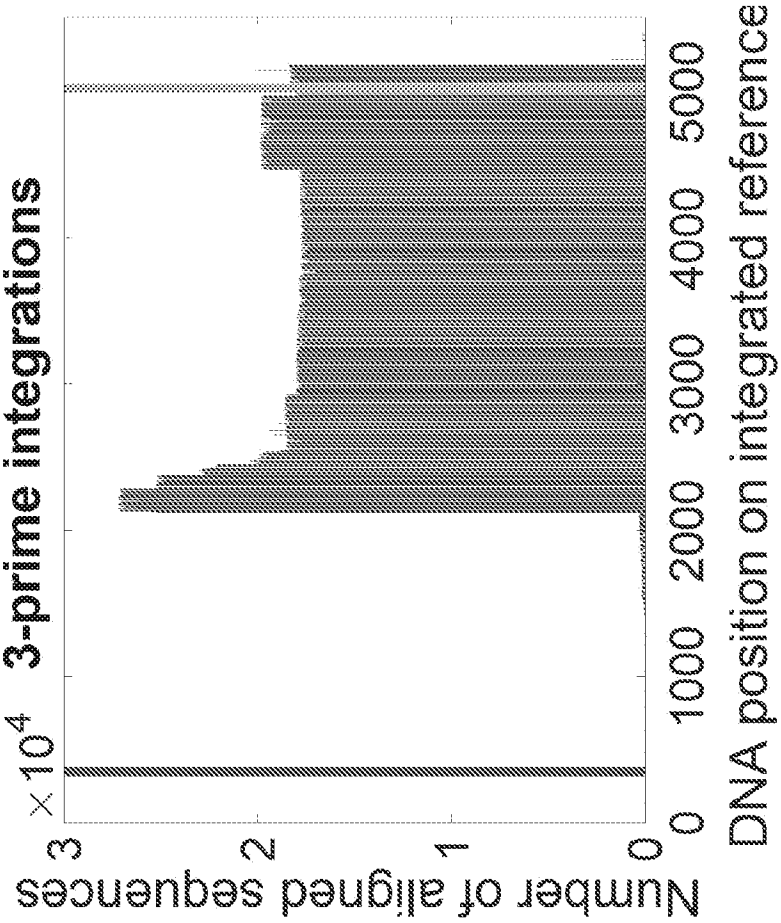

3-prime integrations
×10^4
Number of aligned sequences
0
1
2
3
0
1000
2000
3000
4000
5000
DNA position on integrated reference

FIG. 20B

FIG. 21A

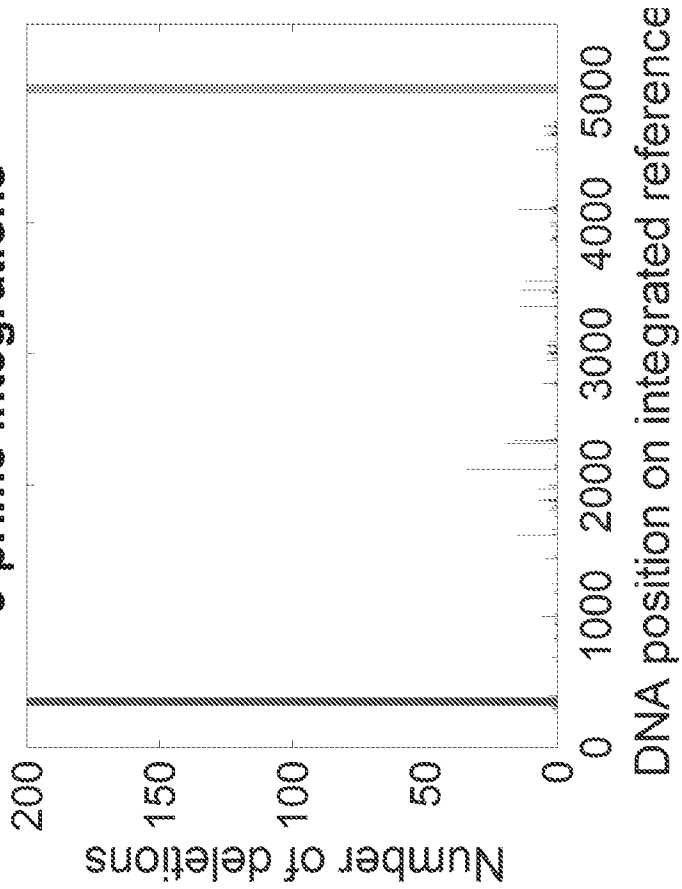

5-prime integrations
Number of deletions
200
150
100
50
0
0
1000
2000
3000
4000
5000
DNA position on integrated reference

