



US 20250257368A1

(19) **United States**

(12) **Patent Application Publication**

Hajjar et al.

(10) **Pub. No.: US 2025/0257368 A1**

(43) **Pub. Date:** Aug. 14, 2025

(54) **IN VITRO ASSEMBLY OF ANELLOVIRUS CAPSIDS ENCLOSING RNA**

(71) Applicant: **FLAGSHIP PIONEERING INNOVATIONS V, INC.**, Cambridge, MA (US)

(72) Inventors: **Roger Joseph Hajjar**, Lexington, MA (US); **Simon Delagrange**, Sudbury, MA (US); **Kurt Adam Swanson**, West Newton, MA (US)

(21) Appl. No.: **19/056,060**

(22) Filed: **Feb. 18, 2025**

**Related U.S. Application Data**

(63) Continuation of application No. 17/846,503, filed on Jun. 22, 2022, now abandoned, which is a continu-

ation of application No. PCT/US2021/064887, filed on Dec. 22, 2021.

(60) Provisional application No. 63/147,064, filed on Feb. 8, 2021, provisional application No. 63/130,360, filed on Dec. 23, 2020.

**Publication Classification**

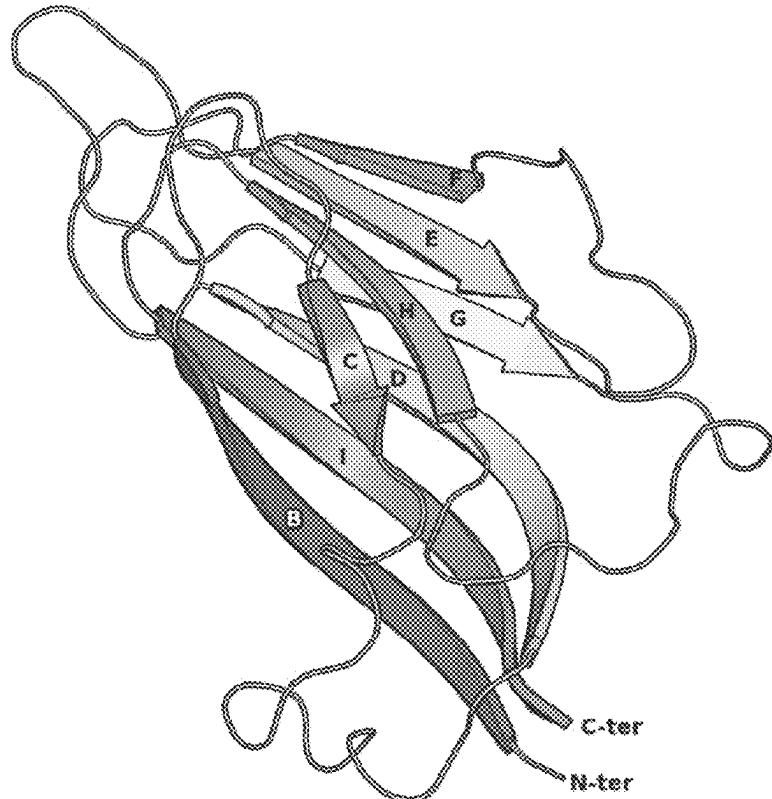
(51) **Int. Cl.**  
**C12N 15/86** (2006.01)

(52) **U.S. Cl.**  
CPC ..... **C12N 15/86** (2013.01)

(57) **ABSTRACT**

This invention relates generally to compositions for making anellovectors and uses thereof.

**Specification includes a Sequence Listing.**



**1. C - H - E - F**

**2. B - I - D - G**

**PRIMARY  
DNA CONTACTS**

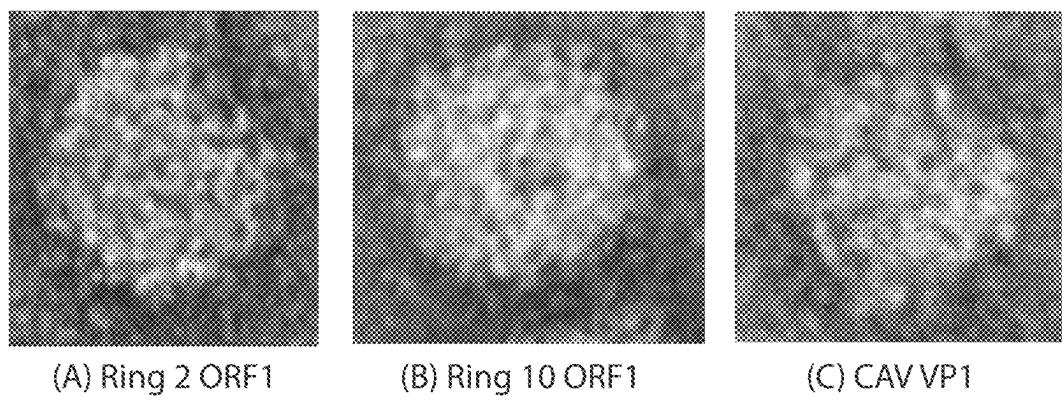
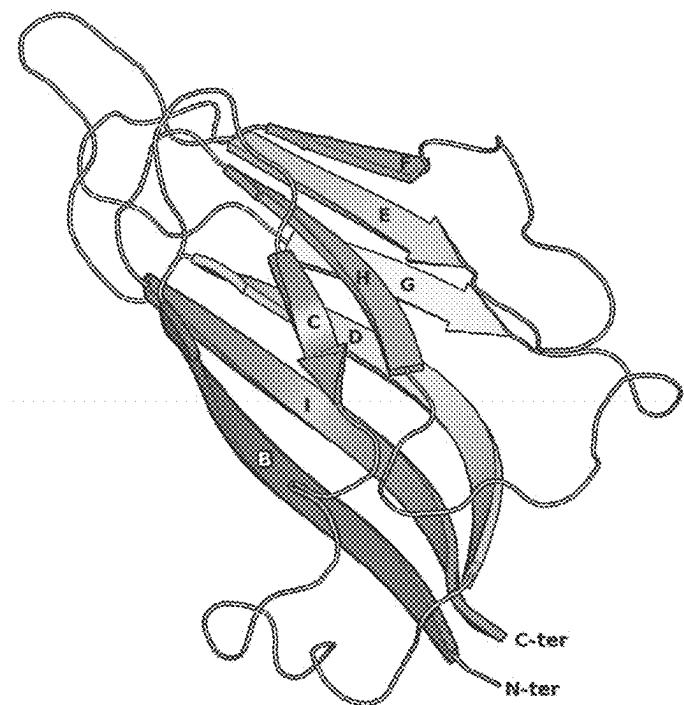


FIG. 1

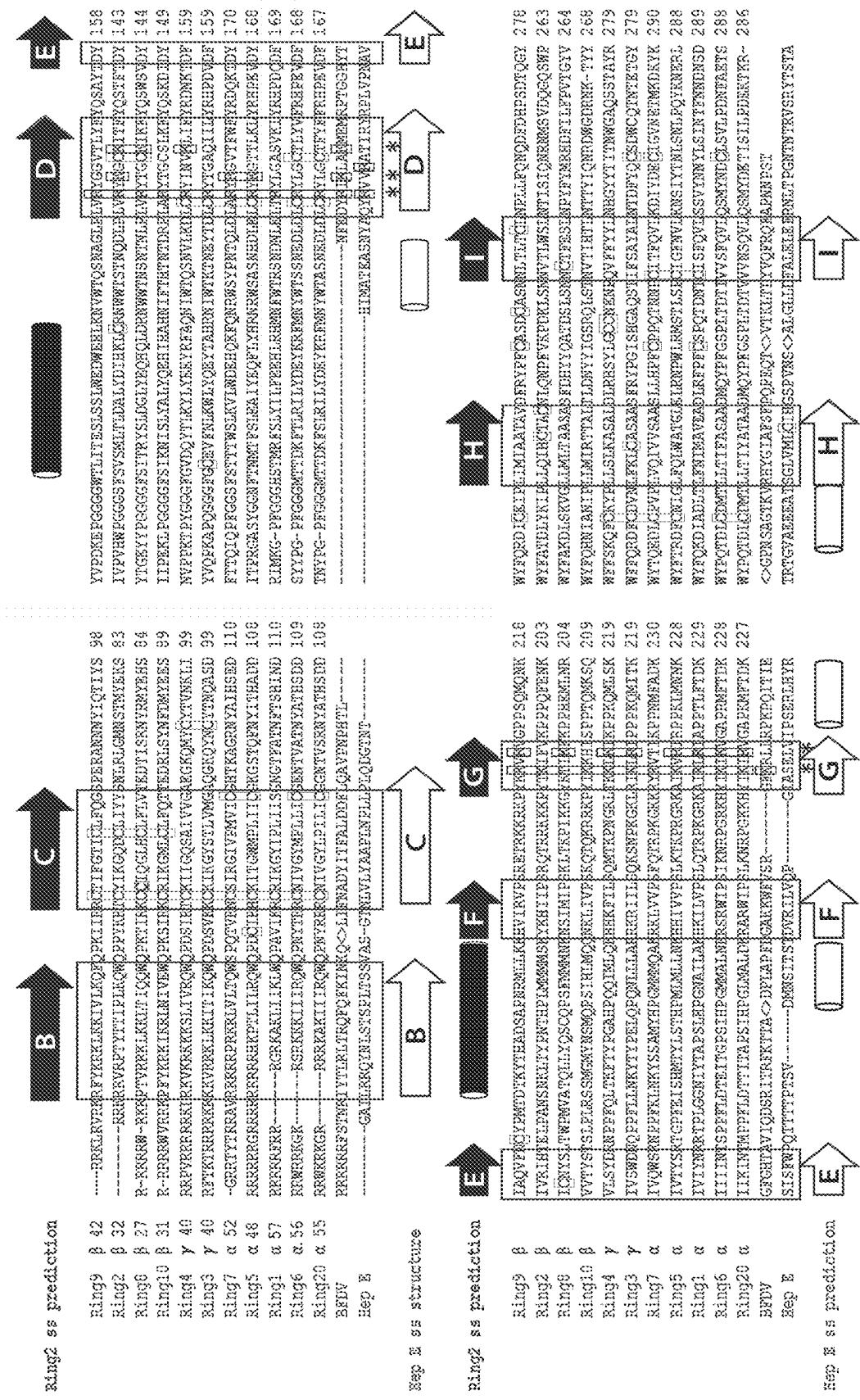


**1. C - H - E - F**

**2. B - I - D - G**

PRIMARY  
DNA CONTACTS

FIG. 2



三  
三  
三

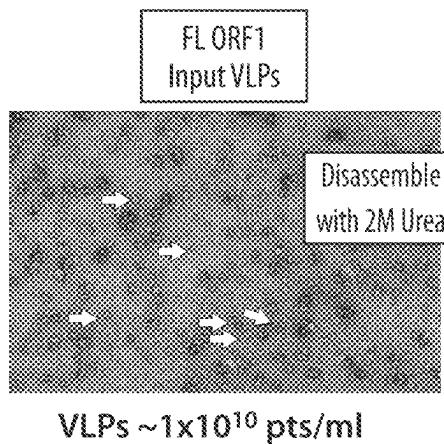


FIG. 4A

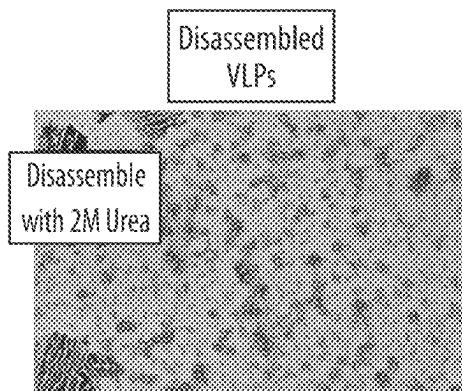


FIG. 4B

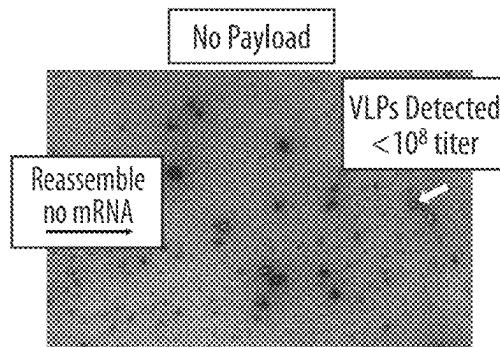


FIG. 4C

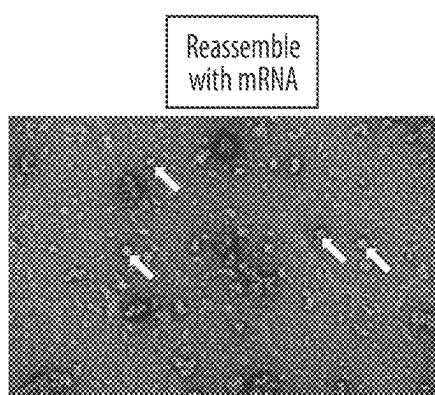


FIG. 4D

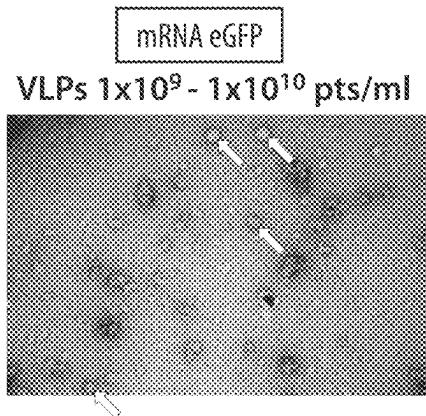


FIG. 4E

## IN VITRO ASSEMBLY OF ANELLOVIRUS CAPSIDS ENCLOSING RNA

### CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. patent application Ser. No. 17/846,503, filed Jun. 22, 2022, which is a continuation of PCT/US2021/064887, filed Dec. 22, 2021, which claims the benefit of U.S. Provisional Application No. 63/130,360, filed Dec. 23, 2020 and U.S. Provisional Application No. 63/147,064, filed Feb. 8, 2021. The contents of the aforesaid applications are hereby incorporated by reference in their entirety.

### SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. The Sequence Listing file, entitled V2057-701721\_SL.xml, was created on Apr. 16, 2025, and is 399,527 bytes in size.

### BACKGROUND

[0003] There is an ongoing need to develop compositions and methods for making suitable vectors to deliver therapeutic effectors to patients.

### SUMMARY

[0004] The present disclosure provides compositions and methods for producing an anellovector (e.g., a synthetic anellovector) that can be used as a delivery vehicle, e.g., for delivering genetic material, for delivering an effector, e.g., a payload, or for delivering a therapeutic agent or a therapeutic effector to a eukaryotic cell (e.g., a human cell or a human tissue). While naturally occurring Anellovirus has a DNA genome, the present disclosure provides anellovectors with a genetic element that comprises RNA.

[0005] An anellovector (e.g., produced using a composition or method as described herein) generally comprises a genetic element (e.g., a genetic element comprising or encoding an effector, e.g., an exogenous or endogenous effector, e.g., a therapeutic effector) encapsulated in a proteinaceous exterior (e.g., a proteinaceous exterior comprising an Anellovirus capsid protein, e.g., an Anellovirus ORF1 protein or a polypeptide encoded by an Anellovirus ORF1 nucleic acid, e.g., as described herein), which is capable of introducing the genetic element into a cell (e.g., a mammalian cell, e.g., a human cell). The genetic element may comprise RNA. In some embodiments, the anellovector is an infectious vehicle or particle comprising a proteinaceous exterior comprising a polypeptide encoded by an Anellovirus ORF1 nucleic acid (e.g., an ORF1 nucleic acid of Alphatorquevirus, Betatorquevirus, or Gammatorquevirus, e.g., an ORF1 of Alphatorquevirus clade 1, Alphatorquevirus clade 2, Alphatorquevirus clade 3, Alphatorquevirus clade 4, Alphatorquevirus clade 5, Alphatorquevirus clade 6, or Alphatorquevirus clade 7, e.g., as described herein). The genetic element of an anellovector of the present disclosure is typically a circular and/or single-stranded RNA molecule (e.g., circular and single stranded) having a protein binding sequence that binds to the proteinaceous exterior enclosing it, or a polypeptide attached thereto, which may facilitate enclosure of the genetic element within the proteinaceous exterior and/or enrichment of the genetic element, relative to

other nucleic acids, within the proteinaceous exterior. In some embodiments, the genetic element of an anellovector is produced using a baculovirus, nucleic acid construct (e.g., a bacmid and/or donor vector), insect cell, and/or animal cell line, e.g., as described herein.

[0006] In some instances, the genetic element comprises or encodes an effector (e.g., comprises a nucleic acid effector, such as a non-coding RNA, or encodes a polypeptide effector, e.g., a protein), e.g., which can be expressed in a target cell. In some embodiments, the effector is a therapeutic agent or a therapeutic effector, e.g., as described herein. In some embodiments, the effector is an endogenous effector or an exogenous effector, e.g., exogenous to a wild-type Anellovirus or a target cell. In some embodiments, the effector is exogenous to a wild-type Anellovirus or a target cell. In some embodiments, the anellovector can deliver an effector into a cell by contacting the cell and introducing a genetic element encoding the effector into the cell, such that the effector is made or expressed by the cell. In certain instances, the effector is an endogenous effector (e.g., endogenous to the target cell but, e.g., provided in increased amounts by the anellovector). In other instances, the effector is an exogenous effector. The effector can, in some instances, modulate a function of the cell or modulate an activity or level of a target molecule in the cell. For example, the effector can decrease levels of a target protein in the cell (e.g., as described in Examples 3 and 4 of PCT/US19/65995). In another example, the anellovector can deliver and express an effector, e.g., an exogenous protein, *in vivo* (e.g., as described in Examples 19 and 28 of PCT/US19/65995). Anellovectors can be used, for example, to deliver genetic material to a target cell, tissue or subject; to deliver an effector to a target cell, tissue or subject; or for treatment of diseases and disorders, e.g., by delivering an effector that can operate as a therapeutic agent to a desired cell, tissue, or subject.

[0007] In some embodiments, the compositions and methods described herein can be used to produce the genetic element of a synthetic anellovector, e.g., in a host cell. A synthetic anellovector has at least one structural difference compared to a wild-type virus (e.g., a wild-type Anellovirus, e.g., as described herein), e.g., a deletion, insertion, substitution, modification (e.g., enzymatic modification), relative to the wild-type virus. Generally, synthetic anellovectors include an exogenous genetic element enclosed within a proteinaceous exterior, which can be used for delivering the genetic element, or an effector (e.g., an exogenous effector or an endogenous effector) encoded therein (e.g., a polypeptide or nucleic acid effector), into eukaryotic (e.g., human) cells. In embodiments, the anellovector does not cause a detectable and/or an unwanted immune or inflammatory response, e.g., does not cause more than a 1%, 5%, 10%, 15% increase in a molecular marker(s) of inflammation, e.g., TNF-alpha, IL-6, IL-12, IFN, as well as B-cell response e.g. reactive or neutralizing antibodies, e.g., the anellovector may be substantially non-immunogenic to the target cell, tissue or subject.

[0008] In some embodiments, the compositions and methods described herein can be used to produce the genetic element of an anellovector comprising: (i) a proteinaceous exterior comprising an ORF1 molecule; and (ii) a genetic element comprising RNA; wherein the genetic element is enclosed within the proteinaceous exterior. In some embodiments, the genetic element consists of at least 10%, 20%,

25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% RNA. In some embodiments, the genetic element does not comprise DNA. In some embodiments, the genetic element does not comprise ssDNA. Alternatively, in some embodiments, the genetic element comprises a DNA region. In some embodiments, a DNA or RNA molecule described herein comprises one or more modified nucleotides (e.g., a base modification, sugar modification, or backbone modification). In some embodiments, the genetic element is a single-stranded. In some embodiments, the genetic element comprises a double stranded region. In some embodiments the genetic element is a linear polypeptide. Alternatively, in some embodiments, the genetic element is a circular polynucleotide. In some embodiments, the nucleic acid sequence is codon-optimized, e.g., for expression in an insect cell. In some embodiments, at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the codons in the nucleic acid sequence are codon-optimized, e.g., for expression in an insect cell. In some embodiments, the nucleic acid sequence is codon-optimized, e.g., for expression in a mammalian (e.g., human) cell. In some embodiments, at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the codons in the nucleic acid sequence are codon-optimized, e.g., for expression in a mammalian (e.g., human) cell. In some embodiments, the genetic element is about 10-20, 20-30, 30-40, 50-60, 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, or 4000-4500 nucleotides in length. In some embodiments, the genetic element is at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, or 4500 nucleotides in length.

[0009] In some embodiments, the compositions and methods described herein can be used to produce the genetic element of an infectious (e.g., to a human cell) Anellovector, vehicle, or particle comprising a capsid (e.g., a capsid comprising an Anellovirus ORF, e.g., ORF1, polypeptide) encapsulating a genetic element comprising a protein binding sequence that binds to the capsid and a heterologous (to the Anellovirus) sequence encoding a therapeutic effector. In embodiments, the Anellovector is capable of delivering the genetic element into a mammalian, e.g., human, cell.

[0010] In an aspect, the invention features a method of making an anellovector by *in vitro* assembly. In some embodiments, a method of making an anellovector comprises: (a) providing a mixture comprising: (i) a genetic element comprising RNA, and (ii) an ORF1 molecule; and (b) incubating the mixture under conditions suitable for enclosing the genetic element within a proteinaceous comprising the ORF1 molecule, thereby making an anellovector; optionally wherein the mixture is not comprised in a cell. In some embodiments, a method further comprises, prior to the providing of (a), expressing the ORF1 molecule, e.g., in a host cell (e.g., an insect cell or a mammalian cell). In some embodiments, the expressing comprises incubating a host cell (e.g., an insect cell or a mammalian cell) comprising a nucleic acid molecule (e.g., a baculovirus expression vector) encoding the ORF1 molecule under conditions suitable for producing the ORF1 molecule. In some embodiments, a method further comprises, prior to the providing of (a), purifying the ORF1 molecule expressed by the host cell.

[0011] In some embodiments, anellovectors, as described herein, can be used as effective delivery vehicles for introducing an agent, such as an effector described herein, to a target cell, e.g., a target cell in a subject to be treated therapeutically or prophylactically.

[0012] In an aspect, the invention features a pharmaceutical composition comprising an anellovector (e.g., a synthetic anellovector) as described herein. In some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier or excipient. In embodiments, the pharmaceutical composition comprises a unit dose comprising about  $10^5$ - $10^{14}$  (e.g., about  $10^6$ - $10^{11}$ ,  $10^7$ - $10^{12}$ ,  $10^8$ - $10^{11}$ , or  $10^9$ - $10^{10}$ ) genome equivalents of the anellovector per kilogram of a target subject. In some embodiments, the pharmaceutical composition comprising the preparation will be stable over an acceptable period of time and temperature, and/or be compatible with the desired route of administration and/or any devices this route of administration will require, e.g., needles or syringes. In some embodiments, the pharmaceutical composition is formulated for administration as a single dose or multiple doses. In some embodiments, the pharmaceutical composition is formulated at the site of administration, e.g., by a healthcare professional. In some embodiments, the pharmaceutical composition comprises a desired concentration of anellovector genomes or genomic equivalents (e.g., as defined by number of genomes per volume).

[0013] In an aspect, the invention features a method of treating a disease or disorder in a subject, the method comprising administering to the subject an anellovector, e.g., a synthetic anellovector, e.g., as described herein.

[0014] In an aspect, the invention features a method of delivering an effector or payload (e.g., an endogenous or exogenous effector) to a cell, tissue or subject, the method comprising administering to the subject an anellovector, e.g., a synthetic anellovector, e.g., as described herein, wherein the anellovector comprises a nucleic acid sequence encoding the effector. In some embodiments, the payload is a nucleic acid. In some embodiments, the payload is a polypeptide.

[0015] In an aspect, the invention features a method of delivering an anellovector to a cell, comprising contacting the anellovector, e.g., a synthetic anellovector, e.g., as described herein, with a cell, e.g., a eukaryotic cell, e.g., a mammalian cell, e.g., *in vivo* or *ex vivo*.

[0016] Additional features of any of the aforesaid anellovectors, compositions or methods include one or more of the following enumerated embodiments.

[0017] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following enumerated embodiments.

#### Enumerated Embodiments

[0018] 1. An anellovector comprising:

[0019] a) a proteinaceous exterior comprising an ORF1 molecule;

[0020] b) a genetic element comprising RNA,

[0021] wherein the genetic element is enclosed within the proteinaceous exterior.

[0022] 2. The anellovector of embodiment 1, wherein the genetic element consists of at least 10%, 20%, 25%, 30%,

40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% RNA.

[0023] 3. The anellovector of embodiment 1 or 2, wherein the RNA comprises one or more chemical modifications.

[0024] 4. The anellovector of any of the preceding embodiments, wherein the genetic element consists of or consists essentially of RNA.

[0025] 5. The anellovector of any of the preceding embodiments, wherein the anellovector does not comprise DNA.

[0026] 6. The anellovector of any of the preceding embodiments, wherein the anellovector does not comprise ssDNA.

[0027] 7. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a DNA region.

[0028] 8. The anellovector of any of the preceding embodiments, wherein all nucleotides of the DNA region are chemically modified.

[0029] 9. The anellovector of embodiment 7 or 8, wherein the DNA region is covalently linked to the RNA of the genetic element.

[0030] 10. The anellovector of any of the preceding embodiments, wherein at least a portion of the DNA region hybridizes to at least a portion of the RNA of the genetic element.

[0031] 11. The anellovector of any of the preceding embodiments, wherein the DNA region is 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 nucleotides in length.

[0032] 12. The anellovector of any of the preceding embodiments, wherein the genetic element is single stranded.

[0033] 13. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a double stranded region (e.g., a region of RNA pairing with RNA or a DNA pairing with RNA).

[0034] 14. The anellovector of any of the preceding embodiments, wherein the genetic element is linear.

[0035] 15. The anellovector of any of the preceding embodiments, wherein the genetic element is circular.

[0036] 16. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a first region and a second region that can hybridize with the first region.

[0037] 17. The anellovector of any of the preceding embodiments, wherein the genetic element does not comprise a 5' end or a 3' end.

[0038] 18. The anellovector of any of embodiments 15-17, wherein the genetic element does not comprise one or both of a free phosphate and a free sugar.

[0039] 19. The anellovector of any of embodiments 15-18, wherein every phosphate in the genetic element is covalently linked to a first sugar by a first oxygen atom comprised by the phosphate and a second sugar by a second oxygen atom comprised by the phosphate.

[0040] 20. The anellovector of any of embodiments 15-19, wherein every sugar in the genetic element is covalently linked to a first phosphate by a first carbon atom comprised by the sugar and a second phosphate by a second carbon atom comprised by the sugar.

[0041] 21. The anellovector of any of embodiments 15-20, wherein the genetic element was produced by circularizing a linear RNA.

[0042] 22. The anellovector of any of the preceding embodiments, wherein the genetic element is about 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, or 4000-4500 nucleotides in length.

[0043] 23. The anellovector of any of the preceding embodiments, wherein the genetic element is at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, or 4500 nucleotides in length.

[0044] 24. The anellovector of any of the preceding embodiments, wherein the genetic element binds the ORF1 molecule.

[0045] 25. The anellovector of any of the preceding embodiments, wherein the genetic element binds a jelly roll domain comprised by the ORF1 molecule.

[0046] 26. The anellovector of any of the preceding embodiments, wherein the genetic element binds an arginine-rich domain comprised by the ORF1 molecule.

[0047] 27. The anellovector of any of the preceding embodiments, wherein the anellovector comprises a plurality of genetic elements, e.g., at least 2, 3, 4, 5, 10, 20, 30, 40, 50, or 60 genetic elements.

[0048] 28. The anellovector of embodiment 27, wherein the genetic elements of the plurality each comprise the same sequence.

[0049] 29. The anellovector of embodiment 27, wherein the genetic elements of the plurality comprise different sequences.

[0050] 30. The anellovector of any of the preceding embodiments, wherein the genetic element encodes an exogenous effector.

[0051] 31. The anellovector of embodiment 30, wherein the sequence encoding the exogenous effector is at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, or 3000 nucleotides in length.

[0052] 32. The anellovector of any of embodiments 30-31, wherein the exogenous effector comprises a therapeutic effector (e.g., a polypeptide or a nucleic acid molecule).

[0053] 33. The anellovector of any of embodiments 30-32, wherein the exogenous effector comprises a human protein, or a polypeptide comprising an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

[0054] 34. The anellovector of any of embodiments 30-33, wherein the exogenous effector comprises a nucleic acid molecule.

[0055] 35. The anellovector of any of embodiments 30-34, wherein the exogenous effector comprises a noncoding nucleic acid molecule, e.g., a functional RNA, e.g., an mRNA, miRNA, or siRNA).

[0056] 36. The anellovector of any of embodiments 30-35, wherein the genetic element is an mRNA molecule encoding the exogenous effector (e.g., a peptide or polypeptide, e.g., a therapeutic peptide or polypeptide).

[0057] 37. The anellovector of embodiment 36, wherein the noncoding nucleic acid molecule is a human noncoding nucleic acid molecule, or a nucleic acid molecule comprising

ing a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

[0058] 38. The anellovector of any of embodiments 30-37, wherein the exogenous effector comprises a cytosolic polypeptide or cytosolic peptide (e.g., a DPP-4 inhibitor, an activator of GLP-1 signaling, or an inhibitor of neutrophil elastase, or a functional fragment thereof).

[0059] 39. The anellovector of embodiment 38, wherein the exogenous effector comprises a regulatory intracellular polypeptide.

[0060] 40. The anellovector of any of embodiments 30-39, wherein the exogenous effector comprises a secreted polypeptide or peptide (e.g., a cytokine, an antibody molecule, a hormone, a growth factor, or a clotting-associated factor, or a functional fragment thereof).

[0061] 41. The anellovector of any of embodiments 30-40, wherein the exogenous effector comprises a protein replacement therapeutic.

[0062] 42. The anellovector of any of embodiments 30-41, wherein the exogenous effector comprises an enzyme.

[0063] 43. The anellovector of any of embodiments 30-42, wherein the exogenous effector comprises erythropoietin (EPO) or human growth hormone (hGH), or a functional fragment thereof.

[0064] 44. The anellovector of any of embodiments 30-43, wherein the exogenous effector comprises a component of a gene editing system (e.g., a component of a CRISPR system, e.g., a Cas9, Cpf1, or a functional fragment thereof).

[0065] 45. The anellovector of any of the preceding embodiments, wherein the RNA comprises chemically modified RNA, e.g., as described herein.

[0066] 46. The anellovector of any of the preceding embodiments, wherein the RNA comprises a cap.

[0067] 47. The anellovector of any of the preceding embodiments, wherein the RNA comprises a poly-A tail, e.g., at least about 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 adenosines in length.

[0068] 48. The anellovector of any of the preceding embodiments, wherein the RNA lacks a poly-A tail, e.g., comprises no more than 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 sequential adenosines.

[0069] 49. The anellovector of any of the preceding embodiments, wherein the proteinaceous exterior comprises about 60 (e.g., about 40, 50, 60, 70, or 80) copies of the ORF1 molecule.

[0070] 50. The anellovector of any of the preceding embodiments, wherein the jelly roll domains of the ORF1 molecules face the interior of the proteinaceous exterior.

[0071] 51. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises an amino acid sequence as listed in any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

[0072] 52. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises an arginine-rich region, e.g., comprising the amino acid sequence of an arginine-rich region as listed in any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

[0073] 53. The anellovector of embodiment 52, wherein the arginine-rich region comprises at least about 70% (e.g.,

at least about 70%, 75%, 80%, 85%, 90%, 95%, or 100%) basic residues (e.g., arginine or lysine).

[0074] 54. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises a jelly roll domain, e.g., comprising the amino acid sequence of a jelly roll domain as listed in any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

[0075] 55. The anellovector of embodiment 54, wherein the jelly roll domain comprises one or more (e.g., 1, 2, 3, or 4) of the following characteristics:

[0076] (i) at least 30% (e.g., at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90%, or more) of the amino acids of the jelly-roll domain are part of one or more  $\beta$ -sheets;

[0077] (ii) the secondary structure of the jelly-roll domain comprises at least four (e.g., at least 4, 5, 6, 7, 8, 9, 10, 11, or 12)  $\beta$ -strands; and/or

[0078] (iii) the tertiary structure of the jelly-roll domain comprises at least two (e.g., at least 2, 3, or 4)  $\beta$ -sheets; and/or

[0079] (iv) the jelly-roll domain comprises a ratio of  $\beta$ -sheets to  $\alpha$ -helices of at least 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1.

[0080] 56. The anellovector of embodiment 54, wherein the jelly roll domain comprises two (3-sheets, e.g., arranged in antiparallel orientation relative to each other.

[0081] 57. The anellovector of embodiment 54, wherein the jelly roll domain comprises eight  $\beta$ -strands.

[0082] 58. The anellovector of any of embodiments 52-57, wherein the jelly roll domain comprises a region having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the sequence of a D  $\beta$ -strand, e.g., as shown in FIG. 3.

[0083] 59. The anellovector of embodiment 52, wherein the D  $\beta$ -strand comprises 1, 2, or 3, or more basic residues (e.g., arginine or lysine).

[0084] 60. The anellovector of any of embodiments 52-59, wherein the jelly roll domain comprises a region having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the sequence of a G  $\beta$ -strand, e.g., as shown in FIG. 3.

[0085] 61. The anellovector of embodiment 52, wherein the G  $\beta$ -strand comprises at least about 1, 2, or 3, or more basic residues (e.g., arginine or lysine).

[0086] 62. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises an N22 domain, e.g., comprising the amino acid sequence of an N22 domain as listed in any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

[0087] 63. The anellovector of any of the preceding embodiments, wherein the N22 domain comprises the amino acid sequence YNPX<sup>2</sup>DGX<sup>2</sup>N (SEQ ID NO: 829), wherein X" is each independently a contiguous sequence of any n amino acids.

[0088] 64. The anellovector of any of the preceding embodiments, wherein the N22 domain comprises a first beta strand and a second beta strand flanking the amino acid sequence YNPX<sup>2</sup>DGX<sup>2</sup>N (SEQ ID NO: 829), e.g., wherein the first beta strand comprises the tyrosine (Y) residue of the amino acid sequence YNPX<sup>2</sup>DGX<sup>2</sup>N (SEQ

ID NO: 829) and/or wherein the second beta strand comprises the second asparagine (N) residue (from N to C) of the amino acid sequence YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829).

[0089] 65. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises a C-terminal domain, e.g., comprising the amino acid sequence of a C-terminal domain as listed in any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

[0090] 66. The anellovector of any of the preceding embodiments, wherein the genetic element lacks a sequence encoding an Anellovirus ORF1 protein (e.g., as described herein).

[0091] 67. The anellovector of any of the preceding embodiments, wherein the genetic element lacks a sequence encoding an Anellovirus ORF2 protein (e.g., as described herein).

[0092] 68. The anellovector of any of the preceding embodiments, wherein the genetic element lacks a sequence encoding an Anellovirus ORF3 protein (e.g., as described herein).

[0093] 69. The anellovector of any of the preceding embodiments, wherein the anellovector is configured to deliver the genetic element to a cell (e.g., a eukaryotic cell, e.g., a mammalian cell, e.g., a human cell).

[0094] 70. The anellovector of embodiment 69, wherein a population of at least 1000 of the anellovectors is capable of delivering at least about 100 copies (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 5000, 10,000, 50,000, 100,000, 500,000, or 1,000,000 copies) of the genetic element into one or more of the cells.

[0095] 71. The anellovector of embodiment 69 or 70, wherein a population of the anellovectors (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 genome equivalents of the genetic element per cell) is capable of delivering the genetic element into at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more of a population of the cells.