FIG. 21B

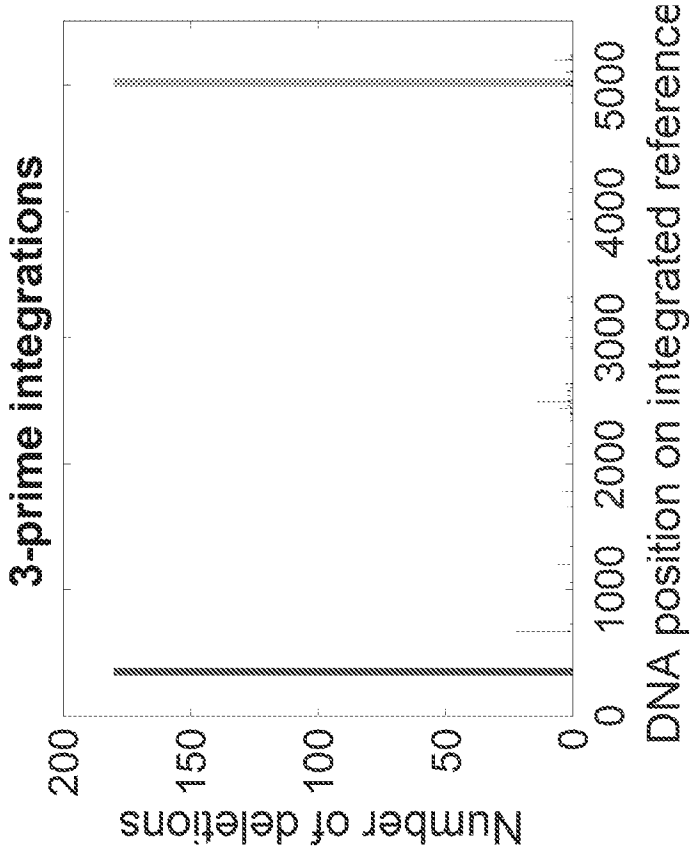

3-prime integrations
Number of deletions
200
150
100
50
0
0
1000
2000
3000
4000
5000
DNA position on integrated reference

FIG. 22

pR21Gfo_WT_28S_Annealing (9362 bp)