[0096] 72. The anellovector of any of embodiments 69-71, wherein a population of the anellovectors (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 genome equivalents of the genetic element per cell) is capable of delivering at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 8,000,  $1 \times 10^4$ ,  $1 \times 10^5$ ,  $1 \times 10^6$ ,  $1 \times 10^7$  or greater copies of the genetic element per cell to a population of the cells.

[0097] 73. The anellovector of any of embodiments 69-72, wherein a population of the anellovectors (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 genome equivalents of the genetic element per cell) is capable of delivering 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 5-10, 10-20, 20-50, 50-100, 100-1000, 1000-104,  $1 \times 10^4$ - $1 \times 10^5$ ,  $1 \times 10^4$ - $1 \times 10^6$ ,  $1 \times 10^4$ - $1 \times 10^7$ ,  $1 \times 10^5$ - $1 \times 10^6$ ,  $1 \times 10^5$ - $1 \times 10^7$ , or  $1 \times 10^6$ - $1 \times 10^7$  copies of the genetic element per cell to a population of the cells.

[0098] 74. The anellovector of any of the preceding embodiments, wherein the anellovector selectively delivers the effector to, or is present at higher levels in (e.g., preferentially accumulates in), a desired cell type, tissue, or organ (e.g., bone marrow, blood, heart, GI, skin, photoreceptors in the retina, epithelial linings, or pancreas).

[0099] 75. The anellovector of any of the preceding embodiments, wherein the genetic element is protected from or resistant to digestion by an RNase (e.g., by the proteinaceous exterior).

[0100] 76. The anellovector of any of the preceding embodiments, wherein the genetic element enclosed within the proteinaceous exterior is resistant to endonuclease digestion, e.g., to RNase digestion.

[0101] 77. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a promoter element.

[0102] 78. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a protein binding sequence.

[0103] 79. The anellovector of embodiment 78, wherein the protein binding sequence is capable of binding to the ORF1 molecule.

[0104] 80. A composition comprising a plurality of the anellovectors of any of the preceding embodiments.

[0105] 81. The composition of embodiment 80, wherein at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of proteinaceous exteriors comprising an ORF1 molecule in the composition comprise at least one copy of the genetic element.

[0106] 82. The composition of embodiment 80, wherein at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of proteinaceous exteriors comprising an ORF1 molecule in the composition comprise at least one copy of an anellovector genetic element.

[0107] 83. The composition of any of embodiments 80-82, wherein the composition comprises at least  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ , or  $10^8$  of the same anellovector.

[0108] 84. The composition of any of embodiments 80-83, wherein the plurality comprises at least  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ ,  $10^{11}$ ,  $10^{12}$ ,  $10^{13}$ ,  $10^{14}$ , or  $10^{15}$  anellovectors (e.g., copies of the anellovector); or wherein the composition comprises at least  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ ,  $10^{11}$ ,  $10^{12}$ ,  $10^{13}$ ,  $10^{14}$ , or  $10^{15}$  anellovector genomes per mL.

[0109] 85. The composition of any of embodiments 80-84, having one or more (e.g., 1, 2, 3, 4, 5, or 6) of the following characteristics:

[0110] a) the composition meets a pharmaceutical or good manufacturing practices (GMP) standard;

[0111] b) the composition was made according to good manufacturing practices (GMP);

[0112] c) the composition has a pathogen level below a predetermined reference value, e.g., is substantially free of pathogens;

[0113] d) the composition has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants (e.g., a denaturant, e.g., urea);

[0114] e) the composition has a predetermined level of non-infectious particles or a predetermined ratio of particles:infectious units (e.g., <300:1, <200:1, <100:1, or <50:1), or

[0115] f) the pharmaceutical composition has low immunogenicity or is substantially non-immunogenic, e.g., as described herein;

[0116] optionally wherein the composition comprises urea at a concentration of less than 0.1M, 0.2M, 0.3M, 0.4M, 0.5M, 0.6M, 0.7M, 0.8M, 0.9M, 1M, 1.1M, 1.2M, 1.3M, 1.5M, 1.5M, 1.6M, 1.7M, 1.8M, 1.9M, or 2M.

[0117] 86. The composition of any of embodiments 80-85, wherein the pharmaceutical composition has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants.

[0118] 87. The composition of any of embodiments 80-86, wherein the composition comprises urea at a concentration of less than 0.1M, 0.2M, 0.3M, 0.4M, 0.5M, 0.6M, 0.7M, 0.8M, 0.9M, 1M, 1.1M, 1.2M, 1.3M, 1.5M, 1.6M, 1.7M, 1.8M, 1.9M, or 2M.

[0119] 88. The composition of embodiment 86 or 87, wherein the contaminant comprises one or more of the following: mycoplasma, endotoxin, host cell nucleic acids (e.g., host cell DNA and/or host cell RNA), animal-derived process impurities (e.g., serum albumin or trypsin), replication-competent agents (RCA), e.g., replication-competent virus or unwanted anellovectors (e.g., an anellovector other than the desired anellovector, e.g., a synthetic anellovector as described herein), free viral capsid protein, adventitious agents, and/or aggregates.

[0120] 89. The composition of any of embodiments 80-88, wherein the composition comprises less than 10% (e.g., less than about 10%, 5%, 4%, 3%, 2%, 1%, 0.5%, or 0.1%) contaminant by weight.

[0121] 90. The composition of any of embodiments 80-89, wherein at least 90% of proteinaceous exteriors comprise the same genetic element (e.g., the genetic element of the anellovector).

[0122] 91. The composition of any of embodiments 80-90, wherein at least 90% of proteinaceous exteriors comprise the same ORF1 molecule.

[0123] 92. A pharmaceutical composition comprising the anellovector or composition of any of the preceding embodiments, and a pharmaceutically acceptable carrier or excipient.

[0124] 93. A method of making an anellovector, the method comprising:

[0125] (a) providing a mixture comprising:

[0126] (i) a genetic element comprising RNA, and

[0127] (ii) an ORF1 molecule; and

[0128] (b) incubating the mixture under conditions suitable for enclosing the genetic element within a proteinaceous exterior comprising the ORF1 molecule, thereby making an anellovector; optionally wherein the mixture is not comprised in a cell.

[0129] 94. The method of embodiment 93, further comprising, prior to the providing of (a), expressing the ORF1 molecule, e.g., in a host cell (e.g., an insect cell or a mammalian cell).

[0130] 95. The method of embodiment 94, wherein the expressing comprising incubating a host cell (e.g., an insect cell or a mammalian cell) comprising a nucleic acid molecule (e.g., a baculovirus expression vector) encoding the ORF1 molecule under conditions suitable for producing the ORF1 molecule.

[0131] 96. The method of embodiment 94 or 95, further comprising, prior to the providing of (a), purifying the ORF1 molecule expressed by the host cell.

[0132] 97. A method of purifying an anellovector, the method comprising:

[0133] (a) providing an anellovector (e.g., as described herein) comprising:

[0134] (i) a genetic element, e.g., a genetic element comprising RNA, and

[0135] (ii) a proteinaceous exterior comprising an ORF1 molecule, the proteinaceous exterior enclosing the genetic element; and

[0136] (b) purifying the anellovector.

[0137] 98. The method of embodiment 96 or 97, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anellovector) comprises affinity purification, e.g., heparin affinity purification.

[0138] 99. The method of any of embodiments 96-98, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anellovector) comprises size exclusion chromatography (e.g., using a Tris buffer mobile phase).

[0139] 100. The method of any of embodiments 96-99, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anellovector) comprises affinity purification (e.g., heparin affinity purification) followed by size exclusion chromatography.

[0140] 101. The method of any of embodiments 96-100, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anellovector) comprises anion exchange chromatography (e.g., Mustang Q membrane chromatography).

[0141] 102. The method of any of embodiments 96-101, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anellovector) comprises mixed mode chromatography (e.g., using a mixed mode resin, e.g., a Cato700 resin).

[0142] 103. The method of any of embodiments 96-102, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anellovector) produces a composition comprising one or more virus-like particles (VLPs) comprising at least about 20, 30, 40, 50, or 60 copies, or 20-30, 30-40, 40-50, or 50-60 copies, of the ORF1 molecule.

[0143] 104. The method of embodiment 103, wherein at least 75%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the virus-like particles comprise proteinaceous exteriors that are 60mers or are particles of at least 30, 31, 32, 33, 34, or 35 nm in diameter.

[0144] 105. The method of embodiment 103, wherein the composition comprises at least  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$  particles/mL, or comprises  $10^5$ - $10^6$ ,  $10^6$ - $10^7$ ,  $10^7$ - $10^9$ ,  $10^8$ - $10^9$ ,  $10^9$ - $10^{10}$ , or  $10^{10}$ - $10^{11}$  particles/mL (e.g., as measured by electron microscopy).

[0145] 106. The method of any of embodiments 93-105, further comprising, prior to the providing of (a), incubating the ORF1 molecule under conditions suitable for disassembly of a proteinaceous exterior (e.g., a virus-like particle (VLP)) comprising the ORF1 molecule.

[0146] 107. The method of embodiment 106, wherein the conditions suitable for disassembly of the proteinaceous exterior comprising the ORF1 molecule comprise incubation in the presence of a denaturant.

[0147] 108. The method of any of embodiments 106-107, wherein the denaturant comprises a chaotropic agent (e.g., urea), or a detergent (e.g., SDS (e.g., 0.1% SDS), Tween, or Triton).

[0148] 109. The method of any of embodiments 106-108, wherein the conditions suitable for disassembly of the proteinaceous exterior comprising the ORF1 molecule comprise a predetermined conductivity, a high salt solution (e.g., a solution comprising NaCl, e.g., at a concentration of at least about 1M, e.g., at least about 0.5, 0.6, 0.7, 0.8, 0.9, 1.0,

1.1, 1.2, 1.3, 1.4, 1.5, 2, 3, 4, or 5M), heat (e.g., temperature above about 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, or 95° C.), or pH (e.g., acidic pH or basic pH).

[0149] 110. The method of any of embodiments 106-109, wherein the conditions suitable for disassembly of the proteinaceous exterior comprising the ORF1 molecule comprise incubation in a solution comprising urea at a concentration of at least 0.1M, 0.2M, 0.3M, 0.4M, 0.5M, 0.6M, 0.7M, 0.8M, 0.9M, 1M, 1.1M, 1.2M, 1.3M, 1.5M, 1.5M, 1.6M, 1.7M, 1.8M, 1.9M, or 2M.

[0150] 111. The method of any of embodiments 106-110, wherein the incubating of (b) results in at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 96%, or 100% of a population of particles comprising the ORF1 molecule or copies thereof being disassembled.

[0151] 112. The method of any of embodiments 106-111, wherein the conditions suitable for disassembly of the proteinaceous exterior are sufficient to disassemble a complex (e.g., a proteinaceous exterior) comprising at least about 20, 30, 40, 50, or 60 copies, or 20-30, 30-40, 40-50, or 50-60 copies, of the ORF1 molecule.

[0152] 113. The method of any of embodiments 106-112, wherein the conditions suitable for disassembly of the proteinaceous exterior result in fewer than  $10^8$  remaining intact particles/mL.

[0153] 114. The method of any of embodiments 106-113, further comprising, prior to the providing of (a), removing the ORF1 molecule from the conditions suitable for disassembly of the proteinaceous exterior (e.g., subjecting the ORF1 molecule to non-denaturing conditions).

[0154] 115. The method of embodiment 114, wherein the removing of the ORF1 molecule from the conditions suitable for disassembly of the proteinaceous exterior comprises reducing the concentration of the denaturant, e.g., reducing the concentration of the denaturant (e.g., urea) to below 0.1M, 0.2M, 0.3M, 0.4M, 0.5M, 0.6M, 0.7M, 0.8M, 0.9M, 1M, 1.1M, 1.2M, 1.3M, 1.5M, 1.5M, 1.6M, 1.7M, 1.8M, 1.9M, or 2M.

[0155] 116. The method of embodiment 114 or 115, wherein the removing results in the formation of one or more anellovectors each enclosing at least one copy (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, or 1000 copies) of the genetic element (e.g., the mRNA).

[0156] 117. The method of embodiment 116, wherein the number of anellovectors enclosing the at least one copy of the genetic element in the resulting solution is at least  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ , or  $10^{11}$ ; or is between  $10^5$ - $10^6$ ,  $10^6$ - $10^7$ ,  $10^7$ - $10^8$ ,  $10^8$ - $10^9$ ,  $10^9$ - $10^{10}$ , or  $10^{10}$ - $10^{11}$  (e.g., as measured by electron microscopy).

[0157] 118. The method of embodiment 116, wherein the number of anellovectors enclosing the at least one copy of the genetic element in the resulting solution is at least  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ , or  $10^{11}$  anellovectors/mL; or is between  $10^5$ - $10^6$ ,  $10^6$ - $10^7$ ,  $10^7$ - $10^8$ ,  $10^8$ - $10^9$ ,  $10^9$ - $10^{10}$ , or  $10^{10}$ - $10^{11}$  anellovectors/mL (e.g., as measured by electron microscopy).

[0158] 119. The method of any of embodiments 114-118, wherein the removing results in a solution comprising at least  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ , or  $10^{11}$  anellovectors/mL; or between  $10^5$ - $10^6$ ,  $10^6$ - $10^7$ ,  $10^7$ - $10^8$ ,  $10^8$ - $10^9$ ,  $10^9$ - $10^{10}$ , or  $10^{10}$ - $10^{11}$  anellovectors/mL (e.g., as measured by electron microscopy), wherein the anellovectors each enclose at least one copy (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25,

30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, or 1000 copies) of the genetic element (e.g., the mRNA) 120. The method of any of embodiments 114-119, wherein at least 75%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the anellovectors comprise proteinaceous exteriors that are 60mers or particles or are particles of at least 30, 31, 32, 33, 34, or 35 nm in diameter.

[0159] 121. The method of any of embodiments 93-120, wherein the genetic element of the anellovector is resistant to an endonuclease (e.g., an RNase).

[0160] 122. The method of any of embodiments 93-121, wherein (a) comprises admixing (i) and (ii).

[0161] 123. The method of any of embodiments 93-122, which is performed in a cell-free system.

[0162] 124. A method of manufacturing an anellovector composition, comprising:

[0163] (a) providing a plurality of anellovectors or compositions according to any of the preceding embodiments;

[0164] (b) optionally evaluating the plurality for one or more of: a contaminant described herein, an optical density measurement (e.g., OD 260), particle number (e.g., by HPLC), infectivity (e.g., particle:infectious unit ratio, e.g., as determined by fluorescence and/or ELISA); and

[0165] (c) formulating the plurality of anellovectors, e.g., as a pharmaceutical composition suitable for administration to a subject, e.g., if one or more of the parameters of (b) meet a specified threshold.

[0166] 125. A method of treating a disease or disorder in a subject in need thereof, the method comprising administering to the subject an anellovector or composition of any of the preceding embodiments, thereby treating a disease or disorder (e.g., as described herein) in the subject.

[0167] 126. A method of modulating, e.g., enhancing or inhibiting, a biological function (e.g., as described herein) in a subject, the method comprising administering the anellovector or the composition of any of the preceding embodiments to the subject.

[0168] 127. A method of delivering a genetic element to a cell, the method comprising contacting the anellovector or composition of any of the preceding embodiments with a cell, e.g., a eukaryotic cell, e.g., a mammalian cell, e.g., a human cell.

[0169] 128. Use of the anellovector or composition of any of the preceding embodiments for treating a disease or disorder (e.g., as described herein) in a subject.

[0170] 129. The anellovector or composition of any of the preceding embodiments for use in a method for treating a disease or disorder (e.g., as described herein) in a subject.

[0171] 130. The anellovector or composition of any of the preceding embodiments for use in the manufacture of a medicament for treating a disease or disorder (e.g., as described herein) in a subject.

[0172] Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

[0173] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. All publications, patent applications, patents, and other references mentioned herein are incorpo-

rated by reference in their entirety. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**[0174]** FIG. 1 is a series of electron micrographs showing that recombinant capsid proteins from anellovirus strains form virus-like particles (VLPs) in vitro. Capsid proteins produced in different cell lines form VLPs in vitro as observed by negative staining electron microscopy. (A) Ring 2 ORF1 VLP purified from insect cells with an observed diameter of approximately 35 nm. (B) Ring 10 ORF1 VLP purified from insect cells with an observed diameter of approximately 35 nm. (C) CAV VP1 VLP purified from mammalian cells with an observed diameter approximately 20 nm.

**[0175]** FIG. 2 is a diagram showing the eight beta-strand jelly roll domain (or jelly roll fold) observed in the structure of Beak and Feather Disease Virus (BFDV). By convention, the beta strands are labeled B through I. The strands form four antiparallel beta sheets with an orientation of B-I-D-G and C-I-E-F. The B-I-D-G sheet forms the interior of the viral capsid.

**[0176]** FIG. 3 is an amino acid sequence alignment depicting the jelly roll sequences for Anellovirus ORF1 protein as compared to jelly roll domain of Beak and Feather Disease Virus (BFDV)/Hepatitis E capsid protein (HePE). FIG. 3 discloses SEQ ID NOs: 956-968, respectively, in order of appearance.

**[0177]** FIGS. 4A-4E are a series of diagrams showing an exemplary method of producing Anellovirus ORF1 protein-based virus-like particles (VLPs) enclosing mRNA molecules encoding eGFP. (A) ORF1 protein was produced in cells and isolated as described herein. (B) VLPs that formed from the ORF1 proteins were then disassembled in a 2 M urea solution. (C) When the urea was removed in the absence of mRNA, few VLPs reformed (titer of less than  $10^8$  particles/mL detected by electron microscopy). (D, E) When the urea was removed in the presence of mRNAs encoding eGFP, substantial quantities of VLPs (titer of  $1 \times 10^9$ - $1 \times 10^{10}$  particles/mL) were detected by electron microscopy (EM). The following detailed description of the embodiments of the invention will be better understood when read in conjunction with the appended drawings. For the purpose of illustrating the invention, there are shown in the drawings embodiments that are presently exemplified. It should be understood, however, that the invention is not limited to the precise arrangement and instrumentalities of the embodiments shown in the drawings.

#### DETAILED DESCRIPTION OF THE INVENTION

##### Definitions

**[0178]** The present invention will be described with respect to particular embodiments and with reference to certain figures but the invention is not limited thereto but only by the claims. Terms as set forth hereinafter are generally to be understood in their common sense unless indicated otherwise.

**[0179]** Where the term "comprising" is used in the present description and claims, it does not exclude other elements. For the purposes of the present invention, the term "con-

sisting of" is considered to be a preferred embodiment of the term "comprising of". If hereinafter a group is defined to comprise at least a certain number of embodiments, this is to be understood to preferably also disclose a group which consists only of these embodiments.

**[0180]** Where an indefinite or definite article is used when referring to a singular noun, e.g. "a", "an" or "the", this includes a plural of that noun unless something else is specifically stated.

**[0181]** The wording "compound, composition, product, etc. for treating, modulating, etc." is to be understood to refer a compound, composition, product, etc. per se which is suitable for the indicated purposes of treating, modulating, etc. The wording "compound, composition, product, etc. for treating, modulating, etc." additionally discloses that, as an embodiment, such compound, composition, product, etc. is for use in treating, modulating, etc.

**[0182]** The wording "compound, composition, product, etc. for use in . . .", "use of a compound, composition, product, etc. in the manufacture of a medicament, pharmaceutical composition, veterinary composition, diagnostic composition, etc. for . . .", or "compound, composition, product, etc. for use as a medicament . . ." indicates that such compounds, compositions, products, etc. are to be used in therapeutic methods which may be practiced on the human or animal body. They are considered as an equivalent disclosure of embodiments and claims pertaining to methods of treatment, etc. If an embodiment or a claim thus refers to "a compound for use in treating a human or animal being suspected to suffer from a disease", this is considered to be also a disclosure of a "use of a compound in the manufacture of a medicament for treating a human or animal being suspected to suffer from a disease" or a "method of treatment by administering a compound to a human or animal being suspected to suffer from a disease". The wording "compound, composition, product, etc. for treating, modulating, etc." is to be understood to refer a compound, composition, product, etc. per se which is suitable for the indicated purposes of treating, modulating, etc.

**[0183]** If hereinafter examples of a term, value, number, etc. are provided in parentheses, this is to be understood as an indication that the examples mentioned in the parentheses can constitute an embodiment. For example, if it is stated that "in embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1-encoding nucleotide sequence of Table 1 (e.g., nucleotides 571-2613 of the nucleic acid sequence of Table 1)", then some embodiments relate to nucleic acid molecules comprising a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to nucleotides 571-2613 of the nucleic acid sequence of Table 1.

**[0184]** The term "amplification," as used herein, refers to replication of a nucleic acid molecule or a portion thereof, to produce one or more additional copies of the nucleic acid molecule or a portion thereof (e.g., a genetic element or a genetic element region). In some embodiments, amplification results in partial replication of a nucleic acid sequence. In some embodiments, amplification occurs via rolling circle replication.

**[0185]** As used herein, the term "anellovector" refers to a vehicle comprising a genetic element, e.g., an RNA, e.g., a

circular RNA, enclosed in a proteinaceous exterior. In some embodiments, the genetic element is substantially protected from digestion with RNase by a proteinaceous exterior. A “synthetic anellovector,” as used herein, generally refers to an anellovector that is not naturally occurring, e.g., has a sequence that is different relative to a wild-type virus (e.g., a wild-type Anellovirus as described herein). In some embodiments, the synthetic anellovector is engineered or recombinant, e.g., comprises a genetic element that comprises a difference or modification relative to a wild-type viral genome (e.g., a wild-type Anellovirus genome as described herein). In some embodiments, enclosed within a proteinaceous exterior encompasses 100% coverage by a proteinaceous exterior, as well as less than 100% coverage, e.g., 95%, 90%, 85%, 80%, 70%, 60%, 50% or less. For example, gaps or discontinuities (e.g., that render the proteinaceous exterior permeable to water, ions, peptides, or small molecules) may be present in the proteinaceous exterior, so long as the genetic element is retained in the proteinaceous exterior or protected from digestion with an RNase, e.g., prior to entry into a host cell. In some embodiments, the anellovector is purified, e.g., it is separated from its original source and/or substantially free (>50%, >60%, >70%, >80%, >90%) of other components. In some embodiments, the anellovector is capable of introducing the genetic element into a target cell (e.g., via infection). In some embodiments, the anellovector is an infective synthetic viral particle containing certain Anellovirus elements, such as an Anellovirus ORF1 molecule.

[0186] As used herein, the term “antibody molecule” refers to a protein, e.g., an immunoglobulin chain or fragment thereof, comprising at least one immunoglobulin variable domain sequence. The term “antibody molecule” encompasses full-length antibodies and antibody fragments (e.g., scFvs). In some embodiments, an antibody molecule is a multispecific antibody molecule, e.g., the antibody molecule comprises a plurality of immunoglobulin variable domain sequences, wherein a first immunoglobulin variable domain sequence of the plurality has binding specificity for a first epitope and a second immunoglobulin variable domain sequence of the plurality has binding specificity for a second epitope. In some embodiments, the multispecific antibody molecule is a bispecific antibody molecule. A bispecific antibody molecule is generally characterized by a first immunoglobulin variable domain sequence which has binding specificity for a first epitope and a second immunoglobulin variable domain sequence that has binding specificity for a second epitope.

[0187] The term “backbone” or “backbone region,” as used herein, refers to a region within a nucleic acid molecule (e.g., within a bacmid or donor vector, e.g., as described herein) that comprises one or more elements involved in (e.g., necessary and/or sufficient for) replication and/or maintenance of the nucleic acid molecule in a host cell. In some embodiments, a backbone region, such as a “baculovirus backbone region,” comprises one or more baculoviral elements (e.g., a baculovirus genome or a functional fragment thereof), e.g., suitable for replication of the nucleic acid construct in insect cells (e.g., Sf9 cells). In some embodiments, the backbone further comprises a selectable marker. In some embodiments, a nucleic acid molecule comprises a genetic element region and a backbone region (e.g., a baculovirus backbone region and/or a backbone region suitable for replication in bacterial cells).

[0188] The term “bacmid”, as used herein, refers to a nucleic acid molecule comprising sufficient baculovirus backbone elements such that it is suitable for replication in insect cells, and furthermore is suitable for replication in bacterial cells. In some embodiments, the nucleic acid molecule is suitable for replication in bacterial cells (e.g., *E. coli* cells, e.g., DH 10Bac cells).

[0189] As used herein, a “circular” nucleic acid refers to a nucleic acid that forms a structure without free 5' or 3' ends. In some embodiments, the circular nucleic acid is closed through covalent or non-covalent bonds. For instance, the circular nucleic acid may be made by covalently linking the ends of a linear nucleic acid, e.g., with a phosphate-sugar bond or a synthetic linker moiety. In other embodiments, the circular nucleic acid comprises two ends that are in proximity and are not free (not substantially accessible to an exonuclease). For instance, the circular nucleic acid may be made by hybridizing the ends of a linear nucleic acid directly or through a nucleic acid splint.

[0190] As used herein, a “DNA region” refers to a portion of a polynucleotide strand comprising a plurality of DNA nucleotides. For example, in some embodiments a DNA region is a plurality of DNA nucleotides incorporated into an RNA strand. For example, a DNA region comprises about 5-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, or 90-100 DNA nucleotides within a polynucleotide strand.

[0191] As used herein, a nucleic acid “encoding” refers to a nucleic acid sequence encoding an amino acid sequence or a polynucleotide, e.g., an mRNA or functional polynucleotide (e.g., a non-coding RNA, e.g., an siRNA or miRNA).

[0192] An “exogenous” agent (e.g., an effector, a nucleic acid (e.g., RNA), a gene, payload, protein) as used herein refers to an agent that is either not comprised by, or not encoded by, a corresponding wild-type virus, e.g., an Anellovirus as described herein. In some embodiments, the exogenous agent does not naturally exist, such as a protein or nucleic acid that has a sequence that is altered (e.g., by insertion, deletion, or substitution) relative to a naturally occurring protein or nucleic acid. In some embodiments, the exogenous agent does not naturally exist in the host cell. In some embodiments, the exogenous agent exists naturally in the host cell but is exogenous to the virus. In some embodiments, the exogenous agent exists naturally in the host cell, but is not present at a desired level or at a desired time.

[0193] A “heterologous” agent or element (e.g., an effector, a nucleic acid sequence, an amino acid sequence), as used herein with respect to another agent or element (e.g., an effector, a nucleic acid sequence, an amino acid sequence), refers to agents or elements that are not naturally found together, e.g., in a wild-type virus, e.g., an Anellovirus. In some embodiments, a heterologous nucleic acid sequence may be present in the same nucleic acid as a naturally occurring nucleic acid sequence (e.g., a sequence that is naturally occurring in the Anellovirus). In some embodiments, a heterologous agent or element is exogenous relative to an Anellovirus from which other (e.g., the remainder of) elements of the anellovector are based.

[0194] As used herein, the term “genetic element” refers to a nucleic acid molecule that is or can be enclosed within (e.g., protected from RNase digestion by) a proteinaceous exterior, e.g., to form an anellovector as described herein. It is understood that the genetic element can be produced as naked RNA and optionally further assembled into a pro-

teinaceous exterior. It is also understood that an anellovector can insert its genetic element into a cell, resulting in the genetic element being present in the cell and the proteinaceous exterior not necessarily entering the cell.

[0195] As used herein, “genetic element construct” refers to a nucleic acid construct (e.g., a plasmid, bacmid, donor vector, cosmid, or minicircle) comprising a genetic element sequence, or fragment thereof. In some embodiments, a bacmid or donor vector as described herein is a genetic element construct comprising a genetic element sequence, or fragment thereof.

[0196] The term “genetic element region,” as used herein, refers to a region of a construct that comprises the sequence of a genetic element. In some embodiments, the genetic element region comprises a sequence having sufficient identity to a wild-type Anellovirus sequence, or a fragment thereof, to be enclosed by a proteinaceous exterior, thereby forming an anellovector (e.g., a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the wild-type Anellovirus sequence or fragment thereof). In embodiments, the genetic element region comprises a protein binding sequence, e.g., as described herein (e.g., a 5' UTR, 3' UTR, and/or a GC-rich region as described herein, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto). In some embodiments, the construct comprising a genetic element region is not enclosed in a proteinaceous exterior, but a genetic element produced from the construct can be enclosed in a proteinaceous exterior. In some embodiments, the construct comprising the genetic element region further comprises a vector backbone (e.g., a bacmid backbone or a donor vector backbone). In some embodiments, the construct (e.g., bacmid) comprises one or more baculovirus elements (e.g., a baculovirus genome, e.g., comprising the genetic element region).

[0197] As used herein, the term “mutant” when used with respect to a genome (e.g., an Anellovirus genome), or a fragment thereof, refers to a sequence having at least one change relative to a corresponding wild-type Anellovirus sequence. In some embodiments, the mutant genome or fragment thereof comprises at least one single nucleotide polymorphism, addition, deletion, or frameshift relative to the corresponding wild-type Anellovirus sequence. In some embodiments, the mutant genome or fragment thereof comprises a deletion of at least one Anellovirus ORF (e.g., one or more of ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and/or ORF1/2) relative to the corresponding wild-type Anellovirus sequence. In some embodiments, the mutant genome or fragment thereof comprises a deletion of all Anellovirus ORFs (e.g., all of ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and ORF1/2) relative to the corresponding wild-type Anellovirus sequence. In some embodiments, the mutant genome or fragment thereof comprises a deletion of at least one Anellovirus noncoding region (e.g., one or more of a 5' UTR, 3' UTR, and/or GC-rich region) relative to the corresponding wild-type Anellovirus sequence. In some embodiments, the mutant genome or fragment thereof comprises or encodes an exogenous effector.

[0198] As used herein the term “ORF molecule” refers to a polypeptide having an activity and/or a structural feature of an Anellovirus ORF protein (e.g., a polypeptide comprising an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and/or ORF1/2 protein), or a functional fragment thereof.

When used generically (i.e., “ORF molecule”), the polypeptide may comprise an activity and/or structural feature of any of the Anellovirus ORFs described herein (e.g., any of an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and/or ORF1/2), or a functional fragment thereof. When used with a modifier to indicate a particular open reading frame (e.g., “ORF1 molecule,” “ORF2 molecule,” “ORF2/2 molecule,” “ORF2/3 molecule,” “ORF1/1 molecule,” or “ORF1/2 molecule”), it is generally meant that the polypeptide comprises an activity and/or structural feature of the corresponding Anellovirus ORF protein, or a functional fragment thereof (for example, as defined below for “ORF1 molecule”). For example, an “ORF2 molecule” comprises an activity and/or structural feature of an Anellovirus ORF2 protein, or a functional fragment thereof.

[0199] As used herein, the term “ORF1 molecule” refers to a polypeptide having an activity and/or a structural feature of an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein), or a functional fragment thereof. An ORF1 molecule may, in some instances, comprise one or more of (e.g., 1, 2, 3 or 4 of): a first region comprising at least 60% basic residues (e.g., at least 60% arginine residues), a second region comprising at least about six beta strands (e.g., at least 4, 5, 6, 7, 8, 9, 10, 11, or 12 beta strands), a third region comprising a structure or an activity of an Anellovirus N22 domain (e.g., as described herein, e.g., an N22 domain from an Anellovirus ORF1 protein as described herein), and/or a fourth region comprising a structure or an activity of an Anellovirus C-terminal domain (CTD) (e.g., as described herein, e.g., a CTD from an Anellovirus ORF1 protein as described herein). In some instances, the ORF1 molecule comprises, in N-terminal to C-terminal order, the first, second, third, and fourth regions. In some instances, an anellovector comprises an ORF1 molecule comprising, in N-terminal to C-terminal order, the first, second, third, and fourth regions. An ORF1 molecule may, in some instances, comprise a polypeptide encoded by an Anellovirus ORF1 nucleic acid. An ORF1 molecule may, in some instances, further comprise a heterologous sequence, e.g., a hypervariable region (HVR), e.g., an HVR from an Anellovirus ORF1 protein, e.g., as described herein. An “Anellovirus ORF1 protein,” as used herein, refers to an ORF1 protein encoded by an Anellovirus genome (e.g., a wild-type Anellovirus genome, e.g., as described herein).

[0200] As used herein, the term “ORF2 molecule” refers to a polypeptide having an activity and/or a structural feature of an Anellovirus ORF2 protein (e.g., an Anellovirus ORF2 protein as described herein), or a functional fragment thereof. An “Anellovirus ORF2 protein,” as used herein, refers to an ORF2 protein encoded by an Anellovirus genome (e.g., a wild-type Anellovirus genome, e.g., as described herein).

[0201] As used herein, the term “proteinaceous exterior” refers to an exterior component that is predominantly (e.g., >50%, >60%, >70%, >80%, >90%) protein.

[0202] As used herein, the term “regulatory nucleic acid” refers to a nucleic acid sequence that modifies expression, e.g., transcription and/or translation, of a DNA sequence that encodes an expression product. In some embodiments, the expression product comprises RNA or protein.

[0203] As used herein, the term “regulatory sequence” refers to a nucleic acid sequence that modifies transcription

of a target gene product. In some embodiments, the regulatory sequence is a promoter or an enhancer.

**[0204]** As used herein, a “substantially non-pathogenic” organism, particle, or component, refers to an organism, particle (e.g., a virus or an anellovector, e.g., as described herein), or component thereof that does not cause or induce an unacceptable disease or pathogenic condition, e.g., in a host organism, e.g., a mammal, e.g., a human. In some embodiments, administration of an anellovector to a subject can result in minor reactions or side effects that are acceptable as part of standard of care.

**[0205]** As used herein, the term “non-pathogenic” refers to an organism or component thereof that does not cause or induce an undesirable condition (e.g., a disease or pathogenic condition), e.g., in a host organism, e.g., a mammal, e.g., a human.

**[0206]** As used herein, a “substantially non-integrating” genetic element refers to a genetic element, e.g., a genetic element in a virus or anellovector, e.g., as described herein, wherein less than about 0.01%, 0.05%, 0.1%, 0.5%, or 1% of the genetic element that enter into a host cell (e.g., a eukaryotic cell) or organism (e.g., a mammal, e.g., a human) integrate into the genome. In some embodiments the genetic element does not detectably integrate into the genome of, e.g., a host cell. In some embodiments, integration of the genetic element into the genome can be detected using techniques as described herein, e.g., nucleic acid sequencing, PCR detection and/or nucleic acid hybridization. In some embodiments, integration frequency is determined by quantitative gel purification assay of genomic DNA separated from free vector, e.g., as described in Wang et al. (2004, *Gene Therapy* 11: 711-721, incorporated herein by reference in its entirety).

**[0207]** As used herein, a “substantially non-immunogenic” organism, particle, or component, refers to an organism, particle (e.g., a virus or anellovector, e.g., as described herein), or component thereof, that does not cause or induce an undesired or untargeted immune response, e.g., in a host tissue or organism (e.g., a mammal, e.g., a human). In some embodiments, the substantially non-immunogenic organism, particle, or component does not produce a clinically significant immune response. In some embodiments, the substantially non-immunogenic anellovector does not produce a clinically significant immune response against a protein comprising an amino acid sequence or encoded by a nucleic acid sequence of an Anellovirus or anellovector genetic element. In some embodiments, an immune response (e.g., an undesired or untargeted immune response) is detected by assaying antibody (e.g., neutralizing antibody) presence or level (e.g., presence or level of an anti-anellovector antibody, e.g., presence or level of an antibody against an anellovector as described herein) in a subject, e.g., according to the anti-TTV antibody detection method described in Tsuda et al. (1999; *J. Virol. Methods* 77: 199-206; incorporated herein by reference) and/or the method for determining anti-TTV IgG levels described in Kakkola et al. (2008; *Virology* 382: 182-189; incorporated herein by reference). Antibodies (e.g., neutralizing antibodies) against an Anellovirus or an anellovector based thereon can also be detected by methods in the art for detecting anti-viral antibodies, e.g., methods of detecting anti-AAV antibodies, e.g., as described in Calcedo et al. (2013; *Front. Immunol.* 4(341): 1-7; incorporated herein by reference).

**[0208]** A “subsequence” as used herein refers to a nucleic acid sequence or an amino acid sequence that is comprised in a larger nucleic acid sequence or amino acid sequence, respectively. In some instances, a subsequence may comprise a domain or functional fragment of the larger sequence. In some instances, the subsequence may comprise a fragment of the larger sequence capable of forming secondary and/or tertiary structures when isolated from the larger sequence similar to the secondary and/or tertiary structures formed by the subsequence when present with the remainder of the larger sequence. In some instances, a subsequence can be replaced by another sequence (e.g., a subsequence comprising an exogenous sequence or a sequence heterologous to the remainder of the larger sequence, e.g., a corresponding subsequence from a different Anellovirus).

**[0209]** This invention relates generally to anellovectors, e.g., synthetic anellovectors, and uses thereof. The present disclosure provides anellovectors, compositions comprising anellovectors, and methods of making or using anellovectors. Anellovectors are generally useful as delivery vehicles, e.g., for delivering a therapeutic agent to a eukaryotic cell. Generally, an anellovector described herein will include a genetic element comprising an RNA sequence (e.g., an RNA sequence encoding an effector, e.g., an exogenous effector or an endogenous effector) enclosed within a proteinaceous exterior. An anellovector may include one or more deletions of sequences (e.g., regions or domains as described herein) relative to an Anellovirus sequence (e.g., as described herein). Anellovectors can be used as a substantially non-immunogenic vehicle for delivering the genetic element, or an effector encoded therein (e.g., a polypeptide or nucleic acid effector, e.g., as described herein), into eukaryotic cells, e.g., to treat a disease or disorder in a subject comprising the cells.

#### TABLE OF CONTENTS

- [0210]** I. Compositions and Methods for Making Anellovectors by In Vitro Assembly
- [0211]** A. Components and Assembly of Anellovectors
  - [0212]** i. ORF1 molecules for assembly of Anellovectors
  - [0213]** ii. ORF2 molecules for assembly of Anellovectors
- [0214]** B. Genetic Elements
  - [0215]** i. Genetic Elements comprising RNA
    - [0216]** a. RNA-only genetic elements
    - [0217]** b. Hybridized RNA-ssDNA genetic elements
    - [0218]** c. RNA/DNA conjugates
  - [0219]** C. Genetic Element Constructs
  - [0220]** D. Production of RNA-based genetic elements
    - [0221]** E. Production of protein components
      - [0222]** i. Baculovirus expression systems
      - [0223]** ii. Insect cell systems
      - [0224]** iii. Mammalian cell systems
    - [0225]** F. Effectors
    - [0226]** G. In vitro assembly methods
    - [0227]** H. Enrichment and Purification
  - [0228]** II. Anellovectors
    - [0229]** A. Anelloviruses
    - [0230]** B. ORF1 molecules
    - [0231]** C. ORF2 molecules

- [0232] D. Genetic elements
  - [0233] E. Protein binding sequences
  - [0234] F. 5' UTR regions
  - [0235] G. GC-rich regions
  - [0236] H. Effectors
  - [0237] I. Regulatory Sequences
  - [0238] J. Other Sequences
  - [0239] K. Proteinaceous exterior
- [0240] III. Methods of use
- [0241] IV. Administration/Delivery

### I. Compositions and Methods for Making Anellovectors by In Vitro Assembly

[0242] The present disclosure provides, in some aspects, compositions and methods that can be used for producing anellovectors, e.g., anellovectors having a genetic element comprising RNA, as described herein. In some embodiments, the compositions and methods described herein can be used to produce a genetic element or a genetic element construct. In some embodiments, the compositions and methods described herein can be used to produce a genetic element or a genetic element construct by in vitro assembly. In some embodiments, the compositions and methods described herein can be used to produce one or more Anellovirus ORF molecules (e.g., an ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 molecule, or a functional fragment or splice variant thereof). In some embodiments, the compositions and methods described herein can be used to produce a proteinaceous exterior or a component thereof (e.g., an ORF1 molecule), e.g., in a host cell (e.g., an insect cell, e.g., an Sf9 cell).

#### Components and Assembly of Anellovectors

[0243] The compositions and methods herein can be used to produce anellovectors. As described herein, an anellovector generally comprises a genetic element (e.g., an RNA molecule) enclosed within a proteinaceous exterior (e.g., comprising a polypeptide encoded by an Anellovirus ORF1 nucleic acid, e.g., as described herein). In some embodiments, the genetic element comprises one or more sequences encoding Anellovirus ORFs (e.g., one or more of an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2). As used herein, an Anellovirus ORF or ORF molecule (e.g., an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2) includes a polypeptide comprising an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a corresponding Anellovirus ORF sequence, e.g., as described in PCT/US2018/037379 or PCT/US19/65995 (each of which is incorporated by reference herein in their entirety). In embodiments, the genetic element comprises a sequence encoding an Anellovirus ORF1, or a splice variant or functional fragment thereof (e.g., a jelly-roll region, e.g., as described herein). In some embodiments, the proteinaceous exterior comprises a polypeptide encoded by an Anellovirus ORF1 nucleic acid (e.g., an Anellovirus ORF1 molecule or a splice variant or functional fragment thereof).

[0244] In some embodiments, an anellovector is assembled by enclosing a genetic element (e.g., as described herein) within a proteinaceous exterior (e.g., as described herein). In some embodiments, the genetic element is enclosed within the proteinaceous exterior in a host cell (e.g., an insect cell, e.g., an Sf9 cell). In some embodiments,

the host cell expresses one or more polypeptides comprised in the proteinaceous exterior (e.g., a polypeptide encoded by an Anellovirus ORF1 nucleic acid, e.g., an ORF1 molecule). For example, in some embodiments, the host cell comprises a nucleic acid sequence encoding an Anellovirus ORF1 molecule, e.g., a splice variant or a functional fragment of an Anellovirus ORF1 polypeptide (e.g., a wild-type Anellovirus ORF1 protein or a polypeptide encoded by a wild-type Anellovirus ORF1 nucleic acid, e.g., as described herein). In embodiments, the nucleic acid sequence encoding the Anellovirus ORF1 molecule is comprised in a nucleic acid construct (e.g., a plasmid, viral vector, virus, minicircle, bacmid, or artificial chromosome) comprised in the host cell. In embodiments, the nucleic acid sequence encoding the Anellovirus ORF1 molecule is integrated into the genome of the host cell.

[0245] In some embodiments, the host cell comprises the genetic element and/or a nucleic acid construct comprising the sequence of the genetic element. In some embodiments, the nucleic acid construct is selected from a plasmid, viral nucleic acid, minicircle, bacmid, or artificial chromosome. In some embodiments, the genetic element is excised from the nucleic acid construct (e.g., bacmid) and, optionally, converted from a double-stranded form to a single-stranded form (e.g., by denaturation). In some embodiments, the genetic element is generated by a polymerase based on a template sequence in the nucleic acid construct (e.g., bacmid). In some embodiments, the polymerase produces a single-stranded copy of the genetic element sequence, which can optionally be circularized to form a genetic element as described herein.

[0246] In some embodiments, the host cell comprises a genetic element construct (e.g., a bacmid, plasmid, or minicircle) and a bacmid comprising one or more sequences encoding Anellovirus ORF molecules (e.g., ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and/or ORF1/2 ORF molecules), or functional fragments thereof. In some embodiments, proteinaceous exterior proteins are expressed from the bacmid. In embodiments, the proteinaceous exterior proteins expressed from the bacmid enclose a genetic element, thereby forming an anellovector. In some embodiments, the bacmid comprises a backbone suitable for replication of the nucleic acid construct in insect cells (e.g., Sf9 cells), e.g., a baculovirus backbone region. In some embodiments, the bacmid comprises a backbone region suitable for replication of the genetic element construct in bacterial cells (e.g., *E. coli* cells, e.g., DH 10Bac cells). In some embodiments, the genetic element construct comprises a backbone suitable for replication of the nucleic acid construct in insect cells (e.g., Sf9 cells), e.g., a baculovirus backbone region. In some embodiments, the genetic element construct comprises a backbone region suitable for replication of the genetic element construct in bacterial cells (e.g., *E. coli* cells, e.g., DH 10Bac cells). In some embodiments, the bacmid is introduced into the host cell via a baculovirus particle. In embodiments, the bacmid is produced by a producer cell, e.g., an insect cell (e.g., an Sf9 cell) or a bacterial cell (e.g., an *E. coli* cell, e.g., a DH 10Bac cell). In embodiments, the producer cell comprises a bacmid and/or a donor vector, e.g., as described herein. In embodiments, the producer cell further comprises sufficient cellular machinery for replication of the bacmid and/or donor vector.

## ORF1 Molecules, e.g., for Assembly of Anellovectors

**[0247]** An anellovector can be made, for example, by enclosing a genetic element within a proteinaceous exterior. In some embodiments, the enclosure occurs in a cell-free system or in a cell. The proteinaceous exterior of an Anellovector generally comprises a polypeptide encoded by an Anellovirus ORF1 nucleic acid (e.g., an Anellovirus ORF1 molecule or a splice variant or functional fragment thereof, e.g., as described herein). An ORF1 molecule may, in some embodiments, comprise one or more of: a first region comprising an arginine rich region, e.g., a region having at least 60% basic residues (e.g., at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% basic residues; e.g., between 60%-90%, 60%-80%, 70%-90%, or 70-80% basic residues), and a second region comprising jelly-roll domain, e.g., at least six beta strands (e.g., 4, 5, 6, 7, 8, 9, 10, 11, or 12 beta strands). In embodiments, the proteinaceous exterior comprises one or more (e.g., 1, 2, 3, 4, or all 5) of an Anellovirus ORF1 arginine-rich region, jelly-roll region, N22 domain, hypervariable region, and/or C-terminal domain. In some embodiments, the proteinaceous exterior comprises an Anellovirus ORF1 jelly-roll region (e.g., as described herein). In some embodiments, the proteinaceous exterior comprises an Anellovirus ORF1 arginine-rich region (e.g., as described herein). In some embodiments, the proteinaceous exterior comprises an Anellovirus ORF1 N22 domain (e.g., as described herein). In some embodiments, the proteinaceous exterior comprises an Anellovirus hypervariable region (e.g., as described herein). In some embodiments, the proteinaceous exterior comprises an Anellovirus ORF1 C-terminal domain (e.g., as described herein).

**[0248]** In some embodiments, the anellovector comprises an ORF1 molecule and/or a nucleic acid encoding an ORF1 molecule. Generally, an ORF1 molecule comprises a polypeptide having the structural features and/or activity of an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein), or a functional fragment thereof. In some embodiments, the ORF1 molecule comprises a truncation relative to an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein). In some embodiments, the ORF1 molecule is truncated by at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, or 700 amino acids of the Anellovirus ORF1 protein. In some embodiments, an ORF1 molecule comprises an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to an Alphatorquevirus, Betatorquevirus, or Gammatorquevirus ORF1 protein, e.g., as described herein. An ORF1 molecule can generally bind to a nucleic acid molecule, such as DNA (e.g., a genetic element, e.g., as described herein). In some embodiments, an ORF1 molecule localizes to the nucleus of a cell. In certain embodiments, an ORF1 molecule localizes to the nucleolus of a cell.

**[0249]** Without wishing to be bound by theory, an ORF1 molecule may be capable of binding to other ORF1 molecules, e.g., to form a proteinaceous exterior (e.g., as described herein). Such an ORF1 molecule may be described as having the capacity to form a capsid. In some embodiments, the proteinaceous exterior may enclose a nucleic acid molecule (e.g., a genetic element as described herein). In some embodiments, a plurality of ORF1 molecules may form a multimer, e.g., to produce a proteinaceous

exterior. In some embodiments, the multimer may be a homomultimer. In other embodiments, the multimer may be a heteromultimer.

## ORF2 Molecules, e.g., for Assembly of Anellovectors

**[0250]** Producing an anellovector using the compositions or methods described herein may involve expression of an Anellovirus ORF2 molecule (e.g., as described herein), or a splice variant or functional fragment thereof. In some embodiments, the anellovector comprises an ORF2 molecule, or a splice variant or functional fragment thereof, and/or a nucleic acid encoding an ORF2 molecule, or a splice variant or functional fragment thereof. In some embodiments, the anellovector does not comprise an ORF2 molecule, or a splice variant or functional fragment thereof, and/or a nucleic acid encoding an ORF2 molecule, or a splice variant or functional fragment thereof. In some embodiments, producing the anellovector comprises expression of an ORF2 molecule, or a splice variant or functional fragment thereof, but the ORF2 molecule is not incorporated into the anellovector.

## Genetic Elements

## Genetic Elements Comprising RNA

**[0251]** In some embodiments, a genetic element is or comprises a nucleic acid. In some embodiments, a genetic element is a single-stranded polynucleotide. In some embodiments, a genetic element comprises one or more double stranded regions. In some embodiments, a genetic element comprises RNA. In some embodiments, the genetic element comprises an RNA hairpin structure. In some embodiments, the genetic element is an mRNA, e.g., a chemically modified mRNA. In some embodiments, a genetic element consists of at least 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% RNA. In some embodiments, the genetic element comprises a DNA strand and an RNA strand, e.g., wherein at least a portion of the DNA strand hybridizes to at least a portion of the RNA strand.

**[0252]** In some embodiments, the genetic element does not encode any of an Anellovirus ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3.

**[0253]** In some embodiments, the RNA genetic element encodes an effector, e.g., an effector protein.

**[0254]** In some embodiments, the RNA genetic element is or comprises an effector, e.g., a functional RNA. In some embodiments, RNA is selected from the group consisting of mRNA, rRNA, tRNA (e.g., a TREM), regulatory RNA, non-coding RNA, long non-coding RNA (lncRNA), circular RNA (circRNA), double stranded RNA (dsRNA), guide RNA (gRNA), small interfering RNA (siRNA), short hairpin RNA (shRNA), piwi-interacting RNA (piRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), extracellular RNA (exRNA), small Cajal body-specific RNA (scaRNA), microRNA (miRNA), and other RNAi molecules.

**[0255]** In some embodiments, the genetic element comprises RNA, e.g., chemically modified RNA. In some embodiments, one or more nucleotides of RNA of a genetic element are chemically modified. In some embodiments, RNA comprises one or more chemical modifications to one or more bases. In some embodiments, RNA comprises one

or more chemical modifications to one or more sugars. In some embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of RNA of a genetic element are chemically modified. In some embodiments, RNA comprises one or more backbone modifications. In some embodiments, a modification comprises a non-naturally occurring modification, e.g., a modification described in any one of Tables 5-9. A non-naturally occurring modification can be made according to methods known in the art.

[0256] In some embodiments, a genetic element described herein comprises a non-naturally occurring modification provided in Table 5, or a combination thereof.

TABLE 5

Exemplary non-naturally occurring modifications			
Name	Symbol	Base	Naturally Occurring
7-deaza-adenosine	—	A	NO
N1-methyl-adenosine	—	A	NO
N6,N6 (dimethyl)adenine	—	A	NO
N6-cis-hydroxy-isopentenyl-adenosine	—	A	NO
a-thio-adenosine	—	A	NO
2 (amino)adenine	—	A	NO
2 (aminopropyl)adenine	—	A	NO
2 (methylthio) N6 (isopentenyl)adenine	—	A	NO
2-(alkyl)adenine	—	A	NO
2-(aminoalkyl)adenine	—	A	NO
2-(aminopropyl)adenine	—	A	NO
2-(halo)adenine	—	A	NO
2-(halo)adenine	—	A	NO
2-(propyl)adenine	—	A	NO
2'-Amino-2'-deoxy-ATP	—	A	NO
2'-Azido-2'-deoxy-ATP	—	A	NO
2'-Deoxy-2'-a-aminoadenosine TP	—	A	NO
2'-Deoxy-2'-a-azidoadenosine TP	—	A	NO
6 (alkyl)adenine	—	A	NO
6 (methyl)adenine	—	A	NO
6-(alkyl)adenine	—	A	NO
6-(methyl)adenine	—	A	NO
7 (deaza)adenine	—	A	NO
8 (alkenyl)adenine	—	A	NO
8 (alkynyl)adenine	—	A	NO
8 (amino)adenine	—	A	NO
8 (thioalkyl)adenine	—	A	NO
8-(alkenyl)adenine	—	A	NO
8-(alkyl)adenine	—	A	NO
8-(alkynyl)adenine	—	A	NO
8-(amino)adenine	—	A	NO
8-(halo)adenine	—	A	NO
8-(hydroxyl)adenine	—	A	NO
8-(thioalkyl)adenine	—	A	NO
8-(thiol)adenine	—	A	NO
8-azido-adenosine	—	A	NO
aza adenine	—	A	NO
deaza adenine	—	A	NO
N6 (methyl)adenine	—	A	NO
N6-(isopentyl)adenine	—	A	NO
7-deaza-8-aza-adenosine	—	A	NO
7-methyladenine	—	A	NO
1-Deazaadenosine TP	—	A	NO
2Fluoro-N6-Bz-deoxyadenosine TP	—	A	NO
2'-OME-2-Amino-ATP	—	A	NO
2'O-methyl-N6-Bz-deoxyadenosine TP	—	A	NO
2'-a-Ethylnyladenosine TP	—	A	NO
2-aminoadenine	—	A	NO
2-Aminoadenosine TP	—	A	NO
2-Amino-ATP	—	A	NO
2'-a-Trifluoromethyladenosine TP	—	A	NO
2-Azidoadenosine TP	—	A	NO
2'-b-Ethylnyladenosine TP	—	A	NO
2-Bromoadenosine TP	—	A	NO
2'-b-Trifluoromethyladenosine TP	—	A	NO
2-Chloroadenosine TP	—	A	NO

TABLE 5-continued

Exemplary non-naturally occurring modifications			
Name	Symbol	Base	Naturally Occurring
2'-Deoxy-2',2'-difluoroadenosine TP	—	A	NO
2'-Deoxy-2'-a-mercaptopadenosine TP	—	A	NO
2'-Deoxy-2'-a-thiomethoxyadenosine TP	—	A	NO
2'-Deoxy-2'-b-aminoadenosine TP	—	A	NO
2'-Deoxy-2'-b-azidoadenosine TP	—	A	NO
2'-Deoxy-2'-b-bromoadenosine TP	—	A	NO
2'-Deoxy-2'-b-chloroadenosine TP	—	A	NO
2'-Deoxy-2'-b-fluoroadenosine TP	—	A	NO
2'-Deoxy-2'-b-iodoadenosine TP	—	A	NO
2'-Deoxy-2'-b-mercaptopadenosine TP	—	A	NO
2'-Deoxy-2'-b-thiomethoxyadenosine TP	—	A	NO
2-Fluoroadenosine TP	—	A	NO
2-Iodoadenosine TP	—	A	NO
2-Mercaptopadenosine TP	—	A	NO
2-methoxy-adenine	—	A	NO
2-methylthio-adenine	—	A	NO
2-Trifluoromethyladenosine TP	—	A	NO
3-Deaza-3-bromoadenosine TP	—	A	NO
3-Deaza-3-chloroadenosine TP	—	A	NO
3-Deaza-3-fluoroadenosine TP	—	A	NO
3-Deaza-3-iodoadenosine TP	—	A	NO
3-Deazaadenosine TP	—	A	NO
4'-Azidoadenosine TP	—	A	NO
4'-Carbocyclic adenosine TP	—	A	NO
4'-Ethylnyladenosine TP	—	A	NO
5'-Homo-adenosine TP	—	A	NO
8-Aza-ATP	—	A	NO
8-bromo-adenosine TP	—	A	NO
8-Trifluoromethyladenosine TP	—	A	NO
9-Deazaadenosine TP	—	A	NO
2-aminopurine	—	A/G	NO
7-deaza-2,6-diaminopurine	—	A/G	NO
7-deaza-8-aza-2,6-diaminopurine	—	A/G	NO
7-deaza-8-aza-2-aminopurine	—	A/G	NO
2,6-diaminopurine	—	A/G	NO
7-deaza-8-aza-adenine, 7-deaza-2-aminopurine	—	A/G	NO
4-methylcytidine	—	C	NO
5-aza-cytidine	—	C	NO
Pseudo-iso-cytidine	—	C	NO
pyrrolo-cytidine	—	C	NO
a-thio-cytidine	—	C	NO
2-(thio)cytosine	—	C	NO
2'-Amino-2'-deoxy-CTP	—	C	NO
2'-Azido-2'-deoxy-CTP	—	C	NO
2'-Deoxy-2'-a-aminocytidine TP	—	C	NO
2'-Deoxy-2'-a-azidocytidine TP	—	C	NO
3 (deaza) 5 (aza)cytosine	—	C	NO
3 (methyl)cytosine	—	C	NO
3-(alkyl)cytosine	—	C	NO
3-(deaza) 5 (aza)cytosine	—	C	NO
3-(methyl)cytidine	—	C	NO
4,2'-O-dimethylcytidine	—	C	NO
5 (halo)cytosine	—	C	NO
5 (methyl)cytosine	—	C	NO
5 (propynyl)cytosine	—	C	NO
5-(trifluoromethyl)cytosine	—	C	NO
5-(alkyl)cytosine	—	C	NO
5-(alkynyl)cytosine	—	C	NO
5-(halo)cytosine	—	C	NO
5-(propynyl)cytosine	—	C	NO
5-(trifluoromethyl)cytosine	—	C	NO
5-bromo-cytidine	—	C	NO
5-iodo-cytidine	—	C	NO
5-propynyl cytosine	—	C	NO
6-(azo)cytosine	—	C	NO
6-aza-cytidine	—	C	NO
aza cytosine	—	C	NO
deaza cytosine	—	C	NO
N4 (acetyl)cytosine	—	C	NO
l-methyl-1-deaza-pseudoisocytidine	—	C	NO
1-methyl-pseudoisocytidine	—	C	NO

TABLE 5-continued

Exemplary non-naturally occurring modifications			
Name	Symbol	Base	Naturally Occurring
2-methoxy-5-methyl-cytidine	—	C	NO
2-methoxy-cytidine	—	C	NO
2-thio-5-methyl-cytidine	—	C	NO
4-methoxy-1-methyl-pseudoisocytidine	—	C	NO
4-methoxy-pseudoisocytidine	—	C	NO
4-thio-1-methyl-1-deaza-pseudoisocytidine	—	C	NO
4-thio-1-methyl-pseudoisocytidine	—	C	NO
4-thio-pseudoisocytidine	—	C	NO
5-aza-zebularaine	—	C	NO
5-methyl-zebularaine	—	C	NO
pyrrolo-pseudoisocytidine	—	C	NO
zebularaine	—	C	NO
(E)-5-(2-Bromo-vinyl)cytidine TP	—	C	NO
2,2'-anhydro-cytidine TP hydrochloride	—	C	NO
2'Fluor-N4-Bz-cytidine TP	—	C	NO
2'Fluoro-N4-Acetyl-cytidine TP	—	C	NO
2'-O-Methyl-N4-Acetyl-cytidine TP	—	C	NO
2'-O-methyl-N4-Bz-cytidine TP	—	C	NO
2'-a-Ethyneylcytidine TP	—	C	NO
2'-a-Trifluoromethylcytidine TP	—	C	NO
2'-b-Ethyneylcytidine TP	—	C	NO
2'-b-Trifluoromethylcytidine TP	—	C	NO
2'-Deoxy-2',2'-difluorocytidine TP	—	C	NO
2'-Deoxy-2'-a-mercaptopcytidine TP	—	C	NO
2'-Deoxy-2'-a-thiomethoxycytidine TP	—	C	NO
2'-Deoxy-2'-b-aminocytidine TP	—	C	NO
2'-Deoxy-2'-b-azidocytidine TP	—	C	NO
2'-Deoxy-2'-b-bromocytidine TP	—	C	NO
2'-Deoxy-2'-b-chlorocytidine TP	—	C	NO
2'-Deoxy-2'-b-fluorocytidine TP	—	C	NO
2'-Deoxy-2'-b-iodocytidine TP	—	C	NO
2'-Deoxy-2'-b-mercaptopcytidine TP	—	C	NO
2'-Deoxy-2'-b-thiomethoxycytidine TP	—	C	NO
2'-O-Methyl-5-(1-propynyl)cytidine TP	—	C	NO
3'-Ethyneylcytidine TP	—	C	NO
4'-Azidocytidine TP	—	C	NO
4'-Carbocyclic cytidine TP	—	C	NO
4'-Ethyneylcytidine TP	—	C	NO
5-(1-Propynyl)ara-cytidine TP	—	C	NO
5-(2-Chloro-phenyl)-2-thiocytidine TP	—	C	NO
5-(4-Amino-phenyl)-2-thiocytidine TP	—	C	NO
5-Aminoallyl-CTP	—	C	NO
5-Cyanocytidine TP	—	C	NO
5-Ethyneylara-cytidine TP	—	C	NO
5-Ethyneylcytidine TP	—	C	NO
5'-Homo-cytidine TP	—	C	NO
5-Methoxycytidine TP	—	C	NO
5-Trifluoromethyl-Cytidine TP	—	C	NO
N4-Amino-cytidine TP	—	C	NO
N4-Benzoyl-cytidine TP	—	C	NO
pseudoisocytidine	—	C	NO
6-thio-guanosine	—	G	NO
7-deaza-guanosine	—	G	NO
8-oxo-guanosine	—	G	NO
N1-methyl-guanosine	—	G	NO
a-thio-guanosine	—	G	NO
2-(propyl)guanine	—	G	NO
2-(alkyl)guanine	—	G	NO
2-Amino-2'-deoxy-GTP	—	G	NO
2'-Azido-2'-deoxy-GTP	—	G	NO
2'-Deoxy-2'-a-aminoguanosine TP	—	G	NO
2'-Deoxy-2'-a-azidoguanosine TP	—	G	NO
6-(methyl)guanine	—	G	NO
6-(alkyl)guanine	—	G	NO
6-(methyl)guanine	—	G	NO
6-methyl-guanosine	—	G	NO
7-(alkyl)guanine	—	G	NO
7-(deaza)guanine	—	G	NO
7-(methyl)guanine	—	G	NO
7-(alkyl)guanine	—	G	NO
7-(deaza)guanine	—	G	NO

TABLE 5-continued

Exemplary non-naturally occurring modifications			
Name	Symbol	Base	Naturally Occurring
7-(methyl)guanine	—	G	NO
8-(alkyl)guanine	—	G	NO
8-(alkynyl)guanine	—	G	NO
8-(halo)guanine	—	G	NO
8-(thioalkyl)guanine	—	G	NO
8-(alkenyl)guanine	—	G	NO
8-(alkyl)guanine	—	G	NO
8-(alkynyl)guanine	—	G	NO
8-(amino)guanine	—	G	NO
8-(halo)guanine	—	G	NO
8-(hydroxyl)guanine	—	G	NO
8-(thioalkyl)guanine	—	G	NO
8-(thiol)guanine	—	G	NO
azaguanine	—	G	NO
deaza guanine	—	G	NO
N (methyl)guanine	—	G	NO
N-(methyl)guanine	—	G	NO
l-methyl-6-thio-guanosine	—	G	NO
6-methoxy-guanosine	—	G	NO
6-thio-7-deaza-8-aza-guanosine	—	G	NO
6-thio-7-deaza-guanosine	—	G	NO
6-thio-7-methyl-guanosine	—	G	NO
7-deaza-8-aza-guanosine	—	G	NO
7-methyl-8-oxo-guanosine	—	G	NO
N2,N2-dimethyl-6-thio-guanosine	—	G	NO
N2-methyl-6-thio-guanosine	—	G	NO
1-Me-GTP	—	G	NO
2'Fluoro-N2-isobutyl-guanosine TP	—	G	NO
2'O-methyl-N2-isobutyl-guanosine TP	—	G	NO
2'a-Ethynguanosine TP	—	G	NO
2'a-Trifluoromethylguanosine TP	—	G	NO
2'b-Ethynguanosine TP	—	G	NO
2'b-Trifluoromethylguanosine TP	—	G	NO
2'Deoxy-2',2'-difluoroguanosine TP	—	G	NO
2'Deoxy-2'-a-mercaptopguanosine TP	—	G	NO
2'Deoxy-2'-a-thiomethoxyguanosine TP	—	G	NO
2'Deoxy-2'-b-aminoguanosine TP	—	G	NO
2'Deoxy-2'-b-azidoguanosine TP	—	G	NO
2'Deoxy-2'-b-bromoguanosine TP	—	G	NO
2'Deoxy-2'-b-chloroguanosine TP	—	G	NO
2'Deoxy-2'-b-fluoroguanosine TP	—	G	NO
2'Deoxy-2'-b-iodoguanosine TP	—	G	NO
2'Deoxy-2'-b-mercaptopguanosine TP	—	G	NO
2'Deoxy-2'-b-thiomethoxyguanosine TP	—	G	NO
4'Azidoguanosine TP	—	G	NO
4'Carbocyclic guanosine TP	—	G	NO
4'Ethynguanosine TP	—	G	NO
5'Homo-guanosine TP	—	G	NO
8-bromo-guanosine TP	—	G	NO
9-Deazaguanosine TP	—	G	NO
N2-isobutyl-guanosine TP	—	G	NO
7-methylinosine	A		NO
allyamoно-thymidine	T		NO
aza thymidine	T		NO
deaza thymidine	T		NO
deoxy-thymidine	T		NO
5-propynyl uracil	U		NO
a-thio-uridine	U		NO
1 (aminoalkylamino-carbonylethylene)-2 (thio)-pseudouracil	U		NO
1 (aminoalkylaminocarbonylethylene)-2,4-(dithio)pseudouracil	U		NO
1 (aminoalkylaminocarbonylethylene)-4-(thio)pseudouracil	U		NO
1 (aminoalkylaminocarbonylethylene)-pseudouracil	U		NO
1 (aminocarbonylethylene)-2(thio)-pseudouracil	U		NO
1 (aminocarbonylethylene)-2,4-(dithio)pseudouracil	U		NO
1 (aminocarbonylethylene)-4-(thio)pseudouracil	U		NO

TABLE 5-continued

Exemplary non-naturally occurring modifications			
Name	Symbol	Base	Naturally Occurring
1 (aminocarbonyl-ethylenyl)-pseudouracil	—	U	NO
1 substituted 2(thio)-pseudouracil	—	U	NO
1 substituted 2,4-(dithio)pseudouracil	—	U	NO
1 substituted 4 (thio)pseudouracil	—	U	NO
1 substituted pseudouracil	—	U	NO
1-(aminoalkylamino-carbonylethylenyl)-	—	U	NO
2-(thio)-pseudouracil	—	U	NO
l-Methyl-3-(3-amino-3-carboxypropyl)	—	U	NO
pseudouridine TP	—	U	NO
l-Methyl-3-(3-amino-3-carboxypropyl)pseudo-UTP	—	U	NO
1-Methyl-pseudo-UTP	—	U	NO
2 (thio)pseudouracil	—	U	NO
2' deoxy uridine	—	U	NO
2' fluorouridine	—	U	NO
2-(thio)uracil	—	U	NO
2,4-(dithio)psuedouracil	—	U	NO
2' methyl, 2'amino, 2'azido, 2'fluoro-guanosine	—	U	NO
2'-Amino-2'-deoxy-UTP	—	U	NO
2'-Azido-2'-deoxy-UTP	—	U	NO
2'-Azido-deoxyuridine TP	—	U	NO
2'-O-methylpsuedouridine	—	U	NO
2' deoxy uridine	2'dU	U	NO
2' fluorouridine	—	U	NO
2'-Deoxy-2'-a-aminouridine TP	—	U	NO
2'-Deoxy-2'-a-azidouridine TP	—	U	NO
2-methylpsuedouridine	m3'P	U	NO
3 (3 amino-3 carboxypropyl)uracil	—	U	NO
4 (thio)pseudouracil	—	U	NO
4-(thio)pseudouracil	—	U	NO
4-(thio)uracil	—	U	NO
4-thiouracil	—	U	NO
5 (1,3-diazole-1-alkyl)uracil	—	U	NO
5 (2-aminopropyl)uracil	—	U	NO
5 (aminoalkyl)uracil	—	U	NO
5 (dimethylaminooalkyl)uracil	—	U	NO
5 (guanidiniumalkyl)uracil	—	U	NO
5 (methoxycarbonylmethyl)-2-(thio)uracil	—	U	NO
5 (methoxycarbonyl-methyl)uracil	—	U	NO
5 (methyl) 2 (thio)uracil	—	U	NO
5 (methyl) 2,4 (dithio)uracil	—	U	NO
5 (methyl) 4 (thio)uracil	—	U	NO
5 (methylaminomethyl)-2 (thio)uracil	—	U	NO
5 (methylaminomethyl)-2,4 (dithio)uracil	—	U	NO
5 (methylaminomethyl)-4 (thio)uracil	—	U	NO
5 (propynyl)uracil	—	U	NO
5 (trifluoromethyl)uracil	—	U	NO
5-(2-aminopropyl)uracil	—	U	NO
5-(alkyl)-2-(thio)pseudouracil	—	U	NO
5-(alkyl)-2,4 (dithio)pseudouracil	—	U	NO
5-(alkyl)-4 (thio)pseudouracil	—	U	NO
5-(alkyl)pseudouracil	—	U	NO
5-(alkyl)uracil	—	U	NO
5-(alkynyl)uracil	—	U	NO
5-(allylamino)uracil	—	U	NO
5-(cyanoalkyl)uracil	—	U	NO
5-(dialkylaminooalkyl)uracil	—	U	NO
5-(dimethylaminooalkyl)uracil	—	U	NO
5-(guanidiniumalkyl)uracil	—	U	NO
5-(halo)uracil	—	U	NO
5-(1,3-diazole-1-alkyl)uracil	—	U	NO
5-(methoxy)uracil	—	U	NO
5-(methoxycarbonylmethyl)-2-(thio)uracil	—	U	NO
5-(methoxycarbonyl-methyl)uracil	—	U	NO
5-(methyl) 2(thio)uracil	—	U	NO
5-(methyl) 2,4 (dithio)uracil	—	U	NO
5-(methyl) 4 (thio)uracil	—	U	NO
5-(methyl)-2-(thio)pseudouracil	—	U	NO
5-(methyl)-2,4 (dithio)pseudouracil	—	U	NO
5-(methyl)-4 (thio)pseudouracil	—	U	NO