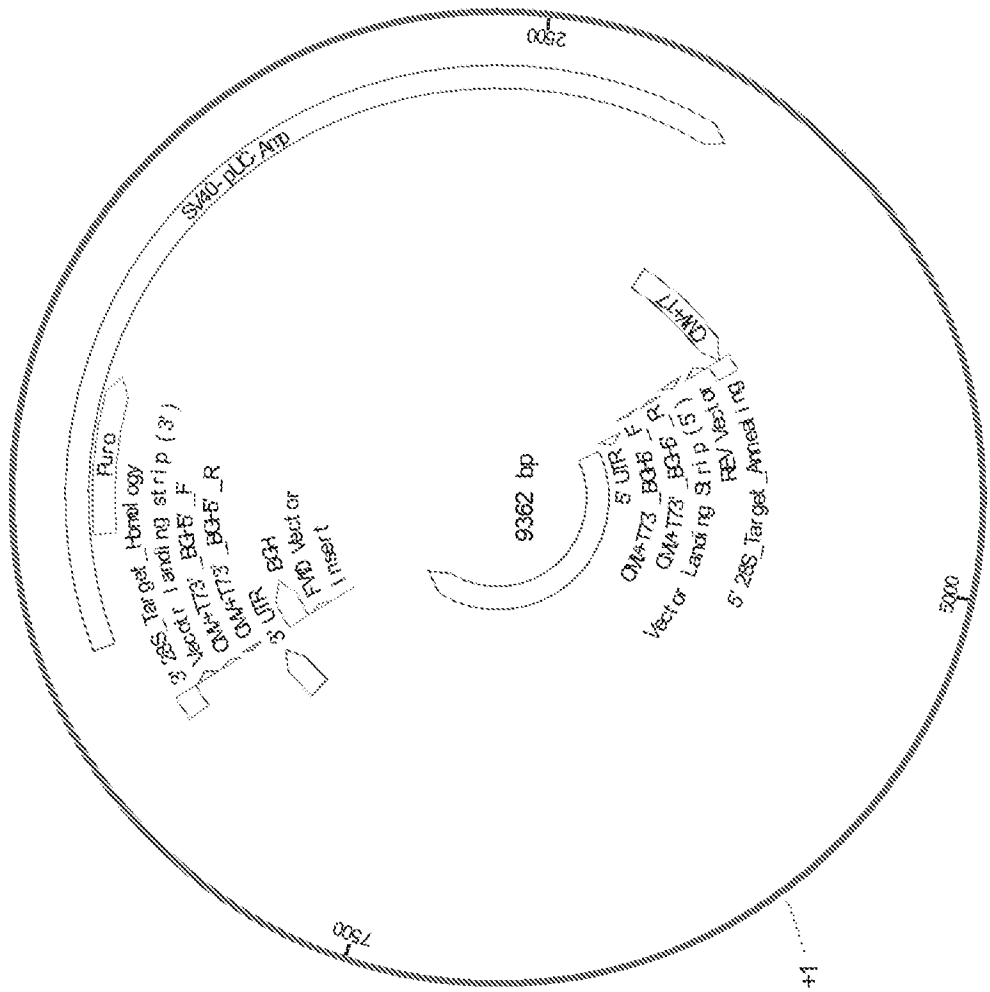


FIG. 23

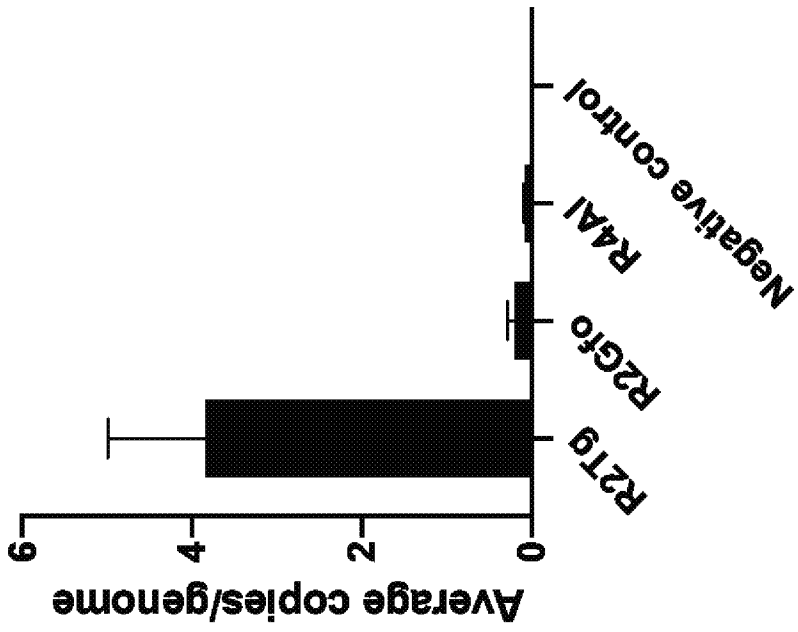

Average copies/genome
0
2
4
6
R2Tg
R2Gfp
R4AI
Negative control

FIG. 24

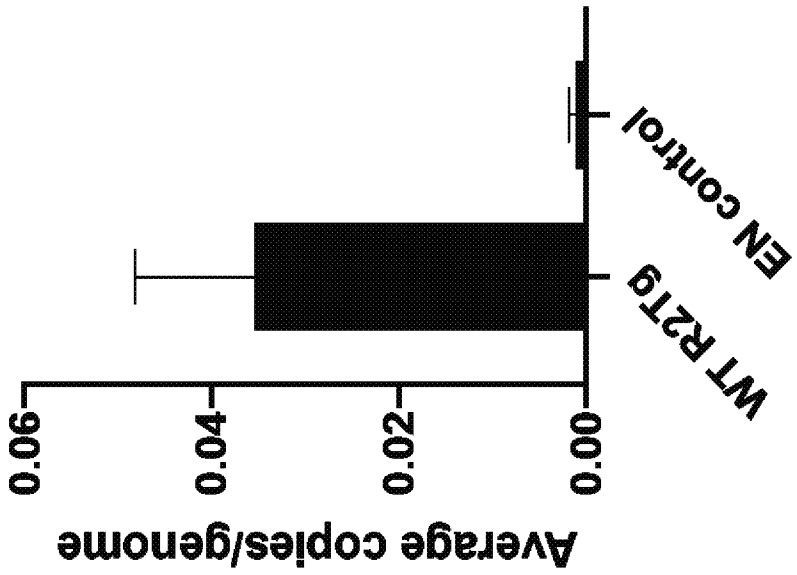
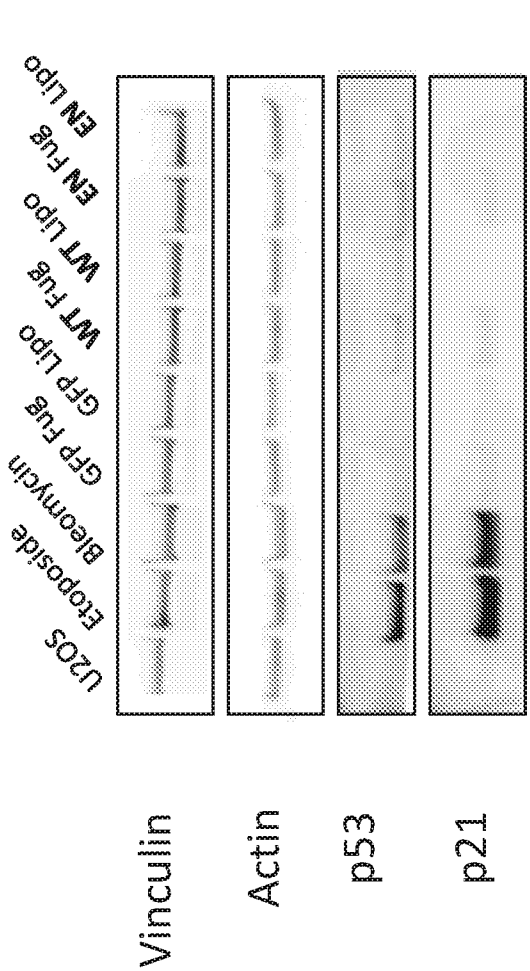