TABLE 5-continued

Exemplary non-naturally occurring modifications			
Name	Symbol	Base	Naturally Occurring
5-(methyl)pseudouracil	—	U	NO
5-(methylaminomethyl)-2 (thio)uracil	—	U	NO
5-(methylaminomethyl)-2,4(dithio)uracil	—	U	NO
5-(methylaminomethyl)-4-(thio)uracil	—	U	NO
5-(propynyl)uracil	—	U	NO
5-(trifluoromethyl)uracil	—	U	NO
5-aminoallyl-uridine	—	U	NO
5-bromo-uridine	—	U	NO
5-iodo-uridine	—	U	NO
5-uracil	—	U	NO
6 (azo)uracil	—	U	NO
6-(azido)uracil	—	U	NO
allyamino-uracil	—	U	NO
aza uracil	—	U	NO
deaza uracil	—	U	NO
N3 (methy)uracil	—	U	NO
Pseudo-UTP-1-2-ethanoic acid	—	U	NO
pseudouracil	—	U	NO
4-Thio-pseudo-UTP	—	U	NO
1-carboxymethyl-pseudouridine	—	U	NO
1-methyl-1-deaza-pseudouridine	—	U	NO
1-propynyl-uridine	—	U	NO
l-taurinomethyl-1-methyl-uridine	—	U	NO
l-taurinomethyl-4-thio-uridine	—	U	NO
l-taurinomethyl-pseudouridine	—	U	NO
2-methoxy-4-thio-pseudouridine	—	U	NO
2-thio-l-methyl-1-deaza-pseudouridine	—	U	NO
2-thio-1-methyl-pseudouridine	—	U	NO
2-thio-5-aza-uridine	—	U	NO
2-thio-dihydropseudouridine	—	U	NO
2-thio-dihydrouridine	—	U	NO
2-thio-pseudouridine	—	U	NO
4-methoxy-2-thio-pseudouridine	—	U	NO
4-methoxy-pseudouridine	—	U	NO
4-thio-1-methyl-pseudouridine	—	U	NO
4-thio-pseudouridine	—	U	NO
5-aza-uridine	—	U	NO
dihydropseudouridine	—	U	NO
(±)1-(2-Hydroxypropyl)pseudouridine TP	—	U	NO
(2R)-1-(2-Hydroxypropyl)pseudouridine	—	U	NO
TP	—	U	NO
(2S)-1-(2-Hydroxypropyl)pseudouridine	—	U	NO
TP	—	U	NO
(E)-5-(2-Bromo-vinyl)ara-uridine TP	—	U	NO
(E)-5-(2-Bromo-vinyl)uridine TP	—	U	NO
(Z)-5-(2-Bromo-vinyl)ara-uridine TP	—	U	NO
(Z)-5-(2-Bromo-vinyl)uridine TP	—	U	NO
1-(2,2,2-Trifluoroethyl)-pseudo-UTP	—	U	NO
1-(2,2,3,3-	—	U	NO
Pentafluoropropyl)pseudouridine TP	—	U	NO
1-(2,2-Diethoxyethyl)pseudouridine TP	—	U	NO
1-(2,4,6-Trimethylbenzyl)pseudouridine	—	U	NO
TP	—	U	NO
1-(2,4,6-Trimethyl-benzyl)pseudo-UTP	—	U	NO
1-(2,4,6-Trimethyl-phenyl)pseudo-UTP	—	U	NO
1-(2-Amino-2-carboxyethyl)pseudo-UTP	—	U	NO
1-(2-Amino-ethyl)pseudo-UTP	—	U	NO
1-(2-Hydroxyethyl)pseudouridine TP	—	U	NO
1-(2-Methoxyethyl)pseudouridine TP	—	U	NO
1-(3,4-Bis-trifluoromethoxybenzyl)pseudouridine	—	U	NO
TP	—	U	NO
1-(3,4-Dimethoxybenzyl)pseudouridine	—	U	NO
TP	—	U	NO
1-(3-Amino-3-carboxypropyl)pseudo-UTP	—	U	NO
1-(3-Amino-propyl)pseudo-UTP	—	U	NO
1-(3-Cyclopropyl-prop-2-enyl)pseudouridine TP	—	U	NO
1-(4-Amino-4-carboxybutyl)pseudo-UTP	—	U	NO
1-(4-Amino-benzyl)pseudo-UTP	—	U	NO

TABLE 5-continued

Exemplary non-naturally occurring modifications			
Name	Symbol	Base	Naturally Occurring
1-(4-Amino-butyl)pseudo-UTP	—	U	NO
1-(4-Amino-phenyl)pseudo-UTP	—	U	NO
1-(4-Azidobenzyl)pseudouridine TP	—	U	NO
1-(4-Bromobenzyl)pseudouridine TP	—	U	NO
1-(4-Chlorobenzyl)pseudouridine TP	—	U	NO
1-(4-Fluorobenzyl)pseudouridine TP	—	U	NO
1-(4-Iodobenzyl)pseudouridine TP	—	U	NO
1-(4-	—	U	NO
Methanesulfonylbenzyl)pseudouridine TP	—	U	NO
1-(4-Methoxybenzyl)pseudouridine TP	—	U	NO
1-(4-Methoxy-benzyl)pseudo-UTP	—	U	NO
1-(4-Methoxy-phenyl)pseudo-UTP	—	U	NO
1-(4-Methylbenzyl)pseudouridine TP	—	U	NO
1-(4-Methyl-benzyl)pseudo-UTP	—	U	NO
1-(4-Nitrobenzyl)pseudouridine TP	—	U	NO
1-(4-Nitro-benzyl)pseudo-UTP	—	U	NO
1-(4-Nitro-phenyl)pseudo-UTP	—	U	NO
1-(4-Thiomethoxybenzyl)pseudouridine TP	—	U	NO
1-(4-	—	U	NO
Trifluoromethoxybenzyl)pseudouridine TP	—	U	NO
1-(4-	—	U	NO
Trifluoromethylbenzyl)pseudouridine TP	—	U	NO
1-(5-Amino-pentyl)pseudo-UTP	—	U	NO
1-(6-Amino-hexyl)pseudo-UTP	—	U	NO
1,6-Dimethyl-pseudo-UTP	—	U	NO
1-[3-(2-{2-[2-(2-Aminoethoxy)-ethoxy}-ethoxy)-ethoxy]-propionyl]pseudouridine TP	—	U	NO
1-[3-{2-(2-Aminoethoxy)-ethoxy}-propionyl] pseudouridine TP	—	U	NO
1-Acetyl)pseudouridine TP	—	U	NO
1-Alkyl-6-(1-propynyl)-pseudo-UTP	—	U	NO
1-Alkyl-6-(2-propynyl)-pseudo-UTP	—	U	NO
1-Alkyl-6-allyl-pseudo-UTP	—	U	NO
1-Alkyl-6-ethynyl-pseudo-UTP	—	U	NO
1-Alkyl-6-homoallyl-pseudo-UTP	—	U	NO
1-Alkyl-6-vinyl-pseudo-UTP	—	U	NO
1-Allylpseudouridine TP	—	U	NO
1-Aminomethyl-pseudo-UTP	—	U	NO
1-Benzoylpseudouridine TP	—	U	NO
1-Benzylloxymethylpseudouridine TP	—	U	NO
1-Benzyl-pseudo-UTP	—	U	NO
1-Biotinyl-PEG2-pseudouridine TP	—	U	NO
1-Biotinylpseudouridine TP	—	U	NO
1-Butyl-pseudo-UTP	—	U	NO
1-Cyanomethylpseudouridine TP	—	U	NO
1-Cyclobutylmethyl-pseudo-UTP	—	U	NO
1-Cyclobutyl-pseudo-UTP	—	U	NO
1-Cycloheptylmethyl-pseudo-UTP	—	U	NO
1-Cycloheptyl-pseudo-UTP	—	U	NO
1-Cyclohexylmethyl-pseudo-UTP	—	U	NO
1-Cyclohexyl-pseudo-UTP	—	U	NO
1-Cyclooctylmethyl-pseudo-UTP	—	U	NO
1-Cyclooctyl-pseudo-UTP	—	U	NO
1-Cyclopentylmethyl-pseudo-UTP	—	U	NO
1-Cyclopentyl-pseudo-UTP	—	U	NO
1-Cyclopropylmethyl-pseudo-UTP	—	U	NO
1-Cyclopropyl-pseudo-UTP	—	U	NO
1-Ethyl-pseudo-UTP	—	U	NO
1-Hexyl-pseudo-UTP	—	U	NO
1-Homoallylpseudouridine TP	—	U	NO
1-Hydroxymethylpseudouridine TP	—	U	NO
1-iso-propyl-pseudo-UTP	—	U	NO
1-Me-2-thio-pseudo-UTP	—	U	NO
1-Me-4-thio-pseudo-UTP	—	U	NO
1-Me-alpha-thio-pseudo-UTP	—	U	NO
1-Methanesulfonylmethylpseudouridine TP	—	U	NO

TABLE 5-continued

Exemplary non-naturally occurring modifications			
Name	Symbol	Base	Naturally Occurring
1-Methoxymethylpseudouridine TP	—	U	NO
1-Methyl-6-(2,2,2-Trifluoroethyl)pseudo-UTP	—	U	NO
1-Methyl-6-(4-morpholino)-pseudo-DTP	—	U	NO
1-Methyl-6-(4-thiomorpholino)-pseudo-UTP	—	U	NO
1-Methyl-6-(substituted phenyl)pseudo-UTP	—	U	NO
1-Methyl-6-amino-pseudo-UTP	—	U	NO
1-Methyl-6-azido-pseudo-UTP	—	U	NO
1-Methyl-6-bromo-pseudo-UTP	—	U	NO
1-Methyl-6-butyl-pseudo-UTP	—	U	NO
1-Methyl-6-chloro-pseudo-UTP	—	U	NO
1-Methyl-6-cyano-pseudo-UTP	—	U	NO
1-Methyl-6-dimethylamino-pseudo-UTP	—	U	NO
1-Methyl-6-ethoxy-pseudo-UTP	—	U	NO
1-Methyl-6-ethylcarboxylate-pseudo-UTP	—	U	NO
1-Methyl-6-ethyl-pseudo-UTP	—	U	NO
1-Methyl-6-fluoro-pseudo-UTP	—	U	NO
1-Methyl-6-formyl-pseudo-UTP	—	U	NO
1-Methyl-6-hydroxyamino-pseudo-UTP	—	U	NO
1-Methyl-6-hydroxy-pseudo-UTP	—	U	NO
1-Methyl-6-iodo-pseudo-UTP	—	U	NO
1-Methyl-6-isopropyl-pseudo-UTP	—	U	NO
1-Methyl-6-methoxy-pseudo-UTP	—	U	NO
1-Methyl-6-methylamino-pseudo-UTP	—	U	NO
1-Methyl-6-phenyl-pseudo-UTP	—	U	NO
1-Methyl-6-propyl-pseudo-UTP	—	U	NO
1-Methyl-6-tert-butyl-pseudo-UTP	—	U	NO
1-Methyl-6-trifluoromethoxy-pseudo-UTP	—	U	NO
1-Methyl-6-trifluoromethyl-pseudo-UTP	—	U	NO
1-Morpholinomethylpseudouridine TP	—	U	NO
1-Pentyl-pseudo-UTP	—	U	NO
1-Phenyl-pseudo-UTP	—	U	NO
1-Pivaloylpseudouridine TP	—	U	NO
1-Propargylpseudouridine TP	—	U	NO
1-Propyl-pseudo-UTP	—	U	NO
1-propynyl-pseudouridine	—	U	NO
1-p-tolyl-pseudo-UTP	—	U	NO
1-tert-Butyl-pseudo-UTP	—	U	NO
1-Thiomethoxymethylpseudouridine TP	—	U	NO
1-Thiomorpholinomethylpseudouridine TP	—	U	NO
1-Trifluoroacetylpsudouridine TP	—	U	NO
1-Trifluoromethyl-pseudo-UTP	—	U	NO
1-Vinylpseudouridine TP	—	U	NO
2,2'-anhydro-uridine TP	—	U	NO
2'-bromo-deoxyuridine TP	—	U	NO
2'-F-5-Methyl-2'-deoxy-UTP	—	U	NO
2'-OMe-5-Me-UTP	—	U	NO
2'-OMe-pseudo-UTP	—	U	NO
2'-a-Ethynyluridine TP	—	U	NO
2'-a-Trifluoromethyluridine TP	—	U	NO
2'-b-Ethynyluridine TP	—	U	NO
2'-b-Trifluoromethyluridine TP	—	U	NO
2'-Deoxy-2',2'-difluorouridine TP	—	U	NO
2'-Deoxy-2'-a-mercaptopuridine TP	—	U	NO
2'-Deoxy-2'-a-thiomethoxyuridine TP	—	U	NO
2'-Deoxy-2'-b-aminouridine TP	—	U	NO
2'-Deoxy-2'-b-azidouridine TP	—	U	NO
2'-Deoxy-2'-b-bromouridine TP	—	U	NO
2'-Deoxy-2'-b-chlorouridine TP	—	U	NO
2'-Deoxy-2'-b-fluorouridine TP	—	U	NO
2'-Deoxy-2'-b-iodouridine TP	—	U	NO
2'-Deoxy-2'-b-mercaptopuridine TP	—	U	NO
2'-Deoxy-2'-b-thiomethoxyuridine TP	—	U	NO
2-methoxy-4-thio-uridine	—	U	NO
2-methoxyuridine	—	U	NO
2'-O-Methyl-5-(1-propynyl)uridine TP	—	U	NO
3-Alkyl-pseudo-UTP	—	U	NO

TABLE 5-continued

Exemplary non-naturally occurring modifications			
Name	Symbol	Base	Naturally Occurring
4'-Azidouridine TP	—	U	NO
4'-Carbocyclic uridine TP	—	U	NO
4'-Ethylnyluridine TP	—	U	NO
5-(1-Propynyl)ara-uridine TP	—	U	NO
5-(2-Furanyl)uridine TP	—	U	NO
5-Cyanouridine TP	—	U	NO
5-Dimethylaminouridine TP	—	U	NO
5'-Homo-uridine TP	—	U	NO
5-iodo-2'-fluoro-deoxyuridine TP	—	U	NO
5-Phenylethylnyluridine TP	—	U	NO
5-Trideuteromethyl-6-deuterouridine TP	—	U	NO
5-Trifluoromethyl-Uridine TP	—	U	NO
5-Vinylarauridine TP	—	U	NO
6-(2,2,2-Trifluoroethyl)-pseudo-UTP	—	U	NO
6-(4-Morpholino)-pseudo-DTP	—	U	NO
6-(Substituted-Phenyl)-pseudo-UTP	—	U	NO
6-Amino-pseudo-UTP	—	U	NO
6-Azido-pseudo-UTP	—	U	NO
6-Bromo-pseudo-UTP	—	U	NO
6-Butyl-pseudo-UTP	—	U	NO
6-Chloro-pseudo-UTP	—	U	NO
6-Cyano-pseudo-UTP	—	U	NO
6-Dimethylamino-pseudo-UTP	—	U	NO
6-Ethoxy-pseudo-UTP	—	U	NO
6-Ethylcarboxylate-pseudo-UTP	—	U	NO
6-Ethyl-pseudo-UTP	—	U	NO
6-Fluoro-pseudo-UTP	—	U	NO
6-Formyl-pseudo-UTP	—	U	NO
6-Hydroxyamino-pseudo-UTP	—	U	NO
6-Hydroxy-pseudo-UTP	—	U	NO
6-Iodo-pseudo-UTP	—	U	NO
6-iso-Propyl-pseudo-UTP	—	U	NO
6-Methoxy-pseudo-UTP	—	U	NO
6-Methylamino-pseudo-UTP	—	U	NO
6-Methyl-pseudo-UTP	—	U	NO
6-Phenyl-pseudo-UTP	—	U	NO
6-Phenyl-pseudo-UTP	—	U	NO
6-Propyl-pseudo-UTP	—	U	NO
6-tert-Butyl-pseudo-UTP	—	U	NO
6-Trifluoromethoxy-pseudo-UTP	—	U	NO
6-Trifluoromethyl-pseudo-UTP	—	U	NO
Alpha-thio-pseudo-UTP	—	U	NO
Pseudouridine 1-(4-methylbenzenesulfonic acid) TP	—	U	NO
Pseudouridine 1-(4-methylbenzoic acid) TP	—	U	NO
Pseudouridine TP 1-[3-(2-ethoxy)]propionic acid	—	U	NO
Pseudouridine TP 1-[3-{2-(2-[2-(2-ethoxy)-ethoxy]-ethoxy)-ethoxy}]propionic acid	—	U	NO
Pseudouridine TP 1-[3-{2-(2-[2-(2-ethoxy)-ethoxy]-ethoxy)-ethoxy}]propionic acid	—	U	NO
Pseudouridine TP 1-[3-{2-(2-[2-ethoxy]-ethoxy)-ethoxy}]propionic acid	—	U	NO
Pseudouridine TP 1-[3-{2-(2-ethoxy)-ethoxy}]propionic acid	—	U	NO
Pseudouridine TP 1-methylphosphonic acid	—	U	NO
Pseudouridine TP 1-methylphosphonic acid diethyl ester	—	U	NO
Pseudo-UTP-N1-3-propionic acid	—	U	NO
Pseudo-UTP-N1-4-butanoic acid	—	U	NO
Pseudo-UTP-N 1-5-pentanoic acid	—	U	NO
Pseudo-UTP-N1-6-hexanoic acid	—	U	NO
Pseudo-UTP-N1-7-heptanoic acid	—	U	NO
Pseudo-UTP-N1-methyl-1-p-benzoic acid	—	U	NO
Pseudo-UTP-N1-p-benzoic acid	—	U	NO

[0257] In some embodiments, a genetic element described herein comprises a modification provided in Table 6, or a combination thereof. The modifications provided in Table 6 occur naturally in RNAs, and may be used herein in a genetic element at a position that does not occur in nature.

TABLE 6

Name	Symbol	Base	Naturally Oc- curring
2-methylthio-N6-(cis-hydroxvinopentenyl)adenosine	ms2i6A	A	YES
2-methylthio-N6-methyladenosine	ms2m6A	A	YES
2-methylthio-N6-threonyl carbamoyladenosine	ms2t6A	A	YES
N6-glycylcarbamoyladenosine	g6A	A	YES
N6-isopentenyladenosine	i6A	A	YES
N6-methyladenosine	m6A	A	YES
N6-threonylcarbamoyladenosine	t6A	A	YES
1,2'-O-dimethyladenosine	mlAm	A	YES
1-methyladenosine	mA	A	YES
2'-O-methyladenosine	Am	A	YES
2'-O-ribosyladenosine (phosphate)	Ar(p)	A	YES
2-methyladenosine	m2A	A	YES
2-methylthio-N6 isopentenyladenosine	ms2i6A	A	YES
2-methylthio-N6-hydroxynorvalyl carbamoyladenosine	ms2hn6A	A	YES
2'-O-methyladenosine	m6A	A	YES
2'-O-ribosyladenosine (phosphate)	Ar(p)	A	YES
isopentenyl adenosine	Iga	A	YES
N6-(cis-hydroxyisopentenyl)adenosine	io6A	A	YES
N6,2'-O-dimethyladenosine	m6Am	A	YES
N6,2'-O-dimethyladenosine	m'6Am	A	YES
N6,N6,2'-O-trimethyladenosine	m62Am	A	YES
N6,N6-dimethyladenosine	m62A	A	YES
N6-acetyladenosine	ac6A	A	YES
N6-hydroxynorvalylcarbamoyladenosine	hn6A	A	YES
N6-methyl-N6-threonylcarbamoyladenosine	m6t6A	A	YES
2-methyladenosine	m <sup>2</sup> A	A	YES
2-methylthio-N <sup>6</sup> -isopentenyl adenosine	ms <sup>2</sup> i6A	A	YES
2-thiocytidine	s2C	C	YES
3-methylcytidine	m3C	C	YES
5-formylcytidine	f5C	C	YES
5-hydroxymethylcytidine	hm5C	C	YES
5-methylcytidine	m5C	C	YES
N4-acetylcytidine	ac4C	C	YES
2'-O-methylcytidine	Cm	C	YES
2'-O-methylcytidine	Cm	C	YES
5,2'-O-dimethylcytidine	m5Cm	C	YES
5-formyl-2'-O-methylcytidine	f5Cm	C	YES
lysidine	k2C	C	YES
N4,2'-O-dimethylcytidine	m4Cm	C	YES
N4-acetyl-2'-O-methylcytidine	ac4Cm	C	YES
N4-methylcytidine	m4C	C	YES
N4,N4-Dimethyl-2'-OMe-Cytidine TP	—	C	YES
7-methylguanosine	m7G	G	YES
N2,2'-O-dimethylguanosine	m2Gm	G	YES
N2-methylguanosine	m2G	G	YES
wysidine	imG	G	YES
1,2'-O-dimethylguanosine	mlGm	G	YES
1-methylguanosine	mlG	G	YES
2'-O-methylguanosine	Gm	G	YES
2'-O-ribosylguanosine (phosphate)	Gr(p)	G	YES
2'-O-methylguanosine	Gm	G	YES
2'-O-ribosylguanosine (phosphate)	Gr(p)	G	YES
7-aminothethyl-7-deazaguanosine	preQ1	G	YES
7-cyano-7-deazaguanosine	preQ0	G	YES
archaeosine	G+	G	YES
methylwyosine	mimG	G	YES
N2,7-dimethylguanosine	m2,7G	G	YES
N2,N2,2'-O-trimethylguanosine	m22Gm	G	YES
N2,N2,7-trimethylguanosine	m2,2,7G	G	YES

TABLE 6-continued

Additional exemplary modifications			
Name	Symbol	Base	Natur-ally Oc-curring
N2,N2-dimethylguanosine	m22G	G	YES
N2,7,2'-O-trimethylguanosine	m2 7	G	YES
	'Gm		
1-methylinosine	mll	A	YES
inosine	I	A	YES
1,2'-O-dimethylinosine	mlim	A	YES
2'-O-methylinosine	Im	A	YES
2'-O-methylinosine	Im	A	YES
epoxyqueuosine	oQ	G	YES
galactosyl-queuosine	galQ	G	YES
mannosyl-queuosine	manQ	G	YES
2'-O-methyluridine	—	U	YES
2-thiouridine	s2U	U	YES
3-methyluridine	m3U	U	YES
5-carboxymethyluridine	cm5U	U	YES
5-hydroxyuridine	ho5U	U	YES
5-methyluridine	m5U	U	YES
5-taurinomethyl-2-thiouridine	rm5s2U	U	YES
5-taurinomethyluridine	rm5U	U	YES
dihydrouridine	D	U	YES
pseudouridine	Q	U	YES
(3-(3-amino-3-carboxypropyl)uridine	acp3U	U	YES
1-methyl-3-(3-amino-5-carboxypropyl)pseudouridine	mlacp3'P	U	YES
1-methylpseudo尿idine	ml'P	U	YES
1-methyl-pseudouridine	—	U	YES
2'-O-methyluridine	Um	U	YES
2'-O-methylpseudo尿idine	P'm	U	YES
2'-O-methyluridine	Um	U	YES
2-thio-2'-O-methyluridine	s2Um	U	YES
3-(3-amino-3-carboxypropyl)uridine	acp3U	U	YES
3,2'-O-dimethyluridine	m3Um	U	YES
3-Methyl-pseudo-Uridine TP	—	U	YES
4-thiouridine	s4U	U	YES
5-(carboxyhydroxymethyl)uridine	chm5U	U	YES
5-(carboxyhydroxymethyl)uridine methyl ester	mchm5U	U	YES
5,2'-O-dimethyluridine	m5Um	U	YES
5,6-dihydro-uridine	—	U	YES
5-aminomethyl-2-thiouridine	nm5s2U	U	YES
5-carbamoylmethyl-2'-O-methyluridine	nem5Um	U	YES
5-carbamoylmethyluridine	nem5U	U	YES
5-carboxyhydroxymethyluridine	—	U	YES
5-carboxyhydroxymethyluridine methyl ester	—	U	YES
5-carboxymethylaminomethyl-2'-O-methyluridine	cmmm5Um	U	YES
5-carboxymethylaminomethyl-2-thiouridine	cmmm5s2U	U	YES
5-carboxymethylaminomethyl-2-thiouridine	—	U	YES
5-carboxymethylaminomethyluridine	cmm5U	U	YES
5-carboxymethylaminomethyluridine	—	U	YES
5-Carbamoylmethyluridine TP	—	U	YES
5-methoxycarbonylmethyl-2'-O-methyluridine	mcm5Um	U	YES
5-methoxycarbonylmethyl-2-thiouridine	mcm5s2U	U	YES
5-methoxycarbonylmethyluridine	mcm5U	U	YES
5-methoxyuridine	m05U	U	YES
5-methyl-2-thiouridine	m5s2U	U	YES
5-methylaminomethyl-2-selenouridine	mm5se2U	U	YES
5-methylaminomethyl-2-thiouridine	mm5s2U	U	YES
5-methylaminomethyluridine	mm5U	U	YES
5-Methyldehydrouridine	—	U	YES
5-Oxyacetic acid-Uridine TP	—	U	YES
5-Oxyacetic acid-methyl ester-Uridine TP	—	U	YES
N1-methyl-pseudo-uridine	—	U	YES
uridine 5-oxyacetic acid	emo5U	U	YES
uridine 5-oxyacetic acid methyl ester	memo5U	U	YES

TABLE 6-continued

Additional exemplary modifications			
Name	Symbol	Base	Natur-ally Oc-curring
3-(3-Amino-3-carboxypropyl)-Uridine	—	U	YES
TP	—	U	YES
5-(iso-Pentenylaminomethyl)-2-thiouridine	—	U	YES
5-(iso-Pentenylaminomethyl)-2'-O-methyluridine	—	U	YES
5-(iso-Pentenylaminomethyl)uridine	—	U	YES
wybutosine	yW	A/T	YES
hydroxywybutosine	OHyW	A/T	YES
isowyosine	imG2	A/T	YES
peroxywybutosine	o2yW	A/T	YES
undermodified hydroxywybutosine	OHyW*	A/T	YES
4-demethylwyosine	imG-14	A/T	YES

[0258] In an embodiment, a genetic element described herein comprises a non-naturally occurring modification provided in Table 7, or a combination thereof.

TABLE 7

Additional exemplary non-naturally occurring modifications	
Name	
2,6-(diamino)purine	
1-(aza)-2-(thio)-3-(aza)-phenoxazin-1-yl	
1,3-(diaz)-2-(oxo)-phenthiazin-1-yl	
1,3-(diaz)-2-(oxo)-phenoxazin-1-yl	
1,3,5-(triaza)-2,6-(dioxa)-naphthalene	
2 (amino)purine	
2,4,5-(trimethyl)phenyl	
2'methyl, 2'amino, 2'azido, 2'fluro-cytidine	
2'methyl, 2'amino, 2'azido, 2'fluro-adenine	
2'methyl, 2'amino, 2'azido, 2'fluro-uridine	
2'-amino-2'-deoxyribose	
2-amino-6-Chloro-purine	
2-aza-inosinyl	
2'-azido-2'-deoxyribose	
2'fluoro-2'-deoxyribose	
2'-fluoro-modified bases	
2'-O-methyl-ribose	
2-oxo-7-aminopyridopyrimidin-3-yl	
2-oxo-pyridopyrimidine-3-yl	
2-pyridinone	
3 nitropyrrole	
3-(methyl)-7-(propynyl)isocarbostyryl	
3-(methyl)isocarbostyryl	
4-(fluoro)-6-(methyl)benzimidazole	
4-(methyl)benzimidazole	
4-(methyl)indolyl	
4,6-(dimethyl)indolyl	
5 nitroindole	
5 substituted pyrimidines	
5-(methyl)isocarbostyryl	
5-nitroindole	
6-(aza)pyrimidine	
6-(azo)thymine	
6-(methyl)-7-(aza)indolyl	
6-chloro-purine	
6-phenyl-pyrrolo-pyrimidin-2-on-3-yl	
7-(aminoalkylhydroxy)-1-(aza)-2-(thio)-3-(aza)-phenothiazin-1-yl	
7-(aminoalkylhydroxy)-1-(aza)-2-(thio)-3-(aza)-phenoxazin-1-yl	
7-(aminoalkylhydroxy)-1,3-(diaz)-2-(oxo)-phenoxazin-1-yl	
7-(aminoalkylhydroxy)-1,3-(diaz)-2-(oxo)-phenothiazin-1-yl	
7-(aminoalkylhydroxy)-1,3-(diaz)-2-(oxo)-phenoxazin-1-yl	

TABLE 7-continued

Additional exemplary non-naturally occurring modifications	
Name	
7-(aza)indolyl	
7-(guanidiniumalkylhydroxy)-1-(aza)-2-(thio)-3-(aza)-phenoxyazin-1-yl	
7-(guanidiniumalkylhydroxy)-1-(aza)-2-(thio)-3-(aza)-phenthiazin-1-yl	
7-(guanidiniumalkylhydroxy)-1-(aza)-2-(thio)-3-(aza)-phenoxyazin-1-yl	
7-(guanidiniumalkylhydroxy)-1,3-(diazia)-2-(oxo)-phenoxyazin-1-yl	
7-(guanidiniumalkyl-hydroxy)-1,3-(diazia)-2-(oxo)-phenthiazin-1-yl	
7-(guanidiniumalkylhydroxy)-1,3-(diazia)-2-(oxo)-phenoxyazin-1-yl	
7-(propynyl)isocarbostyryl	
7-(propynyl)isocarbostyryl, propynyl-7-(aza)indolyl	
7-deaza-inosinyl	
7-substituted 1-(aza)-2-(thio)-3-(aza)-phenoxyazin-1-yl	
7-substituted 1,3-(diazia)-2-(oxo)-phenoxyazin-1-yl	
9-(methyl)-imidizopyridinyl	
aminoindolyl	
anthracenyl	
bis-ortho-(aminoalkylhydroxy)-6-phenyl-pyrrolo-nvrimidin-2-on-3-yl	
bis-ortho-substituted-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl	
difluorotolyl	
hypoxanthine	
imidizopyridinyl	
inosinyl	
isocarbostyryl	
isoguanine	
N2-substituted purines	
N6-methyl-2-amino-purine	
N6-substituted purines	
N-alkylated derivative	
naphthalenyl	
nitrobenzimidazolyl	
nitroimidazolyl	
nitroindazolyl	
nitropyrazolyl	
nubularine	
O6-substituted purines	
O-alkylated derivative	
ortho-(aminoalkylhydroxy)-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl	
ortho-substituted-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl	
Oxoformycin TP	
para-(aminoalkylhydroxy)-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl	
para-substituted-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl	
pentacenyl	
phenanthracenyl	
phenyl	
propynyl-7-(aza)indolyl	
pyrenyl	
pyridopyrimidin-3-yl	
pyridopyrimidin-3-yl, 2-oxo-7-amino-pyridopyrimidin-3-yl	
pyrrolo-pyrimidin-2-on-3-yl	
pyrrolopyrimidinyl	
pyrrolopyrizinyl	
stilbenyl	
substituted 1,2,4-triazoles	
tetracenyl	
tubercidine	
xanthine	
Xanthosine-5'-TP	
2-thio-zebularine	
5-aza-2-thio-zebularine	
7-deaza-2-amino-purine	
pyridin-4-one ribonucleoside	
2-Amino-riboside-TP	
Formycin A TP	
Formycin B TP	

TABLE 7-continued

Additional exemplary non-naturally occurring modifications	
Name	
Pyrrolosine TP	
2'-OH-ara-adenosine TP	
2'-OH-ara-cytidine TP	
2'-OH-ara-uridine TP	
2'-OH-ara-guanosine TP	
5-(2-carbomethoxyvinyl)uridine TP	
N6-(19-Amino-pentaoxanadecyl)adenosine TP	

[0259] In an embodiment, a genetic element described herein comprises a non-naturally occurring modification provided in Table 8, or a combination thereof.

TABLE 8

Exemplary backbone modifications	
Name	
3'-alkylene phosphonates	
3'-amino phosphoramidate	
alkene containing backbones	
aminoalkylphosphoramidates	
aminoalkylphosphotriesters	
boranophosphates	
—CH <sub>2</sub> -O-N(CH <sub>3</sub> )-CH <sub>2</sub> —	
—CH <sub>2</sub> -N(CH <sub>3</sub> )-N(CH <sub>3</sub> )-CH <sub>2</sub> —	
—CH <sub>2</sub> -NH-CH <sub>2</sub> —	
chiral phosphonates	
chiral phosphorothioates	
formacetyl and thioformacetyl backbones	
methylene (methylimino)	
methylene formacetyl and thioformacetyl backbones	
methyleneimino and methylenehydrazino backbones	
morpholino linkages	
—N(CH <sub>3</sub> )-CH <sub>2</sub> -CH <sub>2</sub> —	
oligonucleosides with heteroatom internucleoside linkage	
phosphinates	
phosphoramidates	
phosphorodithioates	
phosphorothioate internucleoside linkages	
phosphorothioates	
phosphotriesters	
PNA	
siloxane backbones	
sulfamate backbones	
sulfide sulfoxide and sulfone backbones	
sulfonamide backbones	
thioalkylphosphonates	
thioalkylphosphotriesters	
thionophosphoramidates	

[0260] In an embodiment, a genetic element described herein comprises a non-naturally occurring modification provided in Table 9, or a combination thereof.

TABLE 9

Exemplary non-naturally occurring backbone modifications	
Name of synthetic backbone modifications	
Phosphorothioate	
Constrained nucleic acid (CNA)	
2'O/methylation	
2'-O-methoxyethylribose (MOE)	
2'Fluoro	
Locked nucleic acid (LNA)	
(S)-constrained ethyl (cEt)	
Fluoro hexitol nucleic acid (FHNA)	
5'phosphorothioate	
Phosphorodiamidate Morpholino Oligomer (PMO)	
Tricyclo-DNA (tcDNA)	

TABLE 9-continued

Exemplary non-naturally occurring backbone modifications
Name of synthetic backbone modifications
(S) 5'-C-methyl
(E)-vinylphosphonate
Methyl phosphonate
(S) 5'-C-methyl with phosphate

[0261] In some embodiments, the genetic element comprises a cap. A cap is typically placed at the 5' end of an mRNA, but a cap can also be positioned at the 3' end of an RNA. In some embodiments, a cap protects the genetic element from exonuclease degradation, and can help in delivery and/or localization within a cell. The cap can be present at the 5'-terminus (5'-cap) or at the 3'-terminal (3'-cap) or can be present on both termini. Non-limiting examples of a 5'-cap include, but are not limited to, glyceryl, inverted deoxy abasic residue (moiety); 4',5'-methylene nucleotide; 1-(beta-D-erythrofuranosyl) nucleotide, 4'-thio nucleotide; carbocyclic nucleotide; 1,5-anhydrohexitol nucleotide; L-nucleotides; alpha-nucleotides; modified base nucleotide; phosphorodithioate linkage; threo-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; acyclic 3,4-dihydroxybutyl nucleotide; acyclic 3,5-dihydroxypentyl nucleotide, 3'-3'-inverted nucleotide moiety; 3'-3'-inverted abasic moiety; 3'-2'-inverted nucleotide moiety; 3'-2'-inverted abasic moiety; 1,4-butanediol phosphate; 3'-phosphoramidate; hexylphosphate; aminohexyl phosphate; 3'-phosphate; 3'-phosphorothioate; phosphorodithioate; or bridging or non-bridging methylphosphonate moiety.

[0262] Non-limiting examples of the 3'-cap include, but are not limited to, glyceryl, inverted deoxy abasic residue (moiety), 4', 5'-methylene nucleotide; 1-(beta-D-erythrofuranosyl) nucleotide; 4'-thio nucleotide, carbocyclic nucleotide; 5'-amino-alkyl phosphate; 1,3-diamino-2-propyl phosphate; 3-aminopropyl phosphate; 6-aminohexyl phosphate; 1,2-aminododecyl phosphate; hydroxypropyl phosphate; 1,5-anhydrohexitol nucleotide; L-nucleotide; alpha-nucleotide; modified base nucleotide; phosphorodithioate; threo-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; 3,4-dihydroxybutyl nucleotide; 3,5-dihydroxypentyl nucleotide, 5'-5'-inverted nucleotide moiety; 5'-5'-inverted abasic moiety; 5'-phosphoramidate; 5'-phosphorothioate; 1,4-butanediol phosphate; 5'-amino; bridging and/or non-bridging 5'-phosphoramidate, phosphorothioate and/or phosphorodithioate, bridging or non bridging methylphosphonate and 5'-mercapto moieties (for more details see Beaucage and Iyer, 1993, Tetrahedron 49, 1925; incorporated by reference herein).

[0263] In some embodiments, the genetic element comprises a poly-A tail. In some embodiments, a poly-A tail comprises at least about 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 adenosines in length. In some embodiments, RNA lacks a poly-A tail. In some embodiments, wherein the RNA lacks a poly-A tail, the RNA comprises no more than about 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 sequential adenosines.

[0264] In some embodiments, a genetic element is linear. In some embodiments, a genetic element is circular. In some embodiments, a genetic element comprises a first region and a second region that can hybridize with the first region. In some embodiments, a genetic element comprises a first region and a second region that can hybridize with the first

region to form a circle. In some embodiments, a genetic element does not comprise a 5' or a 3' end. In some embodiments, a genetic element does not comprise one or both of a free phosphate and a free sugar. In some embodiments, every phosphate in a genetic element is covalently linked to a first sugar by a first oxygen atom comprised by the phosphate and a second sugar by a second oxygen atom comprised by a phosphate. In some embodiments, every sugar in a genetic element is covalently linked to a first phosphate by a first carbon atom comprised by the sugar and a second phosphate by a second carbon atom comprised by the sugar. In some embodiments, a genetic element is produced by circularizing a linear RNA. Circular RNAs are described, e.g., in US Patent Publication 20200306286, which is herein incorporated by reference in its entirety.

[0265] In some embodiments, a genetic element is about 10-20, 20-30, 30-40, 50-60, 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, or 4000-4500 nucleotides in length. In some embodiments a genetic element is at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, or 4500 nucleotides in length.

#### RNA-Only Genetic Elements

[0266] In some embodiments, a genetic element consists of or consists essentially of RNA. For example, in some embodiments, a genetic element is substantially free of DNA. In some embodiments, a genetic element comprises single stranded RNA. In some embodiments, a genetic element comprises at least one double stranded region. In some embodiments, a double stranded region of a genetic element comprises a region of RNA pairing with RNA.

#### Hybridized RNA-ssDNA Genetic Elements

[0267] In some embodiments, a genetic element comprises a DNA region. In some embodiments, a genetic element comprising RNA further comprises a DNA region. For example, a genetic element may be single stranded, wherein a first portion of the single strand comprises ribonucleotides and a second portion of the single strand comprises deoxyribonucleotides. In some embodiments, a genetic element comprising a DNA region comprises one or more DNA nucleotides with chemical modification. In some embodiments, a genetic element comprises a DNA region, wherein all nucleotides of the DNA region are chemically modified.

[0268] In some embodiments, at least a portion of a genetic element is single stranded. In some embodiments, a genetic element is single stranded. In some embodiments, a genetic element comprises ssDNA. In some embodiments, a genetic element comprises a double stranded region. In some embodiments, a double stranded region of a genetic element comprises a region of RNA pairing with RNA. In some such embodiments, a double stranded region of a genetic element comprises a region of DNA pairing with RNA. In some embodiments, at least a portion of the DNA region hybridizes to at least a portion of the RNA of the genetic element.

[0269] In some embodiments, a DNA region is about 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90,

90-100, 100-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 nucleotides in length.

#### RNA/DNA Conjugates

**[0270]** In some embodiments, a genetic element comprises a DNA region. In some embodiments, a genetic element comprising RNA further comprises a DNA region. In some embodiments, a genetic element comprising a DNA region comprises one or more DNA nucleotides with chemical modification. In some embodiments, a genetic element comprising a DNA region, wherein all nucleotides of the DNA region are chemically modified.

**[0271]** In some embodiments, at least a portion of a genetic element is single stranded. In some embodiments, a genetic element is single stranded. In some embodiments, a genetic element comprises ssDNA. In some embodiments, a genetic element comprises a double stranded region. In some embodiments, a double stranded region of a genetic element comprises a region of RNA pairing with RNA. In some such embodiments, a double stranded region of a genetic element comprises a region of DNA pairing with RNA. In some embodiments, wherein a genetic element comprises RNA, a DNA region is covalently linked to the RNA of the genetic element.

**[0272]** In some embodiments, a DNA region is about 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 nucleotides in length.

#### Genetic Element Constructs

**[0273]** In some embodiments, a genetic element is produced from a genetic element construct. For instance, in some embodiments, the genetic element construct is DNA, e.g., double stranded DNA, and the genetic element may be produced by transcription, generating an RNA genetic element.

**[0274]** The genetic element of an anellovector as described herein may be produced from a genetic element construct that comprises a genetic element region and optionally other sequence such as a bacmid (e.g., comprising a baculovirus genome or a fragment thereof, e.g., one or more baculovirus elements) or donor vector backbone. In some embodiments, the genetic element construct comprises an Anellovirus 5' UTR (e.g., as described herein). A genetic element construct may be any nucleic acid construct suitable for delivery of the sequence of the genetic element into a host cell or cell-free system in which the genetic element can be enclosed within a proteinaceous exterior. In some embodiments, the genetic element construct comprises a promoter. In some embodiments, transcription from the genetic element construct produces an RNA genetic element.

**[0275]** In some embodiments, the genetic element construct is a linear nucleic acid molecule. In some embodiments, the genetic element construct is a circular nucleic acid molecule (e.g., a plasmid, bacmid, donor vector, or a minicircle, e.g., as described herein). The genetic element construct may, in some embodiments, be double-stranded. In other embodiments, the genetic element is single-stranded. In some embodiments, the genetic element construct comprises DNA. In some embodiments, the genetic element

construct comprises RNA. In some embodiments, the genetic element construct comprises one or more modified nucleotides.

#### Plasmids

**[0276]** In some embodiments, the genetic element construct is a plasmid. The plasmid will generally comprise the sequence of a genetic element as described herein as well as an origin of replication suitable for replication in a host cell (e.g., a bacterial origin of replication for replication in bacterial cells) and a selectable marker (e.g., an antibiotic resistance gene). In some embodiments, the sequence of the genetic element can be excised from the plasmid. In some embodiments, the plasmid is capable of replication in a bacterial cell. In some embodiments, the plasmid is capable of replication in a mammalian cell (e.g., a human cell). In some embodiments, a plasmid is at least 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, or 5000 bp in length. In some embodiments, the plasmid is less than 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10,000 bp in length. In some embodiments, the plasmid has a length between 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-4000, or 4000-5000 bp.

#### Small Circular Nucleic Acid Constructs

**[0277]** In some embodiments, the genetic element construct is a circular nucleic acid construct, e.g., lacking a vector backbone (e.g., lacking a bacterial origin of replication and/or selectable marker). In embodiments, the genetic element is a single- or double-stranded circular nucleic acid construct. In embodiments, the circular nucleic acid construct is produced by *in vitro* circularization (IVC), e.g., as described herein. In embodiments, the double-stranded circular nucleic acid construct can be introduced into a host cell, in which it can be converted into or used as a template for generating single-stranded circular genetic elements, e.g., as described herein. In some embodiments, the circular nucleic acid construct does not comprise a plasmid backbone or a functional fragment thereof. In some embodiments, the circular nucleic acid construct is at least 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, or 4500 bp in length. In some embodiments, the circular nucleic acid construct is less than 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5500, or 6000 bp in length. In some embodiments, the circular nucleic acid construct is between 2000-2100, 2100-2200, 2200-2300, 2300-2400, 2400-2500, 2500-2600, 2600-2700, 2700-2800, 2800-2900, 2900-3000, 3000-3100, 3100-3200, 3200-3300, 3300-3400, 3400-3500, 3500-3600, 3600-3700, 3700-3800, 3800-3900, 3900-4000, 4000-4100, 4100-4200, 4200-4300, 4300-4400, or 4400-4500 bp in length. In some embodiments, the circular nucleic acid construct is a minicircle.

#### Cis/Trans Constructs

**[0278]** In some embodiments, a genetic element construct (e.g., a bacmid or donor vector) as described herein comprises one or more sequences encoding one or more Anellovirus ORFs, e.g., proteinaceous exterior components (e.g., polypeptides encoded by an Anellovirus ORF1 nucleic acid,

e.g., as described herein). For example, the genetic element construct may comprise a nucleic acid sequence encoding an Anellovirus ORF1 molecule. Such genetic element constructs can be suitable for introducing the genetic element and the Anellovirus ORF(s) into a host cell in cis. In other embodiments, a genetic element construct as described herein does not comprise sequences encoding one or more Anellovirus ORFs, e.g., proteinaceous exterior components (e.g., polypeptides encoded by an Anellovirus ORF1 nucleic acid, e.g., as described herein). For example, the genetic element construct may not comprise a nucleic acid sequence encoding an Anellovirus ORF1 molecule. Such genetic element constructs can be suitable for introducing the genetic element into a host cell, with the one or more Anellovirus ORFs to be provided in trans (e.g., via introduction of a second nucleic acid construct encoding one or more of the Anellovirus ORFs, or via an Anellovirus ORF cassette integrated into the genome of the host cell). In some embodiments, the genetic element construct comprises a backbone suitable for replication of the nucleic acid construct in insect cells (e.g., Sf9 cells), e.g., a baculovirus backbone region. In some embodiments, the genetic element construct comprises a backbone region suitable for replication of the genetic element construct in bacterial cells (e.g., *E. coli* cells, e.g., DH 10Bac cells).

[0279] In some embodiments, the genetic element construct (e.g., bacmid or donor vector) comprises a sequence encoding an Anellovirus ORF1 molecule, or a splice variant or functional fragment thereof (e.g., a jelly-roll region, e.g., as described herein). In embodiments, the portion of the genetic element that does not comprise the sequence of the genetic element comprises the sequence encoding the Anellovirus ORF1 molecule, or splice variant or functional fragment thereof (e.g., in a cassette comprising a promoter and the sequence encoding the Anellovirus ORF1 molecule, or splice variant or functional fragment thereof). In further embodiments, the portion of the construct comprising the sequence of the genetic element comprises a sequence encoding an Anellovirus ORF1 molecule, or a splice variant or functional fragment thereof (e.g., a jelly-roll region, e.g., as described herein). In embodiments, enclosure of such a genetic element in a proteinaceous exterior (e.g., as described herein) produces a replication-component anellovector (e.g., an anellovector that upon infecting a cell, enables the cell to produce additional copies of the anellovector without introducing further nucleic acid constructs, e.g., encoding one or more Anellovirus ORFs as described herein, into the cell).

[0280] In other embodiments, the genetic element does not comprise a sequence encoding an Anellovirus ORF1 molecule, or a splice variant or functional fragment thereof (e.g., a jelly-roll region, e.g., as described herein). In embodiments, enclosure of such a genetic element in a proteinaceous exterior (e.g., as described herein) produces a replication-incompetent anellovector (e.g., an anellovector that, upon infecting a cell, does not enable the infected cell to produce additional anellovectors, e.g., in the absence of one or more additional constructs, e.g., encoding one or more Anellovirus ORFs as described herein).

#### Expression Cassettes

[0281] In some embodiments, a genetic element construct (e.g., bacmid or donor vector) comprises one or more cassettes for expression of a polypeptide or noncoding RNA

(e.g., a miRNA or an siRNA). In some embodiments, the genetic element construct comprises a cassette for expression of an effector (e.g., an exogenous or endogenous effector), e.g., a polypeptide or noncoding RNA, as described herein. In some embodiments, the genetic element construct comprises a cassette for expression of an Anellovirus protein (e.g., an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, or a functional fragment thereof). The expression cassettes may, in some embodiments, be located within the genetic element sequence. In embodiments, an expression cassette for an effector is located within the genetic element sequence. In embodiments, an expression cassette for an Anellovirus protein is located within the genetic element sequence. In other embodiments, the expression cassettes are located at a position within the genetic element construct outside of the sequence of the genetic element (e.g., in the backbone). In embodiments, an expression cassette for an Anellovirus protein is located at a position within the genetic element construct outside of the sequence of the genetic element (e.g., in the backbone).

[0282] A polypeptide expression cassette generally comprises a promoter and a coding sequence encoding a polypeptide, e.g., an effector (e.g., an exogenous or endogenous effector as described herein) or an Anellovirus protein (e.g., a sequence encoding an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, or a functional fragment thereof). Exemplary promoters that can be included in an polypeptide expression cassette (e.g., to drive expression of the polypeptide) include, without limitation, constitutive promoters (e.g., CMV, RSV, PGK, EF1a, or SV40), cell or tissue-specific promoters (e.g., skeletal a-actin promoter, myosin light chain 2A promoter, dystrophin promoter, muscle creatine kinase promoter, liver albumin promoter, hepatitis B virus core promoter, osteocalcin promoter, bone sialoprotein promoter, CD2 promoter, immunoglobulin heavy chain promoter, T cell receptor a chain promoter, neuron-specific enolase (NSE) promoter, or neurofilament light-chain promoter), and inducible promoters (e.g., zinc-inducible sheep metallothionein (MT) promoter; the dexamethasone (Dex)-inducible mouse mammary tumor virus (MMTV) promoter; the T7 polymerase promoter system, tetracycline-repressible system, tetracycline-inducible system, RU486-inducible system, rapamycin-inducible system), e.g., as described herein. In some embodiments, the expression cassette further comprises an enhancer, e.g., as described herein.

#### Production of RNA-based genetic elements

[0283] An RNA-based genetic element may be produced by a variety of methods. For example, a genetic element construct comprising DNA can be transcribed to produce a genetic element that comprises RNA, e.g., as described above. The transcription may take place, e.g., in a cell or a cell-free system. RNA may be synthesized in vitro, for example, by solid phase synthesis.

#### Production of Protein Components

[0284] Protein components of an anellovector, e.g., ORF1, can be produced in a variety of ways described herein.

#### Baculovirus Expression Systems

[0285] A viral expression system, e.g., a baculovirus expression system, may be used to express proteins (e.g., for

production of anellovectors), e.g., as described herein. Baculoviruses are rod-shaped viruses with a circular, supercoiled double-stranded DNA genome. Genera of baculoviruses include: Alphabaculovirus (nucleopolyhedroviruses (NPVs) isolated from Lepidoptera), Betabaculoviruses (granuloviruses (GV) isolated from Lepidoptera), Gammabaculoviruses (NPVs isolated from Hymenoptera) and Deltabaculoviruses (NPVs isolated from Diptera). While GV's typically contain only one nucleocapsid per envelope, NPVs typically contain either single (SNPV) or multiple (MNPV) nucleocapsids per envelope. The enveloped virions are further occluded in granulin matrix in GV's and polyhedrin in NPVs. Baculoviruses typically have both lytic and occluded life cycles. In some embodiments, the lytic and occluded life cycles manifest independently throughout the three phases of virus replication: early, late, and very late phase. In some embodiments, during the early phase, viral DNA replication takes place following viral entry into the host cell, early viral gene expression and shut-off of the host gene expression machinery. In some embodiments, in the late phase late genes that code for viral DNA replication are expressed, viral particles are assembled, and extracellular virus (EV) is produced by the host cell. In some embodiments, in the very late phase the polyhedrin and p10 genes are expressed, occluded viruses (OV) are produced by the host cell, and the host cell is lysed. Since baculoviruses infect insect species, they can be used as biological agents to produce exogenous proteins in baculoviruses-permissive insect cells or larvae. Different isolates of baculovirus, such as *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) and *Bombyx mori* (silkworm) nuclear polyhedrosis virus (BmNPV) may be used in exogenous protein expression. Various baculoviral expression systems are commercially available, e.g., from ThermoFisher.

**[0286]** In some embodiments, the proteins described herein (e.g., an Anellovirus ORF molecule, e.g., ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, or a functional fragment or splice variant thereof) may be expressed using a baculovirus expression vector (e.g., a bacmid) that comprises one or more components described herein. For example, a baculovirus expression vector may include one or more of (e.g., all of) a selectable marker (e.g., kanR), an origin of replication (e.g., one or both of a bacterial origin of replication and an insect cell origin of replication), a recombinase recognition site (e.g., an att site), and a promoter. In some embodiments, a baculovirus expression vector (e.g., a bacmid as described herein) can be produced by replacing the naturally occurring wild-type polyhedrin gene, which encodes for baculovirus occlusion bodies, with genes encoding the proteins described herein. In some embodiments, the genes encoding the proteins described herein are cloned into a baculovirus expression vector (e.g., a bacmid as described herein) containing a baculovirus promoter. In some embodiments, the baculoviral vector comprises one or more non-baculoviral promoters, e.g., a mammalian promoter or an Anellovirus promoter. In some embodiments, the genes encoding the proteins described herein are cloned into a donor vector (e.g., as described herein), which is then contacted with an empty baculovirus expression vector (e.g., an empty bacmid) such that the genes encoding the proteins described herein are transferred (e.g., by homologous recombination or transposase activity) from the donor vector into the baculovirus expression vector (e.g., bacmid). In some embodiments, the baculovirus pro-

moter is flanked by baculovirus DNA from the nonessential polyhedrin gene locus. In some embodiments, a protein described herein is under the transcriptional control of the AcNPV polyhedrin promoter in the very late phase of viral replication. In some embodiments, a strong promoter suitable for use in baculoviral expression in insect cells include, but are not limited to, baculovirus p10 promoters, polyhedrin (polh) promoters, p6.9 promoters and capsid protein promoters. Weak promoters suitable for use in baculoviral expression in insect cells include ie1, ie2, ie0, et1, 39K (aka pp31) and gp64 promoters of baculoviruses.

**[0287]** In some embodiments, a recombinant baculovirus is produced by homologous recombination between a baculoviral genome (e.g., a wild-type or mutant baculoviral genome), and a transfer vector. In some embodiments, one or more genes encoding a protein described herein are cloned into the transfer vector. In some embodiments, the transfer vector further contains a baculovirus promoter flanked by DNA from a nonessential gene locus, e.g., polyhedrin gene. In some embodiments, one or more genes encoding a protein described herein are inserted into the baculoviral genome by homologous recombination between the baculoviral genome and the transfer vector. In some embodiments, the baculoviral genome is linearized at one or more unique sites. In some embodiments, the linearized sites are located near the target site for insertion of genes encoding the proteins described herein into the baculoviral genome. In some embodiments, a linearized baculoviral genome missing a fragment of the baculoviral genome downstream from a gene, e.g., polyhedrin gene, can be used for homologous recombination. In some embodiments, the baculoviral genome and transfer vector are co-transfected into insect cells. In some embodiments, the method of producing the recombinant baculovirus comprises the steps of preparing the baculoviral genome for performing homologous recombination with a transfer vector containing the genes encoding one or more protein described herein and co-transfected the transfer vector and the baculoviral genome DNA into insect cells. In some embodiments, the baculoviral genome comprises a region homologous to a region of the transfer vector. These homologous regions may enhance the probability of recombination between the baculoviral genome and the transfer vector. In some embodiments, the homology region in the transfer vector is located upstream or downstream of the promoter. In some embodiments, to induce homologous recombination, the baculoviral genome, and transfer vector are mixed at a weight ratio of about 1:1 to 10:1.

**[0288]** In some embodiments, a recombinant baculovirus is generated by a method comprising site-specific transposition with Tn7, e.g., whereby the genes encoding the proteins described herein are inserted into bacmid DNA, e.g., propagated in bacteria, e.g., *E. coli* (e.g., DH 10Bac cells). In some embodiments, the genes encoding the proteins described herein are cloned into a pFASTBAC® vector and transformed into competent cells, e.g., DH10BAC® competent cells, containing the bacmid DNA with a mini-attTn7 target site. In some embodiments, the baculovirus expression vector, e.g., pFASTBAC® vector, may have a promoter, e.g., a dual promoter (e.g., polyhedrin promoter, p10 promoter). Commercially available pFASTBAC® donor plasmids include: pFASTBAC 1, pFASTBAC HT, and pFASTBAC DUAL. In some embodiments, recombi-

nant bacmid DNA containing-colonies are identified and bacmid DNA is isolated to transfect insect cells.

[0289] In some embodiments, a baculoviral vector is introduced into an insect cell together with a helper nucleic acid. The introduction may be concurrent or sequential. In some embodiments, the helper nucleic acid provides one or more baculoviral proteins, e.g., to promote packaging of the baculoviral vector. In some embodiments, recombinant baculovirus produced in insect cells (e.g., by homologous recombination) is expanded and used to infect insect cells (e.g., in the mid-logarithmic growth phase) for recombinant protein expression. In some embodiments, recombinant bacmid DNA produced by site-specific transposition in bacteria, e.g., *E. coli*, is used to transfect insect cells with a transfection agent, e.g., Cellfectin® II. Additional information on baculovirus expression systems is discussed in U.S. patent application Ser. Nos. 14/447,341, 14/277,892, and 12/278,916, which are hereby incorporated by reference.

#### Insect Cell Systems

[0290] The proteins described herein may be expressed in insect cells infected or transfected with recombinant baculovirus or bacmid DNA, e.g., as described above. In some embodiments, insect cells include: the Sf9 and Sf21 cells derived from *Spodoptera frugiperda* and the Tn-368 and High Five™ BTI-TN-5B1-4 cells (also referred to as Hi5 cells) derived from *Trichoplusia ni*. In some embodiments, insect cell lines Sf21 and Sf9, derived from the ovaries of the pupal fall army worm *Spodoptera frugiperda*, can be used for the expression of recombinant proteins using the baculovirus expression system. In some embodiments, Sf21 and Sf9 insect cells may be cultured in commercially available serum-supplemented or serum-free media. Suitable media for culturing insect cells include: Grace's Supplemented (TNM-FH), IPL-41, TC-100, Schneider's *Drosophila*, SF-900 II SFM, and EXPRESS-FIVE™ SFM. In some embodiments, some serum-free media formulations utilize a phosphate buffer system to maintain a culture pH in the range of 6.0-6.4 (Licari et al. Insect cell hosts for baculovirus expression vectors contain endogenous exoglycosidase activity. Biotechnology Progress 9: 146-152 (1993) and Drugmand et al. Insect cells as factories for biomanufacturing. Biotechnology Advances 30:1140-1157 (2012)) for both cultivation and recombinant protein production. In some embodiments, a pH of 6.0-6.8 for cultivating various insect cell lines may be used. In some embodiments, insect cells are cultivated in suspension or as a monolayer at a temperature between 25° to 30° C. with aeration. Additional information on insect cells is discussed, for example, in U.S. patent application Ser. Nos. 14/564,512 and 14/775,154, each of which is hereby incorporated by reference.

#### Mammalian Cell Systems

[0291] In some embodiments, the proteins described herein may be expressed in vitro in animal cell lines infected or transfected with a vector encoding the protein, e.g., as described herein. Animal cell lines envisaged in the context of the present disclosure include porcine cell lines, e.g., immortalised porcine cell lines such as, but not limited to the porcine kidney epithelial cell lines PK-15 and SK, the monomyeloid cell line 3D4/31 and the testicular cell line ST. Also, other mammalian cells lines are included, such as CHO cells (Chinese hamster ovaries), MARC-145, MDBK,

RK-13, EEL. Additionally or alternatively, particular embodiments of the methods of the invention make use of an animal cell line which is an epithelial cell line, i.e. a cell line of cells of epithelial lineage. Cell lines suitable for expressing the proteins described herein include, but are not limited to cell lines of human or primate origin, such as human or primate kidney carcinoma cell lines.

#### Effectors

[0292] The compositions and methods described herein can be used to produce a genetic element of an anellovector comprising a sequence encoding an effector (e.g., an exogenous effector or an endogenous effector), e.g., as described herein. In some embodiments, the genetic element is the effector, e.g., the genetic element is a functional RNA. The effector may be, in some instances, an endogenous effector or an exogenous effector. In some embodiments, the effector is a therapeutic effector. In some embodiments, the effector comprises a polypeptide (e.g., a therapeutic polypeptide or peptide, e.g., as described herein). In some embodiments, the effector comprises a non-coding RNA (e.g., an miRNA, siRNA, shRNA, mRNA, lncRNA, RNA, DNA, antisense RNA, or gRNA). In some embodiments, the effector comprises a regulatory nucleic acid, e.g., as described herein.

#### In Vitro Assembly Methods

[0293] An anellovector may be produced, e.g., by in vitro assembly. In some embodiments, the genetic element is contacted to ORF1 in vitro under conditions that allow for assembly.

[0294] In some embodiments, baculovirus constructs are used to produce Anellovirus proteins. These proteins may then be used, e.g., for in vitro assembly to encapsidate a genetic element, e.g., a genetic element comprising RNA. In some embodiments, a polynucleotide encoding one or more Anellovirus protein is fused to a promoter for expression in a host cell, e.g., an insect or animal cell. In some embodiments, the polynucleotide is cloned into a baculovirus expression system. In some embodiments, a host cell, e.g., an insect cell is infected with the baculovirus expression system and incubated for a period of time. In some embodiments, an infected cell is incubated for about 1, 2, 3, 4, 5, 10, 15, or 20 days. In some embodiments, an infected cell is lysed to recover the Anellovirus protein.

[0295] In some embodiments, an isolated Anellovirus protein is purified. In some embodiments, an Anellovirus protein is purified using purification techniques including but not limited to chelating purification, heparin purification, gradient sedimentation purification and/or SEC purification. In some embodiments, a purified Anellovirus protein is mixed with a genetic element to encapsidate the genetic element, e.g., a genetic element comprising RNA. In some embodiments, a genetic element is encapsidated using an ORF1 protein, ORF2 protein, or modified version thereof. In some embodiments two nucleic acids are encapsidated. For instance, the first nucleic acid may be an mRNA e.g., chemically modified mRNA, and the second nucleic acid may be DNA.

[0296] In some embodiments, DNA encoding Anellovirus (AV) ORF1 (e.g., wildtype ORF1 protein, ORF1 proteins harboring mutations, e.g., to improve assembly efficiency, yield or stability, chimeric ORF1 protein, or fragments thereof) are expressed in insect cell lines (e.g., Sf9 and/or

HighFive), animal cell lines (e.g., chicken cell lines (MDCC)), bacterial cells (e.g., *E. coli*) and/or mammalian cell lines (e.g., 293expi and/or MOLT4). In some embodiments, DNA encoding AV ORF1 may be untagged. In some embodiments, DNA encoding AV ORF1 may contain tags fused N-terminally and/or C-terminally. In some embodiments, DNA encoding AV ORF1 may harbor mutations, insertions or deletions within the ORF1 protein to introduce a tag, e.g., to aid in purification and/or identity determination, e.g., through immunostaining assays (including but not limited to ELISA or Western Blot). In some embodiments, DNA encoding AV ORF1 may be expressed alone or in combination with any number of helper proteins. In some embodiments, DNA encoding AV ORF1 is expressed in combination with AV ORF2 and/or ORF3 proteins.

[0297] In some embodiments, ORF1 proteins harboring mutations to improve assembly efficiency may include, but are not limited to, ORF1 proteins that harbor mutations introduced into the N-terminal Arginine Arm (ARG arm) to alter the pI of the ARG arm permitting pH sensitive nucleic acid binding to trigger particle assembly (SEQ ID 3-5). In some embodiments, ORF1 proteins harboring mutations that improve stability may include mutations to an interprotomer contacting beta strands F and G of the canonical jellyroll beta-barrel to alter hydrophobic state of the protomer surface and improve thermodynamic favorability of capsid formation.

[0298] In some embodiments, chimeric ORF1 proteins may include, but are not limited to, ORF1 proteins which have a portion or portions of their sequence replaced with comparable portions from another capsid protein, e.g., Beak and Feather Disease Virus (BFDV) capsid protein, or Hepatitis E capsid protein, e.g., ARG arm or F and G beta strands of Ring 9 ORF1 replaced with the comparable components from BFDV capsid protein. In some embodiments, chimeric ORF1 proteins may also include ORF1 proteins which have a portion or portions of their sequence replaced with comparable portions of another AV ORF1 protein (e.g., jellyroll fragments or the C-terminal portion of Ring 2 ORF1 replaced with comparable portions of Ring 9 ORF1).

[0299] In some embodiments, the present disclosure describes a method of making an anellovector, the method comprising: (a) providing a mixture comprising: (i) a genetic element comprising RNA, and (ii) an ORF1 molecule; and (b) incubating the mixture under conditions suitable for enclosing the genetic element within a proteinaceous exterior comprising the ORF1 molecule, thereby making an anellovector; optionally wherein the mixture is not comprised in a cell. In some embodiments, the method further comprises, prior to the providing of (a), expressing the ORF1 molecule, e.g., in a host cell (e.g., an insect cell or a mammalian cell). In some embodiments, the expressing comprises incubating a host cell (e.g., an insect cell or a mammalian cell) comprising a nucleic acid molecule (e.g., a baculovirus expression vector) encoding the ORF1 molecule under conditions suitable for producing the ORF1 molecule. In some embodiments, the method further comprises, prior to the providing of (a), purifying the ORF1 molecule expressed by the host cell. In some embodiments, the method is performed in a cell-free system. In some embodiments, the present disclosure describes a method of manufacturing an anellovector composition, comprising: (a) providing a plurality of anellovectors or compositions according to any of the preceding embodiments; (b) option-

ally evaluating the plurality for one or more of: a contaminant described herein, an optical density measurement (e.g., OD 260), particle number (e.g., by HPLC), infectivity (e.g., particle:infectious unit ratio, e.g., as determined by fluorescence and/or ELISA); and (c) formulating the plurality of anellovectors, e.g., as a pharmaceutical composition suitable for administration to a subject, e.g., if one or more of the parameters of (b) meet a specified threshold.

#### Enrichment and Purification

[0300] Harvested anellovectors can be purified and/or enriched, e.g., to produce an anellovector preparation. In some embodiments, the harvested anellovectors are isolated from other constituents or contaminants present in the harvest solution, e.g., using methods known in the art for purifying viral particles (e.g., purification by sedimentation, chromatography, and/or ultrafiltration). In some embodiments, the harvested anellovectors are purified by affinity purification (e.g., heparin affinity purification). In some embodiments, the harvested anellovectors are purified by size exclusion chromatography (e.g., using a Tris buffer mobile phase). In some embodiments, the harvested anellovectors are purified by anion exchange chromatography (e.g., Mustang Q membrane chromatography). In some embodiments, the harvested anellovectors are purified by mixed mode chromatography (e.g., using a mixed mode resin, e.g., a Cato700 resin). In some embodiments, the purification steps comprise removing one or more of serum, host cell DNA, host cell proteins, particles lacking the genetic element, and/or phenol red from the preparation. In some embodiments, the harvested anellovectors are enriched relative to other constituents or contaminants present in the harvest solution, e.g., using methods known in the art for enriching viral particles.

[0301] In some embodiments, the resultant preparation or a pharmaceutical composition comprising the preparation will be stable over an acceptable period of time and temperature, and/or be compatible with the desired route of administration and/or any devices this route of administration will require, e.g., needles or syringes.

#### II. Anellovectors

[0302] In some aspects, the disclosure provides compositions and methods of using and making an anellovector, anellovector preparations, and therapeutic compositions. In some embodiments, the anellovector comprises one or more nucleic acids or polypeptides comprising a sequence, structure, and/or function that is based on an Anellovirus (e.g., an Anellovirus as described herein), or fragments or portions thereof, or other substantially non-pathogenic virus, e.g., a symbiotic virus, commensal virus, native virus. In some embodiments, an Anellovirus-based anellovector comprises at least one element exogenous to that Anellovirus, e.g., an exogenous effector or a nucleic acid sequence encoding an exogenous effector disposed within a genetic element of the anellovector. In some embodiments, an Anellovirus-based anellovector comprises at least one element heterologous to another element from that Anellovirus, e.g., an effector-encoding nucleic acid sequence that is heterologous to another linked nucleic acid sequence, such as a promoter element. In some embodiments, an anellovector comprises a genetic element (e.g., circular DNA, e.g., single stranded DNA), which comprise at least one element that is heter-

ologous relative to the remainder of the genetic element and/or the proteinaceous exterior (e.g., an exogenous element encoding an effector, e.g., as described herein). An anellovector may be a delivery vehicle (e.g., a substantially non-pathogenic delivery vehicle) for a payload into a host, e.g., a human. In some embodiments, the anellovector is capable of replicating in a eukaryotic cell, e.g., a mammalian cell, e.g., a human cell. In some embodiments, the anellovector is substantially non-pathogenic and/or substantially non-integrating in the mammalian (e.g., human) cell. In some embodiments, the anellovector is substantially non-immunogenic in a mammal, e.g., a human. In some embodiments, the anellovector is replication-deficient. In some embodiments, the anellovector is replication-competent.

[0303] In some embodiments the anellovector comprises a curon, or a component thereof (e.g., a genetic element, e.g., comprising a sequence encoding an effector, and/or a proteinaceous exterior), e.g., as described in PCT Application No. PCT/US2018/037379, which is incorporated herein by reference in its entirety. In some embodiments the anellovector comprises an anellovector, or a component thereof (e.g., a genetic element, e.g., comprising a sequence encoding an effector, and/or a proteinaceous exterior), e.g., as described in PCT Application No. PCT/US19/65995, which is incorporated herein by reference in its entirety.

[0304] In an aspect, the invention includes an anellovector comprising (i) a genetic element comprising a promoter element, a sequence encoding an effector, (e.g., an endogenous effector or an exogenous effector, e.g., a payload), and a protein binding sequence (e.g., an exterior protein binding sequence, e.g., a packaging signal), wherein the genetic element is a single-stranded DNA, and has one or both of the following properties: is circular and/or integrates into the genome of a eukaryotic cell at a frequency of less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5%, or 2% of the genetic element that enters the cell; and (ii) a proteinaceous exterior; wherein the genetic element is enclosed within the proteinaceous exterior; and wherein the anellovector is capable of delivering the genetic element into a eukaryotic cell.

[0305] In some embodiments of the anellovector described herein, the genetic element integrates at a frequency of less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5%, or 2% of the genetic element that enters a cell. In some embodiments, less than about 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, or 5% of the genetic elements from a plurality of the anellovectors administered to a subject will integrate into the genome of one or more host cells in the subject. In some embodiments, the genetic elements of a population of anellovectors, e.g., as described herein, integrate into the genome of a host cell at a frequency less than that of a comparable population of AAV viruses, e.g., at about a 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, or more lower frequency than the comparable population of AAV viruses.

[0306] In an aspect, the invention includes an anellovector comprising: (i) a genetic element comprising a promoter element and a sequence encoding an effector (e.g., an endogenous effector or an exogenous effector, e.g., a payload), and a protein binding sequence (e.g., an exterior protein binding sequence), wherein the genetic element has at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to a wild-type Anellovirus sequence (e.g., a wild-type Torque

Teno virus (TTV), Torque Teno mini virus (TTMV), or TTMDV sequence, e.g., a wild-type Anellovirus sequence as described herein); and (ii) a proteinaceous exterior; wherein the genetic element is enclosed within the proteinaceous exterior; and wherein the anellovector is capable of delivering the genetic element into a eukaryotic cell.

[0307] In one aspect, the invention includes an anellovector comprising:

[0308] a) a genetic element comprising (i) a sequence encoding an exterior protein (e.g., a non-pathogenic exterior protein), (ii) an exterior protein binding sequence that binds the genetic element to the non-pathogenic exterior protein, and (iii) a sequence encoding an effector (e.g., an endogenous or exogenous effector); and

[0309] b) a proteinaceous exterior that is associated with, e.g., envelops or encloses, the genetic element.

[0310] In some embodiments, the anellovector includes sequences or expression products from (or having >70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 100% homology to) a non-enveloped, circular, single-stranded DNA virus. Animal circular single-stranded DNA viruses generally refer to a subgroup of single strand DNA (ssDNA) viruses, which infect eukaryotic non-plant hosts, and have a circular genome. Thus, animal circular ssDNA viruses are distinguishable from ssDNA viruses that infect prokaryotes (i.e. Microviridae and Inoviridae) and from ssDNA viruses that infect plants (i.e. Geminiviridae and Nanoviridae). They are also distinguishable from linear ssDNA viruses that infect non-plant eukaryotes (i.e. Parvoviridae).

[0311] In some embodiments, the anellovector modulates a host cellular function, e.g., transiently or long term. In certain embodiments, the cellular function is stably altered, such as a modulation that persists for at least about 1 hr to about 30 days, or at least about 2 hrs, 6 hrs, 12 hrs, 18 hrs, 24 hrs, 2 days, 3, days, 4 days, 5 days, 6 days, 7 days, 8 days, 9 days, 10 days, 11 days, 12 days, 13 days, 14 days, 15 days, 16 days, 17 days, 18 days, 19 days, 20 days, 21 days, 22 days, 23 days, 24 days, 25 days, 26 days, 27 days, 28 days, 29 days, 30 days, 60 days, or longer or any time therebetween. In certain embodiments, the cellular function is transiently altered, e.g., such as a modulation that persists for no more than about 30 mins to about 7 days, or no more than about 1 hr, 2 hrs, 3 hrs, 4 hrs, 5 hrs, 6 hrs, 7 hrs, 8 hrs, 9 hrs, 10 hrs, 11 hrs, 12 hrs, 13 hrs, 14 hrs, 15 hrs, 16 hrs, 17 hrs, 18 hrs, 19 hrs, 20 hrs, 21 hrs, 22 hrs, 24 hrs, 36 hrs, 48 hrs, 60 hrs, 72 hrs, 4 days, 5 days, 6 days, 7 days, or any time therebetween.