FIG. 25


R2Tg
U2OS
Etoposide
Bleomycin
GFP Fuse
GFP Lipo
WT Fuse
WT Lipo
EN Fuse
EN Lipo
Vinculin
Actin
p53
p21

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2019/048607

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/10 C12N9/12
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | CHRISTENSEN S ET AL: "Footprint of the Retrotransposon R2Bm Protein on its Target Site Before and After Cleavage", JOURNAL OF MOLECULAR BIOLOGY, ACADEMIC PRESS, UNITED KINGDOM, vol. 336, no. 5, 5 March 2004 (2004-03-05) , pages 1035-1045, XP004490163, ISSN: 0022-2836, DOI: 10.1016/J.JMB.2003.12.077 the whole document<br>-----<br>-/-- | 1-9,11, 12,16, 17,19-28 |

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 18 December 2019 | 15/01/2020 |

| Name and mailing address of the ISA/ | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Weinberg, Suzanna |

3

Form PCT/ISA/210 (second sheet) (April 2005)

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2019/048607

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | AHMET M. DENLI ET AL: "Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity", CELL, vol. 163, no. 3, 1 October 2015 (2015-10-01), pages 583-593, XP055651566, AMSTERDAM, NL ISSN: 0092-8674, DOI: 10.1016/j.cell.2015.09.025 the whole document | 1-28 |
| A | KENJI K. KOJIMA ET AL: "The Wide Distribution and Change of Target Specificity of R2 Non-LTR Retrotransposons in Animals", PLOS ONE, vol. 11, no. 9, 23 September 2016 (2016-09-23), page e0163496, XP055639736, DOI: 10.1371/journal.pone.0163496 the whole document | 1-28 |
| A | MEERS CHANCE ET AL: "DNA repair by RNA: Templated, or not templated, that is the question", DNA REPAIR, ELSEVIER, AMSTERDAM, NL, vol. 44, 16 May 2016 (2016-05-16), pages 17-21, XP029645040, ISSN: 1568-7864, DOI: 10.1016/J.DNAREP.2016.05.002 the whole document | 1-28 |
| A | CATHERINE M. ADE ET AL: "Evaluating different DNA binding domains to modulate L1 ORF2p-driven site-specific retrotransposition events in human cells", GENE., vol. 642, 1 February 2018 (2018-02-01), pages 188-198, XP55514879, NL ISSN: 0378-1119, DOI: 10.1016/j.gene.2017.11.033 the whole document | 1-28 |
| A | WO 2017/173004 A1 (MIKUNI TAKAYASU [US]; NISHIYAMA JUN [US]; YASUDA RYOHEI [US]) 5 October 2017 (2017-10-05) the whole document | 1-28 |
| A | EP 1 700 914 A1 (OSAKA IND PROMOTION ORG [JP]; NAT INST RADIOLOG [JP]) 13 September 2006 (2006-09-13) the whole document | 1-28 |

3

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2019/048607

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 2017173004 A1 | 05-10-2017 | NONE | |
| EP 1700914 A1 | 13-09-2006 | CA 2546848 A1 | 16-06-2005 |
| | | EP 1700914 A1 | 13-09-2006 |
| | | JP WO2005054463 A1 | 06-12-2007 |
| | | US 2008104723 A1 | 01-05-2008 |
| | | WO 2005054463 A1 | 16-06-2005 |