[0312] In some embodiments, the genetic element comprises a promoter element. In embodiments, the promoter element is selected from an RNA polymerase II-dependent promoter, an RNA polymerase III-dependent promoter, a PGK promoter, a CMV promoter, an EF-1 $\alpha$  promoter, an SV40 promoter, a CAGG promoter, or a UBC promoter, TTV viral promoters, Tissue specific, U6 (polIII), minimal CMV promoter with upstream DNA binding sites for activator proteins (TetR-VP16, Gal4-VP16, dCas9-VP16, etc). In embodiments, the promoter element comprises a TATA box. In embodiments, the promoter element is endogenous to a wild-type Anellovirus, e.g., as described herein.

[0313] In some embodiments, the genetic element comprises one or more of the following characteristics: single-stranded, circular, negative strand, and/or DNA. In embodiments, the genetic element comprises an episome. In some

embodiments, the portions of the genetic element excluding the effector have a combined size of about 2.5-5 kb (e.g., about 2.8-4 kb, about 2.8-3.2 kb, about 3.6-3.9 kb, or about 2.8-2.9 kb), less than about 5 kb (e.g., less than about 2.9 kb, 3.2 kb, 3.6 kb, 3.9 kb, or 4 kb), or at least 100 nucleotides (e.g., at least 1 kb).

**[0314]** In some embodiments, an anellovector, or the genetic element comprised in the anellovector, is introduced into a cell (e.g., a human cell). In some embodiments, the effector (e.g., an RNA, e.g., an miRNA), e.g., encoded by the genetic element of an anellovector, is expressed in a cell (e.g., a human cell), e.g., once the anellovector or the genetic element has been introduced into the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, into a cell modulates (e.g., increases or decreases) the level of a target molecule (e.g., a target nucleic acid, e.g., RNA, or a target polypeptide) in the cell, e.g., by altering the expression level of the target molecule by the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, decreases level of interferon produced by the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, into a cell modulates (e.g., increases or decreases) a function of the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, into a cell modulates (e.g., increases or decreases) the viability of the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, into a cell decreases viability of a cell (e.g., a cancer cell).

**[0315]** In some embodiments, an anellovector (e.g., a synthetic anellovector) described herein induces an antibody prevalence of less than 70% (e.g., less than about 60%, 50%, 40%, 30%, 20%, or 10% antibody prevalence). In embodiments, antibody prevalence is determined according to methods known in the art. In embodiments, antibody prevalence is determined by detecting antibodies against an Anellovirus (e.g., as described herein), or an anellovector based thereon, in a biological sample, e.g., according to the anti-TTV antibody detection method described in Tsuda et al. (1999; *J. Virol. Methods* 77: 199-206; incorporated herein by reference) and/or the method for determining anti-TTV IgG seroprevalence described in Kakkola et al. (2008; *Virology* 382: 182-189; incorporated herein by reference). Antibodies against an Anellovirus or an anellovector based thereon can also be detected by methods in the art for detecting anti-viral antibodies, e.g., methods of detecting anti-AAV antibodies, e.g., as described in Calcedo et al. (2013; *Front. Immunol.* 4(341): 1-7; incorporated herein by reference).

**[0316]** In some embodiments, a replication deficient, replication defective, or replication incompetent genetic element does not encode all of the necessary machinery or components required for replication of the genetic element. In some embodiments, a replication defective genetic element does not encode a replication factor. In some embodiments, a replication defective genetic element does not encode one or more ORFs (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3, e.g., as described herein). In some embodiments, the machinery or components not encoded by the genetic element may be provided in trans (e.g., encoded in a nucleic acid comprised by the host cell, e.g., integrated into the genome of the host cell),

e.g., such that the genetic element can undergo replication in the presence of the machinery or components provided in trans.

**[0317]** In some embodiments, a packaging deficient, packaging defective, or packaging incompetent genetic element cannot be packaged into a proteinaceous exterior (e.g., wherein the proteinaceous exterior comprises a capsid or a portion thereof, e.g., comprising a polypeptide encoded by an ORF1 nucleic acid, e.g., as described herein). In some embodiments, a packaging deficient genetic element is packaged into a proteinaceous exterior at an efficiency less than 10% (e.g., less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.01%, or 0.001%) compared to a wild-type Anellovirus (e.g., as described herein). In some embodiments, the packaging defective genetic element cannot be packaged into a proteinaceous exterior even in the presence of factors (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3) that would permit packaging of the genetic element of a wild-type Anellovirus (e.g., as described herein). In some embodiments, a packaging deficient genetic element is packaged into a proteinaceous exterior at an efficiency less than 10% (e.g., less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.01%, or 0.001%) compared to a wild-type Anellovirus (e.g., as described herein), even in the presence of factors (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3) that would permit packaging of the genetic element of a wild-type Anellovirus (e.g., as described herein).

**[0318]** In some embodiments, a packaging competent genetic element can be packaged into a proteinaceous exterior (e.g., wherein the proteinaceous exterior comprises a capsid or a portion thereof, e.g., comprising a polypeptide encoded by an ORF1 nucleic acid, e.g., as described herein). In some embodiments, a packaging competent genetic element is packaged into a proteinaceous exterior at an efficiency of at least 20% (e.g., at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100%, or higher) compared to a wild-type Anellovirus (e.g., as described herein). In some embodiments, the packaging competent genetic element can be packaged into a proteinaceous exterior in the presence of factors (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3) that would permit packaging of the genetic element of a wild-type Anellovirus (e.g., as described herein). In some embodiments, a packaging competent genetic element is packaged into a proteinaceous exterior at an efficiency of at least 20% (e.g., at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100%, or higher) compared to a wild-type Anellovirus (e.g., as described herein) in the presence of factors (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3) that would permit packaging of the genetic element of a wild-type Anellovirus (e.g., as described herein).

#### Anelloviruses

**[0319]** In some embodiments, an anellovector, e.g., as described herein, comprises sequences or expression products derived from an Anellovirus. In some embodiments, an anellovector includes one or more sequences or expression products that are exogenous relative to the Anellovirus. In some embodiments, an anellovector includes one or more sequences or expression products that are endogenous relative to the Anellovirus. In some embodiments, an anellovector includes one or more sequences or expression products

that are heterologous relative to one or more other sequences or expression products in the anellovector. Anelloviruses generally have single-stranded circular DNA genomes with negative polarity.

[0320] In some embodiments, the genetic element comprises a nucleotide sequence encoding an amino acid sequence or a functional fragment thereof or a sequence having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of the amino acid sequences described herein, e.g., an Anellovirus amino acid sequence.

[0321] In some embodiments, an anellovector as described herein comprises one or more nucleic acid molecules (e.g., a genetic element as described herein) comprising a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus sequence, e.g., as described herein, or a fragment thereof.

[0322] In some embodiments, an anellovector as described herein comprises one or more nucleic acid molecules (e.g., a genetic element as described herein) comprising a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more of a TATA box, cap site, initiator element, transcriptional start site, 5' UTR conserved domain, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, three open-reading frame region, poly(A) signal, GC-rich region, or any combination thereof, of an Anellovirus, e.g., as described herein. In some embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein, e.g., an ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3 sequence of any of the Anelloviruses described herein. In embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF1 protein (or a splice variant or functional fragment thereof) or a polypeptide encoded by an Anellovirus ORF1 nucleic acid.

[0323] In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1 nucleic acid sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1/1 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1/2 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2/2 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2/3 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2/3

nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2t/3 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus TATA box nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus initiator element nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus transcriptional start site nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus 5' UTR conserved domain nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus three open-reading frame region nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus poly(A) signal nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus GC-rich nucleotide sequence of Table A1.

[0324] In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1 nucleic acid sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1/1 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1/2 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2/2 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2/3 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2/3



nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus transcriptional start site nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus 5' UTR conserved domain nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus three open-reading frame region nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus poly(A) signal nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus GC-rich nucleotide sequence of Table D1.

[0327] In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1 nucleic acid sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1/1 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF1/2 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2/2 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus ORF2/3 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus TAIP nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus TATA box nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus initiator element nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%,

96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus transcriptional start site nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus 5' UTR conserved domain nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus three open-reading frame region nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus poly(A) signal nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Anellovirus GC-rich nucleotide sequence of Table E1.

[0328] In some embodiments, the genetic element comprises a nucleotide sequence encoding an amino acid sequence or a functional fragment thereof or a sequence having at least about 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of the amino acid sequences described herein, e.g., an Anellovirus amino acid sequence.

[0329] In some embodiments, an anellovector as described herein comprises one or more nucleic acid molecules (e.g., a genetic element as described herein) comprising a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus sequence, e.g., as described herein, or a fragment thereof. In embodiments, the anellovector comprises a nucleic acid sequence selected from a sequence as shown in any of Tables A1-M2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the anellovector comprises a polypeptide comprising a sequence as shown in any of Tables A2-M2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

[0330] In some embodiments, an anellovector as described herein comprises one or more nucleic acid molecules (e.g., a genetic element as described herein) comprising a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more of a TATA box, cap site, initiator element, transcriptional start site, 5' UTR conserved domain, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, three open-reading frame region, poly(A) signal, GC-rich region, or any combination thereof, of any of the Anelloviruses described herein (e.g., an Anellovirus sequence as annotated, or as encoded by a sequence listed, in any of Tables A-M). In some embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein, e.g., an ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3 sequence of any of the Anelloviruses described herein (e.g., an Anellovirus sequence as annotated, or as encoded by a sequence listed, in any of Tables A-M). In embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF1 or

ORF2 protein (e.g., an ORF1 or ORF2 amino acid sequence as shown in any of Tables A2-M2, or an ORF1 or ORF2 amino acid sequence encoded by a nucleic acid sequence as shown in any of Tables A1-M1). In embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF1 protein (e.g., an ORF1 amino acid sequence as shown in any of Tables A2-M2, or an ORF1 amino acid sequence encoded by a nucleic acid sequence as shown in any of Tables A1-M1).

[0331] In some embodiments, an anellovector as described herein is a chimeric anellovector. In some embodiments, a chimeric anellovector further comprises one or more elements, polypeptides, or nucleic acids from a virus other than an Anellovirus.

[0332] In embodiments, the chimeric anellovector comprises a plurality of polypeptides (e.g., Anellovirus ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3) comprising sequences from a plurality of different Anelloviruses (e.g., as described herein). For example, a chimeric anellovector may comprise an ORF1 molecule from one Anellovirus (e.g., a Ring1 ORF1 molecule, or an ORF1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto) and an ORF2 molecule from a different Anellovirus (e.g., a Ring2 ORF2 molecule, or an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto). In another example, a chimeric anellovector may comprise a first ORF1 molecule from one Anellovirus (e.g., a Ring1 ORF1 molecule, or an ORF1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto) and a second ORF1 molecule from a different Anellovirus (e.g., a Ring2 ORF1 molecule, or an ORF1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto).

[0333] In some embodiments, the anellovector comprises a chimeric polypeptide (e.g., Anellovirus ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3), e.g., comprising at least one portion from an Anellovirus (e.g., as described herein) and at least one portion from a different virus (e.g., as described herein).

[0334] In some embodiments, the anellovector comprises a chimeric polypeptide (e.g., Anellovirus ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3), e.g., comprising at least one portion from one Anellovirus (e.g., as described herein) and at least one portion from a different Anellovirus (e.g., as described herein). In embodiments, the anellovector comprises a chimeric ORF1 molecule comprising at least one portion of an ORF1 molecule from one Anellovirus (e.g., as described herein), or an ORF1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF1 molecule from a different Anellovirus (e.g., as described herein), or an ORF1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 jelly-roll domain from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an ORF1 amino acid subsequence (e.g., as described herein) from a different

Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 arginine-rich region from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an ORF1 amino acid subsequence (e.g., as described herein) from a different Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 hypervariable domain from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an ORF1 amino acid subsequence (e.g., as described herein) from a different Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 N22 domain from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an ORF1 amino acid subsequence (e.g., as described herein) from a different Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 C-terminal domain from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an ORF1 amino acid subsequence (e.g., as described herein) from a different Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF1/1 molecule comprising at least one portion of an ORF1/1 molecule from one Anellovirus (e.g., as described herein), or an ORF1/1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF1/1 molecule from a different Anellovirus (e.g., as described herein), or an ORF1/1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises an ORF1/2 molecule comprising at least one portion of an ORF1/2 molecule from one Anellovirus (e.g., as described herein), or an ORF1/2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF1/2 molecule from a different Anellovirus (e.g., as described herein), or an ORF1/2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF2 molecule comprising at least one portion of an ORF2 molecule from one Anellovirus (e.g., as described herein), or an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF2 molecule from a different Anellovirus (e.g., as described herein), or an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF2/2 molecule comprising at least one portion of an ORF2/2 molecule from one Anellovirus (e.g., as described herein), or an ORF2/2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto.

acid sequence identity thereto, and at least one portion of an ORF2/2 molecule from a different Anellovirus (e.g., as described herein), or an ORF2/2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF2/3 molecule comprising at least one portion of an ORF2/3 molecule from one Anellovirus (e.g., as described herein), or an ORF2/3 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF2/3 molecule from a different Anellovirus (e.g., as described herein), or an ORF2/3 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF2T/3 molecule comprising at least one portion of an ORF2T/3 molecule from one Anellovirus (e.g., as described herein), or an ORF2T/3 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF2T/3 molecule from a different Anellovirus (e.g., as described herein), or an ORF2T/3 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto.

[0335] Additional exemplary Anellovirus genomes, for which sequences or subsequences comprised therein can be utilized in the compositions and methods described herein are described, for example, in PCT Application Nos. PCT/US2018/037379 and PCT/US19/65995 (incorporated herein by reference in their entirety). In some embodiments, the exemplary Anellovirus sequences comprise a nucleic acid sequence as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17 of PCT/US19/65995, incorporated herein by reference. In some embodiments, the exemplary Anellovirus sequences comprise an amino acid sequence as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18 of PCT/US19/65995, incorporated herein by reference. In some embodiments, the exemplary Anellovirus sequences comprise an ORF1 molecule sequence, or a nucleic acid sequence encoding same, e.g., as listed in any of Tables 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10, or 37A-37C of PCT/US19/65995, incorporated herein by reference.

TABLE A1  
Exemplary Anellovirus nucleic acid sequence  
(*Alphatorquevirus*, Clade 3)

Name	Ring 1					
Genus/	Alphatorquevirus,					
Clade	Clade 3					
Accession Number	AJ620231.1					
Full Sequence: 3753 bp						
1	10	20	30	40	50	
TGCTACGTCACTAACCCACGTGTCCTCTACAGGCCAATCGCAGTCTATGT						
CGTGCACCTCCGGCATGGTCTACATAATTATAAATGTCGACTTC						
CGAATGGCTGAGTTTGCTGCCGTCCGGAGAGGCCACGGCAGGG						
GATCCGAACGTCCTGAGGGCGGGTGCCTGAGTTACACACCGAAG						
TCAAGGGCAATTGGGCTCAGGACTGGCGGGCTTGGGCAAGGCTT						
AAAAATGCACTTTCTCGAATAAGCAAAAGAAAAGGAAAGTGCCTACTGC						
TTTGCCTGCCAGCAGCTAACGAAACCAACTGCTATGAGCTCTGGAAA						
CCTCCGGTACACATGTCACGGGATCCAACCGCATGGTATGGTACGCTT						
TCACCGTGGCACGCTTCTTGTGGTTGTTGGAATCCTATACCTACA						
TACTGCACTTGTCAAACATGGCATCAACAGGCCGAGACCTTCT						
GGGCCACGGGAGTAGACCCAAACCCCCACATCCGTAGAGGCCAGGCTGC						
CCGGCCGCTCCGGAGGCCCTCACAGGTTGATTCGAGACCAAGCCCTGACAT						
<hr/>						
Annotations:						
Putative Domain						Base range
TATA Box						83-88
Cap Site						104-111
Transcriptional Start Site						111
5' UTR Conserved Domain						170-240
ORF2						336-719
ORF2/2						336-715; 2363-2789
ORF2/3						336-715; 2565-3015

TABLE A1 -continued

TABLE A1-continued

Exemplary Anellovirus nucleic acid sequence (Alphatorquevirus, Clade 3)	
ORF2t/3	336-388; 2565-3015
ORF1	599-2830
ORF1/1	599-715; 2363-2830
ORF1/2	599-715; 2565-2789
Three open-reading frame region	2551-2786
Poly(A) Signal	3011-3016
GC-rich region	3632-3753

TABLE A2

Exemplary Anellovirus amino acid sequences (Alphatorquevirus, Clade 3)	
Ring1 (Alphatorquevirus Clade 3)	
ORF2	MSFWKPPVHNVTGICQRMWYESFHRC HASFCGGNPILHITALAETYGHPT GPRPSGPGVDPNPHIRRARPAPAA PEPSQVDSRPALTWHGDGGSDGGAG GSGSGGPVADFFADDGLDQLVAALDD EE (SEQ ID NO: 17)
ORF2/2	MSFWKPPVHNVTGICQRMWYESFHRC HASFCGGNPILHITALAETYGHPT GPRPSGPGVDPNPHIRRARPAPAA PEPSQVDSRPALTWHGDGGSDGGAG GSGSGGPVADFFADDGLDQLVAALDD EELLKTTPASSPPMKYPPVPTSLSEY KSSTRGSDWRTTRSGHGTCACTHLA EQVLRECQNNKKLLTLYSQAQKSLG STSQNKKPKKKAIHSKENRDRGRP RKKARQKPSRKRAKRSPSNSSCSSS TKSSSSSDRESKSSSSS (SEQ ID NO: 18)
ORF2/3	MSFWKPPVHNVTGICQRMWYESFHRC HASFCGGNPILHITALAETYGHPT GPRPSGPGVDPNPHIRRARPAPAA PEPSQVDSRPALTWHGDGGSDGGAG GSGSGGPVADFFADDGLDQLVAALDD EEPKKASGRHPKTRNPRKLTFPTK RIETVGDGRKRDRSPLAREPRGPL PTAVAAAVPRAAQATGNQSPRLRAA HKDPTRGCKPMPTVGPQWLFPER KPAPAPSSGDWAMEFLAAKIFDRPV RSNLKDTPYYPVKNQYNVYFDLKFE E (SEQ ID NO: 19)
ORF2t/3	MSFWKPPVHNVTGICQRMWPKKASGR HPKTRNPRRLTFTPKRIETVGDRC RKRDRSPLAREPRGPLPTAVAAAVP RAAQATGNQSPRLRAAHKDPTRGPC KPMPTVGPQWLFPERKPAPAPSSG DWAMEFLAAKIFDRPVRSNLKDTPY YPVKNQYNVYFDLKFE (SEQ ID NO: 20)
ORF1	MAWGWWKRRRRWWFRKRWRTRGRLRR RWPRSARRRPRRRVRRRRWRRGR RKTRTYRRRRRFRRRGKAKLIIK WQPAVIKRCRIKGYIPLIISGNGTF ATNFTSHINDRIMKGPFGGHSTM FSLYILFEELRHMNFWTRSNDNLE LTRYLGASVKIYRHPDQDFIVIYNR RTPLGGNIYTAPSLSHPGNAILAKH ILVPSLQTRPKGRKAIRLRIAPPTL FTDKWYFQKDIADLTLFNIMAVEAD LRFPFCSPQTNTCISFQVLSSVYN NYLSINTFNNNDNSDSKLKEFLNKAF PTTGTGKTSNLNTPTEGCHSP

TABLE A2-continued

Exemplary Anellovirus amino acid sequences (Alphatorquevirus, Clade 3)	
Ring1 (Alphatorquevirus Clade 3)	
ORF2t/3	QLKKPNPQINKPLESQYFAPLDALW GDPITYYNDLNENKNSLNDIIEKLILK NMITYHAKLREPPNSYOGNKAFCHL TGIYSPPYLNQGRISPEIFGLYTEI IYNPYTDKGTKGNKVWMDPLTKENNI YKEGQSKCLLTDMPWLTLFGYTDW CKKDTNNDLPLNYLRLVICPYTFP KLYNEKVKDYGYIPIPSYKFGAGQMP DCGSNYIPFQFRAKWYPTVLHQQQVM EDISRSGPFAPKVEKPSTQLVMKYC FNFNWGGNPPIIBQIVKDPSFQPTYE IPGTGNIPRRIQVIDPRVLGPHYSF RSWDMRRHTFSRASI KRVSEQETS DLVFSGPKKPRVDIPKQETQEESSH SLQRESRPWETEESETEALSQESQ EVFPQQQLQQQQYQEQLKLRQGIKVL FEQLIRTQQGVHVNPCLR (SEQ ID NO: 21)
ORF1/1	MAWGWWKRRRRWWFRKRWRTRGRLRR RWPRSARRRPRRRRIVKDPDFQPTY EIPGTGNIPRRIQVIDPRVLGPHYS FRSWDMRRHTFSRASI KRVSEQET SDLVFSGPKKPRVDIPKQETQEESS HSLQRESRPWETEESETEALSQES QEVFPQQQLQQQQYQEQLKLRQGIKV LEQLIRTQQGVHVNPCLR (SEQ ID NO: 22)
ORF1/2	MAWGWWKRRRRWWFRKRWRTRGRLRR RWPRSARRRPRRRRAQKSLGSTSQN KKPKKKAHHSKENRDRGRPRKKAR QKPSRKRAKRSRPSNSSCSSSTKSS SSDRESKSSSSS (SEQ ID NO: 23)

TABLE B1

Exemplary Anellovirus nucleic acid sequence (Betatorquevirus)						
Name	Ring2					
Genus/Clade	Betatorquevirus					
Accession Number	JX134045.1					
Full Sequence: 2797 bp	1	10	20	30	40	50
TAATAAATATTCAACAGAAAACCACCTAATTAAATTGCGCACCAAAA CCGTCACTTAGTCCTCTTGCACAAACTCTGCTTTTCCAACTGC CGGAAAACCACATAATTGCATGGCTAACACAAACTGATATGCTAATTA ACTTCCCACAAAACAACCTCCCTTTAAACACACCTACAAATTAACTTA TTAAACAGACTACCTCTGGAGGTACTACACACTATAATACCAAGTG CACTTCGAATGGCTGAGTTATGCCCTAGACGGAGAACGCATCAGTT CTGACTGGGACTGAACCTGGGGTGGAGGTAGACGCCCTTTCGCAGAAG AACTCAAGGGGAATTGGGCTAGTTAGCTAGCGGAACGGGCAAGAAA CTTAAACATTATTTTATTCAGATGAGCGACTGCTTAAACCAAATCTG TACAACACAAAACAACAACTACTCGGATTAATAACCTGCATTAAAC CCACGACCTGATCTGCTCTGCCAACACCAACTGACACTTATTACTAG CTTCTAGCAGAACACAACAAATTGAGATGTCATAACAAGAAAAAGAA AAATAACAAGATGCTTAACTACAGAAGAACGGTACAACACTACAGA CGTCTAGATGGTATGGACGAGGTGGATTAGACGCCCTTTCGCAGAAG ATTTCAAGAAAAGAAGGGTAAGACCTACTTATACTACTATTCTCTAA AGCAATGGCAACCGCATATAAAAGAACATGCTATAAAAGAACAGAC TGTTTAATACATACGCAACTTAAGACTGGGAATGAATAGTACAATGTA TGAAAAAGTATTGTAACATGGCTAGTATGGCCGGGGAGGGGTTCTTCTG TAAGCATGTTAATTTAGATGCTTGTATGATATACTAAACTTTGTAGA AACTGGGGACATCCACAAACCAAGACTTACCAACTAGTAAGATAAAGG ATGCAAATAACATTATCAAGACATTTACAGACTACATAGTAAGAA TACATACAGAACTACCACTAAAGCTAACAGTAACAAACTACACCAACACA CATCCACTAATGATGATGTCATAACAAACACATTACACCTAGTAG						

TABLE B1-continued

Exemplary Anellovirus nucleic acid sequence ( <i>Betatorquevirus</i> )	
ACAAACAAGAAGAAAAAGAACATACACAAAAAATTGTAAAACCAC	
CTCCGCAATTGAAAACAAATGGTACTTTGCTACAGACCTCTACAAAATT	
CCATTACTACAAATACACTGCACAGCATGCCACTTACAAAACCCATTGT	
AAAACCGACAAATTATCAAAACATGTATGGTCAGTGGATCAAGGACAATCATGGCA	
TAAGCATAACAAATAGAACATGTCAAGGACAATCATGGCA	
TTTAAATTAACAGGAAACACAAGCTTTTACCTTACACCGGAGC	
AAACCTACAGGAGCACAAACACATACAGTAGCAGACCTTACACAC	
TAACACAAACCCAAAGAATAACAGACCAAGAACATCAATAATGAGGAAAA	
ATTACAGACCATATTACTTCACAGAATACAAAAACAAATTACAAATTA	
TTGGGTAACCCATTATAACACATTCAGAACACCTAGATGATGATAC	
TATACTCACTAAAAGTCAGAGCAATAAAAAGCAATGGACAAACAGAA	
AACATGAATGGAAACCAATTAAACATGCAGGAACAAATGGCATTAACACC	
ATTAACGAGCCAATTACACAAATACAATATAACCCAGATAGAGACA	
CAGGAGAAAGACACTTACATTAACCTCTAACGCTACAGGAACAGGA	
TGGGACCCACAGGAATTCCAGAAATTACTAGAAGGATTCCACTATG	
GTTAATATTGGGATTTCAGACTTCACAAAACCTAAAAAAAGTAA	
CAAACATAGACACAAATTACATGTTAGTAGCAAAACAAAATTACAA	
AAACCTGGCACATTCTACTTAGTAACTAAATGACACCTTGTAGAAGG	
CAATAGGCCACAAACCTTACCTTGAAAGACAACATTAAATGGT	
ACCCACAAAGTCAATACCAATTAGAACACAAAACACTACAAACT	
GGGCATTTACACCAACATACAAGGACAACATCAGACAATATCAAT	
GTTTTATAAATTACTTTAAATGGGGAGGAAGGCCACAAAGCAATTAA	
ATGTTGAAAATCTGCCACAGATTCAATATCCCATAACCGTAACGAG	
CATGAAACACTTCGTTACAGTCAGGGGAAGGCCACAGATCCATTCTT	
ATACTCCTCGACTATAGACACCGGAACACTACACAACAGCTTGTCA	
GAATTAGCCAAGACTGGGCATTAAGACACTGTTCTAAATTACAGAG	
CCAGATCGACAGCAACTGCTCAACAAAGCCCTCGAATGCTGCAAATCTC	
GGAAAGAAACGCGAGAAAAAGAAAAGAGTACAGCAGCTCATCAGCA	
ACCTCAGACAGCAGCAGCTGTACAGAGAGCGAATAATATCATTATA	
AAGGACCAATAACTTTAACTGTGAAAAAGGTGAAATTGTTGATGAT	
AAACCAAAAACCGTAGATTACACCTGGAGATTGAAACTGAGTTACA	
ATAGCAAATGGTTAACAGAGCCCCAAGATCCTTGTAAATGATCCTC	
CCTTTTACCCATGGTACCCACCTGAACTGTGAAACTTTAAGCTTAAT	
TTTACTGAATAAGGCCAGCATTAATTCACTTAAGGAGTCTGTTATTAA	
AGTTAACCTTAATAAACGGTCACCGCTCCCTAATACGCAGGGCAGAA	
AGGGGGCTCGCCCCCTTAACCCCCAGGGGGCTCGCCCCCTGAAACCC	
CCAAGGGGCTACGCCCCCTAACCCCCC (SEQ ID NO: 54)	
<b>Annotations:</b>	
Putative Domain	Base range
TATA Box	237-243
Cap Site	260-267
Transcriptional Start Site	267
5' UTR Conserved Domain	323-393
ORF2	424-723
ORF2/2	424-719; 2274-2589
ORF2/3	424-719; 2449-2812
ORF1	612-2612
ORF1/1	612-719; 2274-2612
ORF1/2	612-719; 2449-2589
Three open-reading frame region	2441-2586
Poly(A) Signal	2808-2813
GC-rich region	2868-2929

TABLE B2

Exemplary Anellovirus amino acid sequences ( <i>Betatorquevirus</i> )	
Ring2 ( <i>Betatorquevirus</i> )	
ORF2	MSDCFKPTCYNNTKQTHWINNLHL THDLICFCPTPTRHLLLALAEQQET IEVSKQEKEKITRCLITTEEDGTT DVLGMDDEVGLDALFAEDFEEKEG (SEQ ID NO: 55)
ORF2/2	MSDCFKPTCYNNTKQTHWINNLHL THDLICFCPTPTRHLLLALAEQQET IEVSKQEKEKITRCLITTEEDGTT DVLGMDDEVGLDALFAEDFEEKEGF NIPYPVTSMKQLRYRVQGKPQNPSY

TABLE B2-continued

Exemplary Anellovirus amino acid sequences ( <i>Betatorquevirus</i> )	
Ring2 ( <i>Betatorquevirus</i> )	
ORF2/3	TPSTIDTGTQQQLCHELAKTGHLK TLFLKLQSQIDSNCNSNPKSNACKSR KKRRKKKKYSSSSATSDDSSSSCT ESE (SEQ ID NO: 56)
ORF1	MSDCFKPTCYNNTKQTHWINNLHL THDLICFCPTPTRHLLLALAEQQET IEVSKQEKEKITRCLITTEEDGTT DVLGMDDEVGLDALFAEDFEEKEGA RSTATAQTSPRMMPANLGRNAGEKRK RSTAHHQQPQTAAAAVQRANNIIIK GPITFNCKVKVLFDDPKNRRFTP EEFETELOQIAKWLKRPRPSFVNNDPP FYPWLPPEPVNVFKLNFT (SEQ ID NO: 57)
ORF1/1	MPYYYRRRRNYRRPRWYGRGWIRR PFRRRFRRKRRVPRPTTIPLKQWQ PPYKRTCYIKGQDCLIYYSNRLGM NSTMYEKSIVPVHWPGGGFSVSM TLDALYDIHKLCRNWWTSTNQDPL VRYKGCKITFYQSTFTDYIVRIHTE LPANSNLTYPNTHPLMMMSKYKH IIPSQRTRKKPYTKFVPKPPQF ENKWKYFATDLYKIPLLQIHCTACNL QNPVVKPDKLSNNVTLWSLNTISIQ NRNMSVDQGQSWPFKILGTQSFYFY FYTGANLPGD7TQIPVADLLPLTNP RINRPQSLNEAKITDHITFTEYKN KFTNYWGNPFNKHIQEHDLMILYSL KSPEAIKNEWTTEENMKWNQLNNAGT MALTPFNEPIFTQIQQYNPDRDTGED TOLYLLSNATGTGWDPPGPIELILE GFPLWLIYWGFAFDQKNLKVKVTNID TNYMLVAKTKFTQKPGTFYLVILND TFVEGNSPYEKQPLPEDNIKWPQV QYQLEAQNKLLQTGPFTPNIQGOLS DNIISMPYFKWGGSPPKAINVEN PAHQIQQPIPNEHETTSQSPGEA PESILYSDYRHGNYTTTALSRSIQ DWALKDTVSKITEPDRQQLLKQALE CLQISEETQEKEKEVQQLISNLRQ QQQLYRERIISLLKDQ (SEQ ID NO: 58)
ORF1/2	MPYYYRRRRNYRRPRWYGRGWIRR PFRRRFRRKRRVPRPTTIPLKQWQ LQSPGEAPESILYSFDYRHGNYTTT ALSRISQDWALKDTVSKITEPDRQ LLKQALECLQISEETQEKEKEVQQ LISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 59)
ORF1/2	MPYYYRRRRNYRRPRWYGRGWIRR PFRRRFRRKRRVPRPTTIPLKQWQ CKSRKKRKKKKYSSSSATSDDSS SSCTESE (SEQ ID NO: 60)

TABLE C1

Exemplary Anellovirus nucleic acid sequence ( <i>Gammatorquevirus</i> )						
Name	Ring4					
Genus/Clade	<i>Gammatorquevirus</i>					
Accession Number						
Full Sequence: 3176 bp	1	10	20	30	40	50
TAAAATGGCGGAGGCCAATCATTTACTTTCACTTCCAAATTAAAAAT						
GGCCACGTCAACAAACAGGGGGTGAGCCATTAACTATAACTAAGTG						
GGGTGGCGAATGGCTGAGTTACCCCGTAGACGGTGCAGGGACCGGATC						
GAGCGACGAGGGAGGTCCCGGCTGAGCCATTGGCGGGAGCGGAGGTGAG						
TGAAACCACCGGAGTCTAGGGCAATTGGGCTAGGGCAGTTAGCGGAA						
CGGGCAAGAACCTAAACATATTGGTTACAGATGGTAGTATATCC						
TCAAGTGATTTTAAGAAAACGAATTAAATGAGGAGACGCCAGAACCA						
AGTATGGATGTCATGAAATTGCTGACTCTCATGATAATCTGCAGTGTCT						
GTCATGCCATTGTCACCTTCTGCTTCCATTCTCTCTGGCCACAAA						
GATCGTGTATCTACTATTAAACCAATTCTCTAAGAGATATAAGAAAAA						
ATGCCATTCTGGGAGAAGAAGAGGAGAAAATTCTGGACCAACACAGGTT						
TAATTACACCAAAGAGAAGATATGAAAAGATGCCAGAACAGGCC						
GCAGAAGAACACATACAGACCCCTGTCGGCCCGTAGAAAATT						
CAGAAGGTTAAAGAAAAAAATTTTAAATGTTAGACAATGGCAACC						
AGACAGTATAAGAAACTTGTAATTAGGACAGTCAGCTATAGTTGTTG						
GGGCTGAAGGAAAGCAGGAAATGTACTGTTATACTGTCAATAAGTAAATTAT						
GTGCCCCCAAAACACCATATGGGGAGGCTTGGAGTAGACCAACTACAC						
ACTGAAATACTTATATGAAATACAGATTGCAACAAAATTGGACAC						
AATCTAATGTAATGAAAGACTTATGCGATACATAATGTTAAGCTATA						
TTCTACAGAGAACACAAAACAGACTTGTCTTCTATGACAGAACCC						
ACCTTTCAACTAACAAATTATCACACCAGGAGCACACCCACAAA						
TATGCTTCAAAACACCAAATTATCATACATACACAAATGACAAGGCT						
AATGGAAGACTACAAAAAAACTCAAATTAAACCTCTAACAAATGCT						
TTCTAAATGGTTCTTTCAAAACATTGTAATACCCCTTAACATCTC						
TTAAAGGTTCTCACTGAGACCTTACCTAGGACTCTACCTAGGCTGTTAT						
GAAAATCCACAGGATTTTTATTATTAACCATGGATACACAAAT						
AACAAACTGGGAGCACAATCTCACAGCATACAGACCTAACCTCAAGG						
TGACAGACACACATACTACAGATAACAAAATGACAGAAAAATTTAAC						
ATTAAAAGCCATGAAATCAGAAAAGTATATGAAAACGGTTATT						
TCAATCTAGTTCTACAAACACAGTGTACATATAACAGTGGAGCTGGT						
AAGCCTGTATAGCAGAAAACACCTAGGAATAGCTATTCACTCAGTA						
AAAGACAATGGGAGTGGTAATGATAACCTGTAGCCTAGCTAGCAA						
CACTTGGGACACGCCCTAAAGACAGCTGCTATTTTAAACAGGAGTAC						
CCATATGGCTTACGGCTTATTTGAGACTACTGTAGACAAATTAAA						
GCTGACAAAACATGGCTAGACAGTCTAGTAACTGCTGAGTCTGC						
TATTTTACTTACCCAATCCAGGAGCAGGCAATGGTATTGCACTAT						
CACAAAGTTTATAAAAGCAATGGTCGTTAACACCAACTACACTG						
CTACAAAAGGAAAGTGGTTCCACAAATACACCAACAGAAATT						
TAATAGCTTGTAGAATCAGGACATTGTCACAAATATGCAAATCAA						
CTGAAAGCACTGGGACTAAAATATAATGTTTACATTAAAGTGG						
GTGTCGGACACAAATTCATGAACTGGACAGGAAATTGCTGACCCCTAGCAAAGA						
GCAGTATGATGTCCTCGACTTCTCATCCAAACAAATACAAATGAGATC						
CAGAAGGACAAGACCCAGATCTCATCCATGATTGGGACTACAGACGA						
GGCTTTATAAGAAAGATCTTAAAGAAATGTCAACTTACTCTCAAC						
TCATACAGATCGCAACACTCAGAGGAAGACATTCCAAAAAGAAAAA						
AGAGAATTGGACCCAACTCACAGTCACAGGAAACAAAAGAGGAGACA						
CTGTCATGTCCTCTCTCTGCAAAAGAGATACCTTCAAGAACAGA						
GACACAAGAACCTCCAGCAGCTCATCAAGCAGCAGCAGGAGCAGC						
TCTCTCTCAAGGAAACATCTCCAGCTCATCCACAAACTAAAGAGAAT						
CAACAAATGCTTACGGCTTACAGGCAATGTTACCTTAACCAAGATTT						
CCTGGATTGGAGACAAACAGAGAGAATTAGCAATTATATTTCTAG						
CCCCCTCTAGAACCTACAAAGAGGACCTTCCATTCTATCCCTGGCTACAC						
CTGCACCCCTTGTACATTAACTCTAACCTTCAACTGCAAGGCTAGGCCAACAT						
GTACACTTAGTAAAGCATGTTTATTAAGCACAACCCCAAATAATG						
AAAAAATAAAAAAAAAAATTTGCAAAATTTCGGCGCT						
CGCGCGCATGTGCGCTCTGGCGAAATCACGCAACGCTCGCGGCCGC						
GTATGTCCTTACACGGCAGCTAGATTGGGTGCGCGCTAGCGCG						
CACCCCAATGGCCCCCGCCCTGTCGACCCGCTTGCAGGGTCGGAC						
ACTTCGGGCTCGGGGGCGCCCTGGGGCTTTTACTAAACAGACT						
CCGAGCCGCCATTGGCCCTTAAGCTCCGCCCTCATGAATATTCA						
AAAGGAAACCAATAATTAGAATTGCCGACCCAAACTGCCATATGCTAA						
TTAGTCCCCTTACAAAGTAAGGGGAAGTGAACATAGGCCACACC						
CGCAGGGCCAAGGCCCGCACCCCTACGTCACTAACACGCCGCCGC						
CATCTTGGGTGCGCAGGGCGGGGC (SEQ ID NO: 886)						
Annotations:						
Putative Domain		Base range				
TATA Box		87-93				

TABLE C1-continued

Exemplary Anellovirus nucleic acid sequence ( <i>Gammatorquevirus</i> )	
Cap Site	110-117
Transcriptional Start Site	117
5' UTR Conserved Domain	185-254
ORF2	286-660
ORF2/2	286-656; 1998-2442
ORF2/3	286-656; 2209-2641
TAIP	385-484
ORF1	501-2489
ORF1/1	501-656; 1998-2489
ORF1/2	501-656; 2209-2442
Three open-reading frame region	2209-2439
Poly(A) Signal	2672-2678
GC-rich region	3076-3176

Exemplary Anellovirus amino acid sequences ( <i>Gammatorquevirus</i> )	
Ring4	( <i>Gammatorquevirus</i> )
ORF2	MVSISSSDFKKTFNEETQNQVWM SQIADSHDNIICSCWHPFAHLLASIF PPGHKDRDLTINQILLRDKKEKCHS GGEEGENSGPTTGLITPKEEDIKD GPEGAAEDHTDALFAAAVENFER (SEQ ID NO: 887)
ORF2/2	MVSISSSDFKKTFNEETQNQVWM SQIADSHDNIICSCWHPFAHLLASIF PPGHKDRDLTINQILLRDKKEKCHS GGEEGENSGPTTGLITPKEEDIKD GPEGAAEDHTDALFAAAVENFESG VDHNSNMQKLLTLANKSSMMPSILS TKQYKLKIQDKTPDLSMIGTTDE ALLKKDLLKECQLTSQLIQISKQLQ RKTFPKRELDPNSQSHNKKRRH CHVSSLASAKKIPSKKQRHKKTSSSS SSSSRSSSSSRETSSSSSTN (SEQ ID NO: 888)
ORF2/3	MVSISSSDFKKTFNEETQNQVWM SQIADSHDNIICSCWHPFAHLLASIF PPGHKDRDLTINQILLRDKKEKCHS GGEEGENSGPTTGLITPKEEDIKD GPEGAAEDHTDALFAAAVENFERS ASNFRGRHSQKEKENWTPTHSPPTK RRGDTVMSPSLQKRYLPRNRDTRR PPAHQAAGAAAPPQEKHPPAHPQ TKRESTNASASHRHTLTRPKPGFE EQTERELAIIFHRPPRTYKEDLPFY PWLPPAPLVQFNLFNFKG (SEQ ID NO: 889)
TAIP	MRRRTKYGCLKLLTLMIIISAVAGI HLLTFLPYFLLATKIVILLLTKFF (SEQ ID NO: 890)
ORF1	MPPWWRRRKFWTNRNFNYTKRRRY RKRWPRRRRRPPYRFPVRRRRKRL RKVKRKKSILIVRQWQPDLSIRTCKI IQOSAIVVGAEGKQMYCYTVNKLIN VPPKTPYGGGFGVQYTLKYLVEEY RPAQNIWTQSNVLKDLCRYINVKLI FYRDNKTDVFLSYDRNPFQLTKFT YPGAHPQOIMLQKHHKFILSOMTKP NGRLTKKLKIKPPKQMLSKWFFSKQ FCKYPLLSLKAASALDLRHSYLGCCN ENPVQFYYLNHGYYTITNWGAQSS TAYRPNSKVTDYYRYKNDRKNIN IKSHEYEKSIISYENGYFQSSFLQTQ CIYTSERGEACIAEKPLGIAIYNPV

TABLE C2-continued

Exemplary Anellovirus amino acid  
sequences (*Gammatorquevirus*)  
Ring4 (*Gammatorquevirus*)

---

TABLE D1

**Exemplary Anellovirus nucleic acid sequence (Betatorquevirus)**

Name	Ring9					
Genus/Clade	Betatorquevirus					
Accession Number	MH649263.1					
Full Sequence: 2845 bp	1	10	20	30	40	50
TTATTAATATTCAACAGGAAACCACCTAATTAAAATTCCGGCACCACAAA						
CCGTCACTAATTCCTTATTAACATTACTTCCCTTTAACCATGAATA						
TTCATACAAACACATCACACTTCTGGGAGGAGCATAAAACTATATAACT						
AACATACACAGACAATGGCTGAGTTTATGCCGTAGACGGAGGCACACA						
GCTACTCTGGCACCTGAACCTTGGCGGGTCCGAGGTGAGTGTAAACCA						
CCGTAGTCAGGGCAATTGGCTAGTTAGTCAGTAGCGGAACGGGCAAG						
ATTATTAATACAAACATTATTTACAGATGAGAACAAACTAAACCAAC						
TTTATACAAAGACAAACATTGGAAATTCAATGGTCAAACACATTTTA						
GCTCTCACGACTGTGCTGGCGTGCACAGCTTACAGTTTACATTACTG						
ATTTAATTAACAAAACCGAGAACGCACCTAACACAGAAGAGACATTAA						
AAATATAAAATACTGGCTCTTACTGGCGCCAAAAAACTACCGAACAGAAGATA						
TAGACCTTTCTCTGGAGAACTAGAAGAATTATTCAGAACAGAAGAAGAT						
GGAGATACCGCAACAAAGAACACATCTGGAGAACAAACTCGGGGTA						
AGAAAACGTTTATAAAAGAAAGTTAAAAAAATGTACTAAACAGTT						
TCAACCAAAATATTAGAAGATGTACAATATTGGAAACATCTGCCCTAT						
TTCAAGGCTCTCCAGAACACAAATTATTCACAAACATCTAC						
TCTCTAGTACCAAGAGATAAGAACAGGCCAACAAATTATTCACAAACATCTAC						
TGAAAGCTTAAGTAGTTATGGGAAGACTGGAACATTAAAAATGTAT						
GGACTCAAGTAACGCTGGTTTACACTTGTAAGATACGGGGAGTAACAA						
TTATACATTTCATCAACTCTGCTTACACTGACTATATTGCTCAAGTTTCA						
CTGGTATCTCATGACAGACACAAAATCACACATGCGACTCAGCACCACAA						
ACAGAACATGTTATTAAAAACATGTAATAAGAGTACCTAGCAGAGAACAC						
CGCAAAAAAAAGAACAGGCCATACAAAAGACTTAGAGTAGGACCTCTCTCA						
AATGCAACAAACATGGTACTTCAAGAGACATATGTAATAACATTAA						
TAATGATTGCGACCCACAGCGTGTGACTTTAGATATCCTTCTTGTCAAGC						

TABLE D1-continued

Exemplary Anellovirus nucleic acid  
sequence (Betatorquevirus)

### Annotations:

Putative Domain	Base range
TATA Box	142-148
Initiation Element	162-177
Transcriptional Start Site	172
5' UTR Conserved Domain	226-296
ORF2	328-651
ORF2/2	328-647; 2121-2457
ORF2/3	328-647; 2296-2680
ORF1	510-2477
ORF1/1	510-647; 2121-2477
ORF1/2	510-647; 2296-2457
Three open-reading frame region	2296-2454
GC-rich region	2734-2845

TABLE D2

Exemplary Anellovirus amino acid  
sequences (Betatorquevirus)  
Ring9 (Betatorquevirus)

---

TABLE D2-continued

Exemplary Anellovirus amino acid sequences ( <i>Betatorquevirus</i> ) Ring9 ( <i>Betatorquevirus</i> )	
	KHHRGRTQKRTRKRKTKHKRSRSS STSSSESINSISSSSSSST (SEQ ID NO: 1003)
ORF2/3	MSQLKPTLYKDKSLELQLWNNIFS SHDLCCGCNDPVHLLLILINKTGEA PKPEEDIKNIKCLLTGAKNTTEEDI DLSPGELEELFKEEKTDGTANQEKH TGEENCFGQTQPTSTQTNITVDGLR NGLGRGKRPNTRDPDPAQQAQKAST ASQAAAQAVPETPKYRIVASNIKVE LFPTKKPFKNRRFTPSETERQCA KAFCRPERHFFYDPPPYPYCVPPEPI VNFalGYKI (SEQ ID NO: 1004)
ORF1	MPPYWRQKYRRRYRPFSWRTRRII QRKRKRWRYRKPRKTYWRRKLRLRKR FYKRKLKKIVLKQFQPKIIIRCTIF GTICLFLQGSUPERANNYIQTIYSYV PDKEPFGGGWTLITESLSSLWEDWE HLKNVWTQSNAGLPLVRYGGVTLYF YQSAYTDYIAQVFNCYPMTDKYTH ADSAAPNRMLLKKHVIRVPSRETRKK RKPYKRVVRVGPPSQMNWKWFQRFDI CEIPLIMIAATAVDFRYPFCASDCA SNNLTLCNLPLLFQNQDFDHPSDT QGYFPKPGVYLYSTORSNKPSSSDC IYLGNTKDNQEGKSASSLMLKTQK ITDWGNPFWHYIDGSKKIFSYFKP

TABLE D2-continued

Exemplary Anellovirus amino acid sequences ( <i>Betatorquevirus</i> ) Ring9 ( <i>Betatorquevirus</i> )	
	PSQLDSSDFEHMTELAEPMFIQVRY NPERDTQGNLIYVTENFRGQJHWDP PSSDNLLKLDGFPLYDMCWGFIDWIE KVHETENLLTNYCFCIRSSAFNEKK TVFIPVDSFLTGFSFYETPVKSSD QAHWHPQIRFQTKSINDICLTGPGC ARSPYGNYMQAKMSYKPHVKWGGCP KTYEKPYDPCSQPWNWTIPHNLNETI Q1QNPNTCPQTELQEWWDWRRIDVTK KAIERIQHTEPHETLQISTGSKHN PPVHRQTSPWTDSETDSEEKDQTO EIQIQLNKLRLKHQQHLKQQLKQYLK PQNIE (SEQ ID NO: 1005)
ORF 1/1	MPPYWRQKYRRRYRPFSWRTRRII QRKRKRWRYRKPRKTYWRRKLRLRPNWT IPHNLNETIQQNPNTCPQTELQEW DWRRDIVTKKAIERIQHTEPHETL QISTGSKHNPPVHRQTSPWTDSETD SEEKDQTOEQIQLNKLRLKHQQHL KQQLKQYLKPQNIE (SEQ ID NO: 1006)
ORF 1/2	MPPYWRQKYRRRYRPFSWRTRRII QRKRKRWRYRKPRKTYWRRKLRLRPNWT THQYTDKHHGRRTQKPRTRKRKTKH KRSRSSSTSSESINSISSSSSSST (SEQ ID NO: 1007)

TABLE E1

Exemplary Anellovirus nucleic acid sequence (*Betatorquevirus*)

Name	Ring 10
Genus/Clade	<i>Betatorquevirus</i>
Accession Number	JX134044.1
Full Sequence: 2912 bp	
1            10            20            30            40            50                                                                    TAATAAAATTCAACAGGAAAACCACCTAATTAAATTGCGCAGCACAAA CCGTCACTTAGTTCCCTTTCCACAACCTCCTTTACTAATGAATA TTCATGTAATTAAATAATCACCGTAATTGGGGAGGCCTTTAAA CTATAAAACTAACATACATCGAATGGCTGAGTTATGCCGCCAGACGG AGACGGGATCACTTCAAGTGACTCCAGGCTGATCAAGGGGGGGTCCCGAAG GTGAGTGAACCACCGTAGTCAGGGCAATTGGCTAGATCAGTCTGG CGGAACGGGAAGAAACTTAAATGTTACTTTATTACAGAAAATGTTCAA ATCTCCAACATACTTAACAACTAAAGGCAAAACATGCCCTAATCAACT GCTTCGTTGGAGACACGATCTTGTGAGCTGTAACAAATCCTGCCCTAC CATTGCCCTAAATACTTGCAACTACCTTAGCACCTCAACTAAAACAAGA AGAAAACAAACAAATAATCAATGCCCTGGTACAGAGCCCTAGCTA CAACCCGGAGACGAAGAAATTGGTTAGAAGACCTAGAAAAACTATT ACAGAGATAACAGAAGAAGACGCCCTGGTAAGAAGAAAACCTTTAC AAACGTAAAATTAGAGACTAAATATAGTAAATGGCAACCTAAATCAAT TAGAAAATGAGAATAAAAGGAATGCTATGTTCAACAGCAGAGAAG ACAGACTGTATATAACTTTGATATGTATGAAGAGTCTTATACAGAA AAACTGCCGGAGGGGGGATTAGCATTAAGAATATAAGCTTATATGC CTTATACCAAGAACACATACATGCACACAAACATATTACACACAAACA CAGACAGACCAACTAGCAAGATAACAGGCTGTTTTAAATTCTACCAA AGCAAAGACATAGACTACGTAGTAACATATTCTACATCACTCCACTAAG AAGCTCAATGGGAATGTACAACACTCCATGCAACCATCCATACATCTAATGC AACAAAACAAACTATTGTACCAAGCAACAAACACAAAAAGAGAAAA CCATATATTTAAACATATATCACCACAAACAAATGAAATCTCAATG GTACTTTCAACATAACATTGCAAACATACCGCTACTAATGATAAGAACCA CAGCATTAAACATTAGATAATTACTATATAGGAAGCAGACAATTAGTACA AATGTCACTATACACACTTAACACAAACATACATCCAAAACAGAGACTG GGGAGACAGAAATAAAACTACTACTGCCCCAAACATTAGGAACACAAAGAT	

TABLE E1 -continued

Exemplary Anellovirus nucleic acid sequence ( <i>Betatorquevirus</i> )	
ACTTCCTATATGGAACACATTCAACTGCACAAAATATTATGACATAAAG	
CTACAAGAACTAATACTTTAACAAACACACAAGACTATGTACAAGGCTT	
TGATTGGACAGAAAAAGACAAACATAACATAACAACCTAACAAAGATTCT	
TAACCTAACGGGACAGGAATCATTTCAGCGAGAATGGATAACAGCACAA	
AACCCAGTAATACACACAGAACAGTCCTACACAATAGAACAAATATA	
CACCGCTTCAACAACAAACATTCCAAAACAAAAACTAACAGACCTACCAA	
CGCCAGGATATATTATTTATACTCCAACAGTAAGCTTAAGATACACACCA	
TACAAAGACCTAGCAGAAAAGAACAAATGCTACTTGTAAAGAAGCAAAAT	
AAATGACACGGGTGGGACCCAGAACACACCAAGAATTAAACAGTG	
ACCTACCACAATGGTTACTATTATGGCTACCCAGACTACATAAAAAGA	
ACACACAAACTTGCATTAGTAGACACAAATTACATACTAGTAGACCACTG	
CCCATAACAAATCCAGAAAAAAACACCATTTACCTTAAGGCACATCAT	
TTATAGAAGGTAGAAGGCCATACAGTCCTCAGACACACATGAACCCAGAT	
GAAGAAGACCAAAACAGGTGGTACCCATGCTACCAATATCACAAAGAATC	
ATAAAATTCAATATGCTTAGGGTCCAGGCACACAAAAATACCAAAAG	
GAATAACAGCAGAAGCAAAGTAAAATATTCCCTTAATTAAAGTGGGGT	
GTTGACCTTACCCAATGTCTACAATTACAAACCCGACAGACCCAGC	
ATATGTTGTTCCAATACTTCATGAAACAACTTCGTTACAGAATCCAA	
CCACCAAGACCAAGAGCAGCTTCTGTACTCCCTTGACGAAAGGGGGACAA	
CTTACAGAAAAGCTACAAAAAGCTTAAAGACTGGAAACTAAAGA	
AACTTCTTATTGTCTACAGATAACAGATTCCGGAGCCACACAAACAC	
AAGCCCCACAAGAGGACCGTCTCGGAAGAAGAAGAGAGGCAACCTC	
TTCGAGCGACTCTCCGACAGCGAACAGCAGCTCCAGCTCAAGGGCAG	
ATAATACAACAAACTGAAAGACCTACAAAATTAGAATAACTAACAGCAA	
AAACACCGTTTACCTATTCCACCTGAACAAAAGAACAGAGACTAACAC	
CATGGGAATAACAGAAGACAAAGAAATAGCCATTATTGGCAGACCA	
CATAGATACTTTAAAGACATTCCCTTCTATTGGGATAACCCCCAGA	
GCCTAAAGTAAACTTGTGATTTCAATTAAGAAATAAAGGGCAAG	
GCCCCATTAACCTAAAGTCGGTGTCTACCTCTTAAGTTAACCTTACTA	
AACGGACTCCGCCCTCCATAATTGGCGCCAAAAGGGGGCTCGCCCCC	
TTAAACCCCAGGGGGCTCGCCCCCTAAACCCCCAAGGGGGCTACGCC	
CCTTACACCCCC (SEQ ID NO: 1008)	
Annotations:	
Putative Domain	Base range
TATA Box	152-158
Initiation Element	172-187
Transcriptional Start Site	182
5' UTR Conserved Domain	239-309
ORF2	343-633
ORF2/2	343-629; 2196-2505
ORF2/3	343-629; 2371-2734
ORF1	522-2540
ORF1/1	522-629; 2196-2540
ORF1/2	522-629; 2371-2505
Three open-reading frame region	2276-2502
GC-rich region	2803-2912

TABLE E2

Exemplary Anellovirus amino acid sequences ( <i>Betatorquevirus</i> )	
Ring 10 ( <i>Betatorquevirus</i> )	
ORF2	MFKSPTYLTTKGKNNALINCFVGDHDLLCSNNPAYHCLQILATTIPLQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGQHMLFPID NO: 1009)
ORF2/2	MFKSPTYLTTKGKNNALINCFVGDHDLLCSNNPAYHCLQILATTIPLQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGQHMLFPID TSMKQLRYRIQPDPQSTSCTPLTKGGDNLQKKLQNACLKTGKLKKLYCLQ NTDSRSQHKHPKRTRPRKKRATSSSDSEPSSSSAE (SEQ ID NO: 1010)
ORF2/3	MFKSPTYLTTKGKNNALINCFVGDHDLLCSNNPAYHCLQILATTIPLQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGQHMLFPID NTSPTRGPVLGRRRREQPLRATPPTANQAAPQAQNNTNIERPTKIRITNSKN TVYLFPPEQKNRRLTPWEIQEDKEIANLFGRPHRYFLKDIPFWDIPPEPKVN FDLNFO (SEQ ID NO: 1011)
ORF1	MPWWYRRRSYNPWRNNWFRPRKTIYRRYRRRRWVRKPFYKRKIKR LNIVEWQPKSIRKCRKGMLCLFQTTEDRLSYNFDMYEESIIPEKLPGGGFSI

TABLE E2-continued

Exemplary Anellovirus amino acid sequences ( <i>Betatorquevirus</i> )	
	KNISLYALYQEHIHAHNIFTHTTDRPLARYTGCSLKFYQSKDIDYVVVYSTS LPLRSSMGMNSMQPSIHLMQQNKLIVPSKQTQKRKPYIKKHISPPTMKS QWYFQHNIANIPLLMIRTTAALTLDNNYYIGSRQLSTNVNTIHTLNNTTYIQNRDW GDRNKTYYCOTLGTQRYFLYGHSTAQNINDIKLQELIPLINTQDYVQGFD WTEKDKNITTYKEFLTKGAGNPFAEWITAQNPNVIHTANSPTQIEQIYTAS TTTFQNKKLTDLPTPGYIFITPTVSLRPNYPKDLAERNKCYFVRSKINAHGW DPEQHQELINSQNLQWLQWLLFGYDPYIKRTQNQFALVDNYIIILVDHCPYTNPEK TPFIPPLSTSFIERGRSPYSPSDTHEPDEEDQNRWYPCYQQQESINSICLSPGPGT KIPKGITAEEKVKYSFNPKWGGDLPPMSTITNPTDQTYVVPNNFNNETTSLO NPTTRPEHFLYSFDERRGQLEKATKRLLKDWETKETSSLSTEYRFAEPTQT QAPQEDPSSEEEESNLFERLLRQRTKQLQLKRRRIIQTALKDLQKLE (SEQ ID NO: 1012)
ORF1/1	MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWPTYVVPNNFNETT SLQNPMTTRPEHFLYSFDERRGQLEKATKRLLKDWETKETSSLSTEYRFAEPT TQTQAPQEDPSSEEEESNLFERLLRQRTKQLQLKRRRIIQTALKDLQKLE
ORF 1/2	MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWNTDSRSQHKPH KRTRPRKKKKRATSSSDSEPSSSSSAE (SEQ ID NO: 1013)

**[0336]** In some embodiments, an anellovector comprises a nucleic acid comprising a sequence listed in PCT Application No. PCT/US2018/037379, incorporated herein by reference in its entirety. In some embodiments, an anellovector comprises a polypeptide comprising a sequence listed in PCT Application No. PCT/US2018/037379, incorporated herein by reference in its entirety. In some embodiments, an anellovector comprises a nucleic acid comprising a sequence listed in PCT Application No. PCT/US19/65995, incorporated herein by reference in its entirety. In some embodiments, an anellovector comprises a polypeptide comprising a sequence listed in PCT Application No. PCT/US19/65995, incorporated herein by reference in its entirety.

#### ORF1 Molecules

**[0337]** In some embodiments, the anellovector comprises an ORF1 molecule and/or a nucleic acid encoding an ORF1 molecule. Generally, an ORF1 molecule comprises a polypeptide having the structural features and/or activity of an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein). In some embodiments, the ORF1 molecule comprises a truncation relative to an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein). An ORF1 molecule may be capable of binding to other ORF1 molecules, e.g., to form a proteinaceous exterior (e.g., as described herein), e.g., a capsid. In some embodiments, the proteinaceous exterior may enclose a nucleic acid molecule (e.g., a genetic element as described herein). In some embodiments, a plurality of ORF1 molecules may form a multimer, e.g., to form a proteinaceous exterior. In some embodiments, the multimer may be a homomultimer. In other embodiments, the multimer may be a heteromultimer.

**[0338]** An ORF1 molecule may, in some embodiments, comprise one or more of: a first region comprising an arginine rich region, e.g., a region having at least 60% basic residues (e.g., at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% basic residues; e.g., between 60%-90%, 60%-80%, 70%-90%, or 70-80% basic residues), and a second region comprising jelly-roll domain, e.g., at least six beta strands (e.g., 4, 5, 6, 7, 8, 9, 10, 11, or 12 beta strands).

**[0339]** In some embodiments, an ORF1 molecule as described herein comprises one or more lysine-to-histidine

mutations relative to a wild-type ORF1 protein sequence (e.g., as described herein). In certain embodiments, the ORF1 molecule comprises one or more lysine-to-histidine mutations in the arginine-rich region and/or the first beta strand.

#### Arginine-Rich Region

**[0340]** An arginine rich region has at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an arginine-rich region sequence described herein or a sequence of at least about 40 amino acids comprising at least 60%, 70%, or 80% basic residues (e.g., arginine, lysine, or a combination thereof).

#### Jelly Roll Domain

**[0341]** A jelly-roll domain or region comprises (e.g., consists of) a polypeptide (e.g., a domain or region comprised in a larger polypeptide) comprising one or more (e.g., 1, 2, or 3) of the following characteristics:

**[0342]** (i) at least 30% (e.g., at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90%, or more) of the amino acids of the jelly-roll domain are part of one or more  $\beta$ -sheets;

**[0343]** (ii) the secondary structure of the jelly-roll domain comprises at least four (e.g., at least 4, 5, 6, 7, 8, 9, 10, 11, or 12)  $\beta$ -strands; and/or

**[0344]** (iii) the tertiary structure of the jelly-roll domain comprises at least two (e.g., at least 2, 3, or 4)  $\beta$ -sheets; and/or

**[0345]** (iv) the jelly-roll domain comprises a ratio of  $\beta$ -sheets to  $\alpha$ -helices of at least 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1.

**[0346]** In certain embodiments, a jelly-roll domain comprises two  $\beta$ -sheets.

**[0347]** In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the  $\beta$ -sheets comprises about eight (e.g., 4, 5, 6, 7, 8, 9, 10, 11, or 12)  $\beta$ -strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the  $\beta$ -sheets comprises eight  $\beta$ -strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the  $\beta$ -sheets comprises seven  $\beta$ -strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the  $\beta$ -sheets comprises six  $\beta$ -strands.

10) of the  $\beta$ -sheets comprises six  $\beta$ -strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the  $\beta$ -sheets comprises five  $\beta$ -strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the  $\beta$ -sheets comprises four  $\beta$ -strands.

[0348] In some embodiments, the jelly-roll domain comprises a first  $\beta$ -sheet in antiparallel orientation to a second  $\beta$ -sheet. In certain embodiments, the first  $\beta$ -sheet comprises about four (e.g., 3, 4, 5, or 6)  $\beta$ -strands. In certain embodiments, the second  $\beta$ -sheet comprises about four (e.g., 3, 4, 5, or 6)  $\beta$ -strands. In embodiments, the first and second  $\beta$ -sheet comprise, in total, about eight (e.g., 6, 7, 8, 9, 10, 11, or 12)  $\beta$ -strands.

[0349] In certain embodiments, a jelly-roll domain is a component of a capsid protein (e.g., an ORF1 molecule as described herein). In certain embodiments, a jelly-roll domain has self-assembly activity. In some embodiments, a polypeptide comprising a jelly-roll domain binds to another copy of the polypeptide comprising the jelly-roll domain. In some embodiments, a jelly-roll domain of a first polypeptide binds to a jelly-roll domain of a second copy of the polypeptide.

#### N22 Domain

[0350] An ORF1 molecule may also include a third region comprising the structure or activity of an Anellovirus N22 domain (e.g., as described herein, e.g., an N22 domain from an Anellovirus ORF1 protein as described herein), and/or a fourth region comprising the structure or activity of an Anellovirus C-terminal domain (CTD) (e.g., as described herein, e.g., a CTD from an Anellovirus ORF1 protein as described herein). In some embodiments, the ORF1 molecule comprises, in N-terminal to C-terminal order, the first, second, third, and fourth regions.

#### Hypervariable Region (HVR)

[0351] The ORF1 molecule may, in some embodiments, further comprise a hypervariable region (HVR), e.g., an HVR from an Anellovirus ORF1 protein, e.g., as described herein. In some embodiments, the HVR is positioned between the second region and the third region. In some embodiments, the HVR comprises at least about 55 (e.g., at least about 45, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, or 65) amino acids (e.g., about 45-160, 50-160, 55-160, 60-160, 45-150, 50-150, 55-150, 60-150, 45-140, 50-140, 55-140, or 60-140 amino acids).

#### Exemplary ORF1 Sequences

[0352] Exemplary Anellovirus ORF1 amino acid sequences, and the sequences of exemplary ORF1 domains, are provided in the tables below. In some embodiments, a

polypeptide (e.g., an ORF1 molecule) described herein comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more Anellovirus ORF1 subsequences, e.g., as described in any of Tables N-Z. In some embodiments, an anellovector described herein comprises an ORF1 molecule comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more Anellovirus ORF1 subsequences, e.g., as described in any of Tables N-Z. In some embodiments, an anellovector described herein comprises a nucleic acid molecule (e.g., a genetic element) encoding an ORF1 molecule comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more Anellovirus ORF1 subsequences, e.g., as described in any of Tables N-Z.

[0353] In some embodiments, the one or more Anellovirus ORF1 subsequences comprises one or more of an arginine (Arg)-rich domain, a jelly-roll domain, a hypervariable region (HVR), an N22 domain, or a C-terminal domain (CTD) (e.g., as listed in any of Tables N-Z), or sequences having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the ORF1 molecule comprises a plurality of subsequences from different Anelloviruses (e.g., any combination of ORF1 subsequences selected from the Alphatorquevirus Clade 1-7 subsequences listed in Tables N-Z). In embodiments, the ORF1 molecule comprises one or more of an Arg-rich domain, a jelly-roll domain, an N22 domain, and a CTD from one Anellovirus, and an HVR from another. In embodiments, the ORF1 molecule comprises one or more of a jelly-roll domain, an HVR, an N22 domain, and a CTD from one Anellovirus, and an Arg-rich domain from another. In embodiments, the ORF1 molecule comprises one or more of an Arg-rich domain, an HVR, an N22 domain, and a CTD from one Anellovirus, and a jelly-roll domain from another. In embodiments, the ORF1 molecule comprises one or more of an Arg-rich domain, a jelly-roll domain, an HVR, and a CTD from one Anellovirus, and an N22 domain from another. In embodiments, the ORF1 molecule comprises one or more of an Arg-rich domain, a jelly-roll domain, an HVR, and an N22 domain from one Anellovirus, and a CTD from another.

[0354] Additional exemplary Anelloviruses for which the ORF1 molecules, or splice variants or functional fragments thereof, can be utilized in the compositions and methods described herein (e.g., to form the proteinaceous exterior of an anellovector, e.g., by enclosing a genetic element) are described, for example, in PCT Application Nos. PCT/US2018/037379 and PCT/US19/65995 (incorporated herein by reference in their entirety).

TABLE N

Exemplary Anellovirus ORF1 amino acid subsequence (Alphatorquevirus, Clade 3)	
Name	Ring 1
Genus/Clade	Alphatorquevirus, Clade 3
Accession Number	AJ620231.1
Protein Accession Number	CAF05750.1

TABLE N-continued

Exemplary Anellovirus ORF1 amino acid subsequence (Alphatorquevirus, Clade 3)						
Full Sequence: 743 AA						
1	10	20	30	40	50	
MAWGWWKRRRRWWFRKRWTRGRRLRRRWPRSARRP RRRRVRRRRWRGR						
RKTRTYR RRRRFRRRGRKAKL11KLWQPAVIKRCRICKGYIPLII SGNGTF						
ATNFTSHINDRIMKGPFGGGHSTMRFLYILFEELRHMHNFWTRSNDNLE						
LTRYLGASVKIYRHPDQDFIVIYNRTPLGGNIYTAPS LHPGNAILAKHK						
ILVPSLQTRPKRKAIRLRIAPPTLFTDKWYFQKDIADLTLFNIMAVEAD						
LRPPFCSPQT DNTCISFQVLSSVYNNYLSINTFNNNDNSDSKLKEFLNKAF						
PTTGTGKTSNLNALTFRTEGCISHPQLKKPNPQINKPLES QYFAPLDALW						
GDP IYYNDLNENKS LNDIIEKILIKNMITYHAKLREFPNSYQGNKAFCHL						
TGIYSPPYLNLQGRISPEIFGLYTEIIYNPYTDKG TGNKVWMDPLTKENNI						
YKEGQSKCLLTDMPLWTLFGYTDWCKKD TNWDLPLNYRLVLCIPYTFP						
KLYNEKVKD YGIPYSYKFGAGQMPDG SNYIPFQFRAKWYPTVLHQQQVM						
EDISRGPFAPKVEKPSTQLMVKYCFN FNWGGNP II EQIVKDPSPFQPTYE						
I PG TGNI PRR IQVIDP RVL GPHYS FRS WDMR RHTFSRASIKR VSE QQETS						
DIVFSGPKKPRVDIPKQETQE ESSHSLQRESRPWETEESETEAL SQESQ						
EVPFQQQLQQQYQEQQLKLRQGIKVLFEQLIRTQQGVHVNPCLR						
(SEQ ID NO: 185)						
Annotations:						
Putative Domain	AA range					
Arg-Rich Region	1-68					
Jelly-roll domain	69-280					
Hypervariable Region	281-413					
N22	414-579					
C-terminal Domain	580-743					

TABLE O

Exemplary Anellovirus ORF1 amino acid subsequence (Alphatorquevirus, Clade 3)	
Ring1 ORF1 (Alphatorquevirus Clade 3)	
Arg-Rich Region	MAWGWWKRRRRWWFRKRWTRGRRLRRRWPRSARRP RRRRVRRRRWRR GRRKRTTYR RRRRFRRRGRK (SEQ ID NO: 186)
Jelly-roll Domain	AKLIIKLWQPAVIKRCRICKGYIPLII SGNGTFATNFTSHINDRIMKGPFGGHST MRFQDQDFIVIYNR TPLGGNIYTAPS LHPGNAILAKHKILVPSLQTRPKRKAIRLRIAPPTLFTDKWY FQKDIADLTLFNIMAVEADLRLPFCSPQT DNTCISFQVLSSVYNNYLS (SEQ ID NO: 187)
Hypervariable domain	NTFNNNDNSDSKLKEFLNKAFP TTGKGTSLNALNTFRTEGCISHPQLKKPNPQI NKP LESQYFAPLDALWDGPIYNDLNENKS LNDIIEKILIKNMITYHAKLREFP NSYQGNKAFCHLTG IYSPPYLNQGR (SEQ ID NO: 188)
N22	ISPEIFGLYTEIIYNPYTDKG TGNKVWMDPLTKENNIYKEGQSKCLLTDMPLW TLLFGYTDWCKKD TNWDLPLNYRLVLCIPYTFPKLYNEKVKD YGIPYSYK FGAGQMPDG SNYIPFQFRAKWYPTVLHQQQVM EDISRGPFAPKVEKPSTQL VMKYCFN FN (SEQ ID NO: 189)
C-terminal domain	WGGNP II EQIVKDPSPFQPTYEI PG TGNI PRR IQVIDP RVL GPHYS FRS WDMR HRTFSRASIKR VSE QQETS D LVFSGPKKPRVDIPKQETQE ESSHSLQRESRPWETEE SETEAL SQESQEVFPQQQLQQQYQEQQLKLRQGIKVLFEQLIRTQQGVHVNPCL R (SEQ ID NO: 190)

TABLE P

Exemplary Anellovirus ORF1 amino acid subsequence (Betatorquevirus)	
Name	Ring 2
Genus/Clade	Betatorquevirus
Accession Number	JX134045.1
Protein Accession Number	AGG91484.1

TABLE P-continued

**Exemplary Anellovirus ORF1 amino acid subsequence (*Betatorquevirus*)**

Full Sequence: 666 AA

1	10	20	30	40	50

MPYYYYRRRYNYRPRWYGRGWIRRPFRRFRKRRVRPTTYTTIPLKQWQ  
 PPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSM  
 TLDALYDIHKLCRNWWSTTNQDPLVRYKGCKITFYQSTFTDYIVRIHT  
 LPANSNKLTPNTHPLMMMSKYKHIIPSRQTRRKPKYTIFVKPPPQF  
 ENKWKWFATDLYKIPLLQIHCTACNLQNPFVVKPDKLSNNVTLWSLN  
 TISIQNRNSVDQGQSWPFKILGTQSFYFYFTGANLPGDTTQIPVADLLPL  
 TNP  
 RINRPQGSNLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHD  
 MLYSLKSPEAIKNEWTENMKWNQLNNAGTMALTPPNEPIFTQI  
 QYNPDRDTGEDTQLYLLSNATGTGWDPPGIP  
 ELEGFPLWLIYWGFA  
 DFQKNLKKV  
 TNIDTNYMLVAKTKFTQKPGTFYLV  
 ILNDTFVEGN  
 SPYEKQPLPEDNIK  
 WYQVQYQLEAQN  
 KLLQTGP  
 FPTPN  
 IQGQLSDN  
 ISMFYK  
 PAHQI  
 QYPIP  
 RNEHE  
 TTS  
 LQSPGEAP  
 ESI  
 LYSDYR  
 HGN  
 TT  
 ALSRIS  
 QDWALK  
 DTVSK  
 ITEPDR  
 QOLL  
 KQALEC  
 LQISEET  
 QEKEKE  
 EVQQL  
 ISNLRQ  
 QQOLYRERIISLLKDQ (SEQ ID NO: 215)

## Annotations:

Putative Domain	AA range
Arg-Rich Region	1-38
Jelly-roll domain	39-246
Hypervariable Region	247-374
N22	375-537
C-terminal Domain	538-666

TABLE Q

**Exemplary Anellovirus ORF1 amino acid subsequence (*Betatorquevirus*)****Ring2 ORF1 (*Betatorquevirus*)**

Arg-Rich Region	MPYYYYRRRYNYRPRWYGRGWIRRPFRRFRKRRVR (SEQ ID NO: 216)
Jelly-roll Domain	PTYTTIPLKQWOPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWP GGSFSVSMLTLDALYDIHKLCRNWWSTTNQDPLVRYKGCKITFYQSTFTDYI VRJHTELPANSNKLTPNTHPLMMMSKYKHIIPSRQTRRKPKYTIFVKPP PQFENKWFATDLYKIPLLQIHCTACNLQNPFVVKPDKLSNNVTLWSLN T (SEQ ID NO: 217)
Hypervariable domain	ISIQNRNSVDQGQSWPFKILGTQSFYFYFTGANLPGDTTQIPVADLLPL TNP RINRPQGSNLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHD MLYSLKSPEAIKNEWTENMKWNQLNNAG (SEQ ID NO: 218)
N22	TMALTPFNEPIFTQI QYNPDRDTGEDTQLYLLSNATGTGWDPPGIP ELEGFPLWLIYWGFA DFQKNLKKV TNIDTNYMLVAKTKFTQKPGTFYLV ILNDTFVEGN SPYEKQPLPEDNIK WYQVQYQLEAQN KLLQTGP FPTPN IQGQLSDN ISMFYK PAHQI QYPIP RNEHE TTS LQSPGEAP ESI LYSDYR HGN TT ALSRIS QDWALK DTVSK ITEPDR QOLL KQALEC LQISEET QEKEKE EVQQL ISNLRQ QQOLYRERIISLLKDQ (SEQ ID NO: 219)
C-terminal domain	WGGSPPKAINVENPAHQI QYPIP RNEHE TTS LQSPGEAP ESI LYSDYR HGN TT ALSRIS QDWALK DTVSK ITEPDR QOLL KQALEC LQISEET QEKEKE EVQQL ISNLRQ QQOLYRERIISLLKDQ (SEQ ID NO: 220)

TABLE R

Exemplary Anellovirus ORF1 amino acid subsequence ( <i>Gammatorquevirus</i> )						
Name	Ring 4					
Genus/Clade	<i>Gammatorquevirus</i>					
Accession Number						
Protein Accession Number						
Full Sequence: 662 AA						
1	10	20	30	40	50	
MPFWWRRRKFWTNNRFNYTKRRRYRKWRPRRRRRPyRRPVRRRRKL						
RKVKKKSLIVRQWPDSIRTCKIIGQSIAVGAEGKQMYCYTVNKLIN						
VPPKTPYGGGFGVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLI						
FYRDNKTDVFVLSYDRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSOMTLP						
NGRLTKKLKIKPPKQMLSKWFFSKQFCYPLLSLKASALDLRHSHYLGCCN						
ENPQVFFYYLNHGYYTITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNIN						
IKSHEYEKSIISYENGYFQSSFLQTQCIYTSEERGEACIAEKPLGIAIYNPV						
KDNGDGNNIYLVSTLANTWDQPPKDSAILIQGVPIWLGLFGYLDYCROIK						
ADKTWLDHSVVLVIQSPAIFTYPNPGAGKWCPLSQSFINGNGPFNQPPTL						
LQKAKWFPQIQQEIIINSFVESGPFPVKYANQTESNWELEYKVKYVTFKW						
GGPQFHEPEIADPSKQEQYDVPDTFYQTQIEDPEGQDPRSLIHWDYRR						
GFIKERSLKRMSTYFSTHTDQATSEEDIPKKKRIGPQLTVPQQKEET						
LSCLLSLCKDKTFQETETQEDLQQLIKQQQEQQLLLKRNILOLIHKLKEN						
QQMLQLHTGMLP (SEQ ID NO: 925)						
Annotations:						
Putative Domain	AA range					
Arg-Rich Region	1-58					
Jelly-roll domain	59-260					
Hypervariable Region	261-339					
N22	340-499					
C-terminal Domain	500-662					

TABLE S

Exemplary Anellovirus ORF1 amino acid subsequence ( <i>Gammatorquevirus</i> )	
Ring4 ( <i>Gammatorquevirus</i> )	
Arg-Rich Region	MPFWWRRRKFWTNNRFNYTKRRRYRKWRPRRRRRPyRRPVRRRRKL RKVKRKKK (SEQ ID NO: 926)
Jelly-roll Domain	SLIVRQWPDSIRTCKIIGQSIAVVAEGKQMYCYTVNKLINVPPKTPYGGF GVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLIIFYRDNKTDVFVLSY DRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSOMTLPNGRLLTKKLKIKPPKQM LSKWFFSKQFCYPLLSLKASALDLRHSHYLGCCNENPQVFFYL (SEQ ID NO: 927)
Hypervariable domain	NHGYYTIITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNINIKSHEYEKSIYE NGYFQSSFLQTQCIYTSEERGEACIAE (SEQ ID NO: 928)
N22	KPLGIAIYNPVKDNGDGNNIYLVSTLANTWDQPPKDSAILIQGVPIWLGLFGY LDYCROIKADKTWLDHSVVLVIQSPAIFTYPNPGAGKWCPLSQSFINGNGPFN QFPTLLOKAKWFPQIQQEIIINSFVESGPFPVKYANQTESNWELEYKVKYVTF K (SEQ ID NO: 929)
C-terminal domain	WGGPQFHEPEIADPSKQEQYDVPDTFYQTQIEDPEGQDPRSLIHWDYRRGFI KERSLKRMSTYFSTHTDQATSEEDIPKKKRIGPQLTVPQQKEETLSCLLSL CKKDTFQETETQEDLQQLIKQQQEQQLLLKRNILOLIHKLKENQQMLQLHTG MLP (SEQ ID NO: 930)

**[0355]** In some embodiments, the first region can bind to a nucleic acid molecule (e.g., DNA). In some embodiments, the basic residues are selected from arginine, histidine, or lysine, or a combination thereof. In some embodiments, the first region comprises at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% arginine residues (e.g., between 60%-90%, 60%-80%, 70%-90%, or 70-80% arginine residues). In some embodiments, the first region comprises about 30-120 amino acids (e.g., about 40-120, 40-100, 40-90, 40-80, 40-70, 50-100, 50-90, 50-80, 50-70, 60-100, 60-90, or 60-80 amino acids). In some embodiments, the first region comprises the structure or activity of a viral ORF1 arginine-rich region (e.g., an arginine-rich region from an Anellovirus ORF1 protein, e.g., as described herein). In some embodiments, the first region comprises a nuclear localization signal.

**[0356]** In some embodiments, the second region comprises a jelly-roll domain, e.g., the structure or activity of a viral ORF1 jelly-roll domain (e.g., a jelly-roll domain from an Anellovirus ORF1 protein, e.g., as described herein). In some embodiments, the second region is capable of binding to the second region of another ORF1 molecule, e.g., to form a proteinaceous exterior (e.g., capsid) or a portion thereof.

**[0357]** In some embodiments, the fourth region is exposed on the surface of a proteinaceous exterior (e.g., a proteinaceous exterior comprising a multimer of ORF1 molecules, e.g., as described herein).

**[0358]** In some embodiments, the first region, second region, third region, fourth region, and/or HVR each comprise fewer than four (e.g., 0, 1, 2, or 3) beta sheets.

**[0359]** In some embodiments, one or more of the first region, second region, third region, fourth region, and/or HVR may be replaced by a heterologous amino acid sequence (e.g., the corresponding region from a heterologous ORF1 molecule). In some embodiments, the heterologous amino acid sequence has a desired functionality, e.g., as described herein.

**[0360]** In some embodiments, the ORF1 molecule comprises a plurality of conserved motifs (e.g., motifs comprising about 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, or more amino acids) (e.g., as shown in FIG. 34 of PCT/US19/65995). In some embodiments, the conserved motifs may show 60, 70, 80, 85, 90, 95, or 100% sequence identity to an ORF1 protein of one or more wild-type Anellovirus clades (e.g., Alphatorquevirus, clade 1; Alphatorquevirus, clade 2; Alphatorquevirus, clade 3; Alphatorquevirus, clade 4; Alphatorquevirus, clade 5; Alphatorquevirus, clade 6; Alphatorquevirus, clade 7; Betatorquevirus; and/or Gammatorquevirus). In embodiments, the conserved motifs each have a length between 1-1000 (e.g., between 5-10, 5-15, 5-20, 10-15, 10-20, 15-20, 5-50, 5-100, 10-50, 10-100, 10-1000, 50-100, 50-1000, or 100-1000) amino acids. In certain embodiments, the conserved motifs consist of about 2-4% (e.g., about 1-8%, 1-6%, 1-5%, 1-4%, 2-8%, 2-6%, 2-5%, or 2-4%) of the sequence of the ORF1 molecule, and each show 100% sequence identity to the corresponding motifs in an ORF1 protein of the wild-type Anellovirus clade. In certain embodiments, the conserved motifs consist of about 5-10% (e.g., about 1-20%, 1-10%, 5-20%, or 5-10%) of the sequence of the ORF1 molecule, and each show 80% sequence identity to the corresponding motifs in an ORF1 protein of the wild-type Anellovirus clade. In certain embodiments, the conserved motifs consist of about

10-50% (e.g., about 10-20%, 10-30%, 10-40%, 10-50%, 20-40%, 20-50%, or 30-50%) of the sequence of the ORF1 molecule, and each show 60% sequence identity to the corresponding motifs in an ORF1 protein of the wild-type Anellovirus clade. In some embodiments, the conserved motifs comprise one or more amino acid sequences as listed in Table 19.

**[0361]** In some embodiments, an ORF1 molecule or a nucleic acid molecule encoding same comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type ORF1 protein, e.g., as described herein.

#### Conserved ORF1 Motif in N22 Domain

**[0362]** In some embodiments, a polypeptide (e.g., an ORF1 molecule) described herein comprises the amino acid sequence YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829), wherein X<sup>2</sup> is a contiguous sequence of any n amino acids. For example, X<sup>2</sup> indicates a contiguous sequence of any two amino acids. In some embodiments, the YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829) is comprised within the N22 domain of an ORF1 molecule, e.g., as described herein. In some embodiments, a genetic element described herein comprises a nucleic acid sequence (e.g., a nucleic acid sequence encoding an ORF1 molecule, e.g., as described herein) encoding the amino acid sequence YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829), wherein X<sup>2</sup> is a contiguous sequence of any n amino acids.

**[0363]** In some embodiments, a polypeptide (e.g., an ORF1 molecule) comprises a conserved secondary structure, e.g., flanking and/or comprising a portion of the YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829) motif, e.g., in an N22 domain. In some embodiments, the conserved secondary structure comprises a first beta strand and/or a second beta strand. In some embodiments, the first beta strand is about 5-6 (e.g., 3, 4, 5, 6, 7, or 8) amino acids in length. In some embodiments, the first beta strand comprises the tyrosine (Y) residue at the N-terminal end of the YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829) motif. In some embodiments, the YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829) motif comprises a random coil (e.g., about 8-9 amino acids of random coil). In some embodiments, the second beta strand is about 7-8 (e.g., 5, 6, 7, 8, 9, or 10) amino acids in length. In some embodiments, the second beta strand comprises the asparagine (N) residue at the C-terminal end of the YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829) motif.

**[0364]** Exemplary YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829) motif-flanking secondary structures are described in Example 47 and FIG. 48 of PCT/US19/65995; incorporated herein by reference in its entirety. In some embodiments, an ORF1 molecule comprises a region comprising one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or all) of the secondary structural elements (e.g., beta strands) shown in FIG. 48 of PCT/US19/65995. In some embodiments, an ORF1 molecule comprises a region comprising one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or all) of the secondary structural elements (e.g., beta strands) shown in FIG. 48 of PCT/US19/65995, flanking a YNPX<sup>2</sup>DXGX<sup>2</sup>N (SEQ ID NO: 829) motif (e.g., as described herein).

#### Conserved Secondary Structural Motif in ORF1 Jelly-Roll Domain

**[0365]** In some embodiments, a polypeptide (e.g., an ORF1 molecule) described herein comprises one or more

secondary structural elements comprised by an Anellovirus ORF1 protein (e.g., as described herein). In some embodiments, an ORF1 molecule comprises one or more secondary structural elements comprised by the jelly-roll domain of an Anellovirus ORF1 protein (e.g., as described herein). Generally, an ORF1 jelly-roll domain comprises a secondary structure comprising, in order in the N-terminal to C-terminal direction, a first beta strand, a second beta strand, a first alpha helix, a third beta strand, a fourth beta strand, a fifth beta strand, a second alpha helix, a sixth beta strand, a seventh beta strand, an eighth beta strand, and a ninth beta strand. In some embodiments, an ORF1 molecule comprises a secondary structure comprising, in order in the N-terminal to C-terminal direction, a first beta strand, a second beta strand, a first alpha helix, a third beta strand, a fourth beta strand, a fifth beta strand, a second alpha helix, a sixth beta strand, a seventh beta strand, an eighth beta strand, and a ninth beta strand.

**[0366]** In some embodiments, a pair of the conserved secondary structural elements (i.e., the beta strands and/or alpha helices) are separated by an interstitial amino acid sequence, e.g., comprising a random coil sequence, a beta strand, or an alpha helix, or a combination thereof. Interstitial amino acid sequences between the conserved secondary structural elements may comprise, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more amino acids. In some embodiments, an ORF1 molecule may further comprise one or more additional beta strands and/or alpha helices (e.g., in the jelly-roll domain). In some embodiments, consecutive beta strands or consecutive alpha helices may be combined. In some embodiments, the first beta strand and the second beta strand are comprised in a larger beta strand. In some embodiments, the third beta strand and the fourth beta strand are comprised in a larger beta strand. In some embodiments, the fourth beta strand and the fifth beta strand are comprised in a larger beta strand. In some embodiments, the sixth beta strand and the seventh beta strand are comprised in a larger beta strand. In some embodiments, the seventh beta strand and the eighth beta strand are comprised in a larger beta strand. In some embodiments, the eighth beta strand and the ninth beta strand are comprised in a larger beta strand.

**[0367]** In some embodiments, the first beta strand is about 5-7 (e.g., 3, 4, 5, 6, 7, 8, 9, or 10) amino acids in length. In some embodiments, the second beta strand is about 15-16 (e.g., 13, 14, 15, 16, 17, 18, or 19) amino acids in length. In some embodiments, the first alpha helix is about 15-17 (e.g., 13, 14, 15, 16, 17, 18, 19, or 20) amino acids in length. In some embodiments, the third beta strand is about 3-4 (e.g.,

1, 2, 3, 4, 5, or 6) amino acids in length. In some embodiments, the fourth beta strand is about 10-11 (e.g., 8, 9, 10, 11, 12, or 13) amino acids in length. In some embodiments, the fifth beta strand is about 6-7 (e.g., 4, 5, 6, 7, 8, 9, or 10) amino acids in length. In some embodiments, the second alpha helix is about 8-14 (e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17) amino acids in length. In some embodiments, the second alpha helix may be broken up into two smaller alpha helices (e.g., separated by a random coil sequence). In some embodiments, each of the two smaller alpha helices are about 4-6 (e.g., 2, 3, 4, 5, 6, 7, or 8) amino acids in length. In some embodiments, the sixth beta strand is about 4-5 (e.g., 2, 3, 4, 5, 6, or 7) amino acids in length. In some embodiments, the seventh beta strand is about 5-6 (e.g., 3, 4, 5, 6, 7, 8, or 9) amino acids in length. In some embodiments, the eighth beta strand is about 7-9 (e.g., 5, 6, 7, 8, 9, 10, 11, 12, or 13) amino acids in length. In some embodiments, the ninth beta strand is about 5-7 (e.g., 3, 4, 5, 6, 7, 8, 9, or 10) amino acids in length.

**[0368]** Exemplary jelly-roll domain secondary structures are described in Example 47 and FIG. 47 of PCT/US19/65995. In some embodiments, an ORF1 molecule comprises a region comprising one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or all) of the secondary structural elements (e.g., beta strands and/or alpha helices) of any of the jelly-roll domain secondary structures shown in FIG. 47 of PCT/US19/65995.

#### Consensus ORF1 Domain Sequences

**[0369]** In some embodiments, an ORF1 molecule, e.g., as described herein, comprises one or more of a jelly-roll domain, N22 domain, and/or C-terminal domain (CTD). In some embodiments, the jelly-roll domain comprises an amino acid sequence having a jelly-roll domain consensus sequence as described herein (e.g., as listed in any of Tables 37A-37C). In some embodiments, the N22 domain comprises an amino acid sequence having a N22 domain consensus sequence as described herein (e.g., as listed in any of Tables 37A-37C). In some embodiments, the CTD domain comprises an amino acid sequence having a CTD domain consensus sequence as described herein (e.g., as listed in any of Tables 37A-37C). In some embodiments, the amino acids listed in any of Tables 37A-37C in the format “(X<sub>a-b</sub>)” comprise a contiguous series of amino acids, in which the series comprises at least a, and at most b, amino acids. In certain embodiments, all of the amino acids in the series are identical. In other embodiments, the series comprises at least two (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 21) different amino acids.

TABLE 37A

Alphatorquevius ORF1 domain consensus sequences		
Domain	Sequence	SEQ ID NO:
Jelly-Roll	LVLTOWQPNTVRRCYIRGYLPLIICGEN (X <sub>0-3</sub> ) TTSRNYATHS DDTIQKGPFGGGMSTTFSLRVLYDEYQRFMNRWTYSNED LDLARYLGCKFTFYRHPDXDFIVQYNTNPFFKDTKLAPSIIH P (X <sub>1-5</sub> ) GMMLMSKRKILIPSLKTRPKGKHVVKVRIGPPKLFED KWYTQSDLCDVPLVXLYATAADLQHPFGSPQTDNPCVTFO VLGSXYNKHLSISP; wherein X = any amino acid.	227
N22	SNFEFPGAYTDITYNPLTDKGVNMMWIQYLTKPDTIXDKT QS (X <sub>0-3</sub> ) KCLIEDLPLWAALYGVDFCEKETGDSAIIXNXGRV LIRCPYTKPPLYDKT (X <sub>0-4</sub> ) NKGFVPVYSTNFNGKMPGGSGY	228

TABLE 37A-continued

<i>Alphatorquevius</i> ORF1 domain consensus sequences		
Domain	Sequence	SEQ ID NO:
	VPIYWRARWYPTLHFHQKEVLEDIVQSGPPAYKDEKPSTQLV MKYCFNFn; wherein X = any amino acid.	
CTD	WGGNPISQQVVRNPCKDSG (X <sub>0-3</sub> ) SGXGRQPRSVQVVDPKY MGPEYTFHSHWDWRRLGFGEKAIRMS EQPTDDEIFTGGXPK RPRRDPTXQXPEE (X <sub>1-4</sub> ) QKESSFR (X <sub>2-14</sub> ) PWESSSQEXESES QEEE (X <sub>0-30</sub> ) EQTVQQQLRQQLREQRRLRVQLQLLFQQLLKT (X <sub>0-4</sub> ) QAGLHINPPLLSQA (X <sub>0-40</sub> ) *; wherein X = any amino acid.	229

TABLE 37B

<i>Betatorquevius</i> ORF1 domain consensus sequences		
Domain	Sequence	SEQ ID NO:
Jelly-Roll	LKQWQPSTIRKCKIKGYLPLFQCGKGRISNNYTQYKESIVPH HEPGGGGWSIQQFTL GALYEEHLKLRNWWTKSNDGLPLV YLGCTIKLYRSEDTDYIVTYQRCYPMATAKLTYLSTQPSRM LMNKHKIIIVPSKXT (X <sub>1-4</sub> ) NKKKKPYKKI FIKPPSQMQNKWYF QDQIANTPLLQLTXXTACSLDRMYLSSDSISNNITFTSLNTNF QNPNFQ; wherein X = any amino acid.	230
N22	(X <sub>4-10</sub> ) TPLYFECRYNPFPDKKGTKGNKVYLVSN (X <sub>1-8</sub> ) TGWDPP TDPDLIIEGFPLWLLLWGWLWQKKGKIQNI DTDYILVIQS XYYIPP (X <sub>1-3</sub> ) KLPYVVPDXD (X <sub>0-2</sub> ) FLHGRSPY (X <sub>3-16</sub> ) PSDKQH WHPKVRFQXETINNIALTGPGTPKLPNQKS IQAHMKYKFYF K; wherein X = any amino acid.	231
CTD	WGGCPAPM ETIDPCQPKYPIPNNLQTTSLQXPPTPIETY YKFDERRGLLT KKA AKRIKKDXTTETTLFTDTGXXTSTLPT XXQTETTQEEXTSEEE (X <sub>0-5</sub> ) ETLLQQLQQLRRQKQLRXRIL QLLQLLXL (X <sub>0-26</sub> ) *; wherein X = any amino acid.	232

TABLE 37C

<i>Gammatorquevius</i> ORF1 domain consensus sequences		
Domain	Sequence	SEQ ID NO:
Jelly-Roll	TIPLKQWQPESIRKCKIKGYGTLVLGAEGRQFYCYTNEKDE YTPPKAPGGGGFVELFSLEYLYEQWKARNNIWTKSNXYK DLCRYTGCKITFYRHTTDFTVXYSRQPPFEIDKXTYMXXP QXLLLRKHKKIILSKATNPKGK LKKKIKKPKQMLNKWPF QKQFAXYGLVQLQAAACBLRYPRLGCCNENR LITLYYL; wherein X = any amino acid.	233
N22	LPIVVARYNPA DTGKGNKWLXSTLNGSXWAPPDKDL II EGLPLWLALYGYWSYJJKVKKDKGILQSHMFVVKSPAIQP LXTATTQXTFYPXIDNSFIQGKXPYDEPJT XNQKKLWYPTLE HQQETINAIVESGPYVPKLDNQKNSTWELXYXYTFYFK; wherein X = any amino acid.	234
CTD	WGGPQI PDPQPV EDPKXQGTYPV PDPDXQQT I QIXNPLKQKPE TMFHDWDYRRGI ITSTALKRMQENLET DSSFXSDSEETP (X <sub>0-2</sub> ) KKKKR LTXELPXPQEEETEEIQSCLL SLC EESTCQEE (X <sub>1-6</sub> ) ENL QQLIHQQQQQQQLKHNLKLLSDLKZKQRLLQLQTGILE (X <sub>1-10</sub> ) *; wherein X = any amino acid.	235

**[0370]** In some embodiments, the jelly-roll domain comprises a jelly-roll domain amino acid sequence as listed in any of Tables 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10, or 37A-37C, or an amino acid sequence having at least 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the N22 domain comprises a N22 domain amino acid sequence as listed in any of Tables 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10, or 37A-37C, or an amino acid sequence having at least 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the CTD domain comprises a CTD domain amino acid sequence as listed in any of Tables 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10, or 37A-37C, or an amino acid sequence having at least 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

#### Identification of ORF1 Protein Sequences

**[0371]** In some embodiments, an Anellovirus ORF1 protein sequence, or a nucleic acid sequence encoding an ORF1 protein, can be identified from the genome of an Anellovirus (e.g., a putative Anellovirus genome identified, for example, by nucleic acid sequencing techniques, e.g., deep sequencing techniques). In some embodiments, an ORF1 protein sequence is identified by one or more (e.g., 1, 2, or all 3) of the following selection criteria:

**[0372]** (i) Length Selection: Protein sequences (e.g., putative Anellovirus ORF1 sequences passing the criteria described in (ii) or (iii) below) may be size-selected for those greater than about 600 amino acid residues to identify putative Anellovirus ORF1 proteins. In some embodiments, an Anellovirus ORF1 protein sequence is at least about 600, 650, 700, 750, 800, 850, 900, 950, or 1000 amino acid residues in length. In some embodiments, an Alphatorquevirus ORF1 protein sequence is at least about 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 900, or 1000 amino acid residues in length. In some embodiments, a Betatorquevirus ORF1 protein sequence is at least about 650, 660, 670, 680, 690, 700, 750, 800, 900, or 1000 amino acid residues in length. In some embodiments, a Gammatorquevirus ORF1 protein sequence is at least about 650, 660, 670, 680, 690, 700, 750, 800, 900, or 1000 amino acid residues in length. In some embodiments, a nucleic acid sequence encoding an Anellovirus ORF1 protein is at least about 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 nucleotides in length. In some embodiments, a nucleic acid sequence encoding an Alphatorquevirus ORF1 protein sequence is at least about 2100, 2150, 2200, 2250, 2300, 2400, or 2500 nucleotides in length. In some embodiments, a nucleic acid sequence encoding a Betatorquevirus ORF1 protein sequence is at least about 1900, 1950, 2000, 2500, 2100, 2150, 2200, 2250, 2300, 2400, or 2500 or 1000 nucleotides in length. In some embodiments, a nucleic acid sequence encoding a Gammatorquevirus ORF1 protein sequence is at least about 1900, 1950, 2000, 2500, 2100, 2150, 2200, 2250, 2300, 2400, or 2500 or 1000 nucleotides in length.

**[0373]** (ii) Presence of ORF1 motif Protein sequences (e.g., putative Anellovirus ORF1 sequences passing the criteria described in (i) above or (iii) below) may be filtered to identify those that contain the conserved ORF1 motif in the N22 domain described above. In some embodiments, a putative Anellovirus ORF1 sequence comprises the

sequence YNPXXDXGXXN (SEQ ID NO: 829). In some embodiments, a putative Anellovirus ORF1 sequence comprises the sequence Y[NCS]PXXDX[GASKR]XX[NTS-VAK](SEQ ID NO: 950).

**[0374]** (iii) Presence of arginine-rich region: Protein sequences (e.g., putative Anellovirus ORF1 sequences passing the criteria described in (i) and/or (ii) above) may be filtered for those that include an arginine-rich region (e.g., as described herein). In some embodiments, a putative Anellovirus ORF1 sequence comprises a contiguous sequence of at least about 30, 35, 40, 45, 50, 55, 60, 65, or 70 amino acids that comprises at least 30% (e.g., at least about 20%, 25%, 30%, 35%, 40%, 45%, or 50%) arginine residues. In some embodiments, a putative Anellovirus ORF1 sequence comprises a contiguous sequence of about 35-40, 40-45, 45-50, 50-55, 55-60, 60-65, or 65-70 amino acids that comprises at least 30% (e.g., at least about 20%, 25%, 30%, 35%, 40%, 45%, or 50%) arginine residues. In some embodiments, the arginine-rich region is positioned at least about 30, 40, 50, 60, 70, or 80 amino acids downstream of the start codon of the putative Anellovirus ORF1 protein. In some embodiments, the arginine-rich region is positioned at least about 50 amino acids downstream of the start codon of the putative Anellovirus ORF1 protein.

**[0375]** ORF2 Molecules In some embodiments, the anellovector comprises an ORF2 molecule and/or a nucleic acid encoding an ORF2 molecule. Generally, an ORF2 molecule comprises a polypeptide having the structural features and/or activity of an Anellovirus ORF2 protein (e.g., an Anellovirus ORF2 protein as described herein, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18), or a functional fragment thereof. In some embodiments, an ORF2 molecule comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF2 protein sequence as shown in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18.

**[0376]** In some embodiments, an ORF2 molecule comprises an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to an Alphatorquevirus, Betatorquevirus, or Gammatorquevirus ORF2 protein. In some embodiments, an ORF2 molecule (e.g., an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to an Alphatorquevirus ORF2 protein) has a length of 250 or fewer amino acids (e.g., about 150-200 amino acids). In some embodiments, an ORF2 molecule (e.g., an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to a Betatorquevirus ORF2 protein) has a length of about 50-150 amino acids. In some embodiments, an ORF2 molecule (e.g., an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to a Gammatorquevirus ORF2 protein) has a length of about 100-200 amino acids (e.g., about 100-150 amino acids). In some embodiments, the ORF2 molecule comprises a helix-turn-helix motif (e.g., a helix-turn-helix motif comprising two alpha helices flanking a turn region). In some embodiments, the ORF2 molecule does not comprise the amino acid sequence of the ORF2 protein of TTV isolate TA278 or TTV isolate SANBAN. In some embodiments, an ORF2 molecule has protein phosphatase activity. In some embodiments, an ORF2 molecule or a nucleic acid molecule encoding same

comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type ORF2 protein, e.g., as described herein (e.g., as shown in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18).

#### Conserved ORF2 Motif

**[0377]** In some embodiments, a polypeptide (e.g., an ORF2 molecule) described herein comprises the amino acid sequence [W/F]X<sup>7</sup>HX<sup>3</sup>CX<sup>1</sup>CX<sup>5</sup>H (SEQ ID NO: 949), wherein X<sup>n</sup> is a contiguous sequence of any n amino acids. In embodiments, X<sup>7</sup> indicates a contiguous sequence of any seven amino acids. In embodiments, X<sup>3</sup> indicates a contiguous sequence of any three amino acids. In embodiments, X<sup>1</sup> indicates any single amino acid. In embodiments, X<sup>5</sup> indicates a contiguous sequence of any five amino acids. In some embodiments, the [W/F] can be either tryptophan or phenylalanine. In some embodiments, the [W/F] X<sup>7</sup>HX<sup>3</sup>CX<sup>1</sup>CX<sup>5</sup>H (SEQ ID NO: 949) is comprised within the N22 domain of an ORF2 molecule, e.g., as described herein. In some embodiments, a genetic element described herein comprises a nucleic acid sequence (e.g., a nucleic acid sequence encoding an ORF2 molecule, e.g., as described herein) encoding the amino acid sequence [W/F] X<sup>7</sup>HX<sup>3</sup>CX<sup>1</sup>CX<sup>5</sup>H (SEQ ID NO: 949), wherein X<sup>n</sup> is a contiguous sequence of any n amino acids.

#### Genetic Elements

**[0378]** In some embodiments, the anellovector comprises a genetic element. In some embodiments, the genetic element has one or more of the following characteristics: is substantially non-integrating with a host cell's genome, is an episomal nucleic acid, is a single stranded RNA, is circular, is about 1 to 10 kb, exists within the nucleus of the cell, can be bound by endogenous proteins, produces an effector, such as a polypeptide or nucleic acid (e.g., an RNA, iRNA, microRNA) that targets a gene, activity, or function of a host or target cell. In one embodiment, the genetic element is a substantially non-integrating. In some embodiments, the genetic element comprises a packaging signal, e.g., a sequence that binds a capsid protein. In some embodiments, outside of the packaging or capsid-binding sequence, the genetic element has less than 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5% sequence identity to a wild type Anellovirus nucleic acid sequence, e.g., has less than 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5% sequence identity to an Anellovirus nucleic acid sequence, e.g., as described herein. In some embodiments, outside of the packaging or capsid-binding sequence, the genetic element has less than 500, 450, 400, 350, 300, 250, 200, 150, or 100 contiguous nucleotides that are at least 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to an Anellovirus nucleic acid sequence.

**[0379]** In some embodiments, the genetic element has a length less than 20 kb (e.g., less than about 19 kb, 18 kb, 17 kb, 16 kb, 15 kb, 14 kb, 13 kb, 12 kb, 11 kb, 10 kb, 9 kb, 8 kb, 7 kb, 6 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, or less). In some embodiments, the genetic element has, independently or in addition to, a length greater than 1000b (e.g., at least about 1.1 kb, 1.2 kb, 1.3 kb, 1.4 kb, 1.5 kb, 1.6 kb, 1.7 kb, 1.8 kb, 1.9 kb, 2 kb, 2.1 kb, 2.2 kb, 2.3 kb, 2.4 kb, 2.5 kb, 2.6 kb, 2.7 kb, 2.8 kb, 2.9 kb, 3 kb, 3.1 kb, 3.2 kb, 3.3 kb, 3.4 kb, 3.5 kb, 3.6 kb, 3.7 kb, 3.8 kb, 3.9 kb, 4 kb, 4.1 kb, 4.2 kb,

4.3 kb, 4.4 kb, 4.5 kb, 4.6 kb, 4.7 kb, 4.8 kb, 4.9 kb, 5 kb, or greater). In some embodiments, the genetic element has a length of about 2.5-4.6, 2.8-4.0, 3.0-3.8, or 3.2-3.7 kb. In some embodiments, the genetic element has a length of about 1.5-2.0, 1.5-2.5, 1.5-3.0, 1.5-3.5, 1.5-3.8, 1.5-3.9, 1.5-4.0, 1.5-4.5, or 1.5-5.0 kb. In some embodiments, the genetic element has a length of about 2.0-2.5, 2.0-3.0, 2.0-3.5, 2.0-3.8, 2.0-3.9, 2.0-4.0, 2.0-4.5, or 2.0-5.0 kb. In some embodiments, the genetic element has a length of about 2.5-3.0, 2.5-3.5, 2.5-3.8, 2.5-3.9, 2.5-4.0, 2.5-4.5, or 2.5-5.0 kb. In some embodiments, the genetic element has a length of about 3.0-5.0, 3.5-5.0, 4.0-5.0, or 4.5-5.0 kb. In some embodiments, the genetic element has a length of about 1.5-2.0, 2.0-2.5, 2.5-3.0, 3.0-3.5, 3.1-3.6, 3.2-3.7, 3.3-3.8, 3.4-3.9, 3.5-4.0, 4.0-4.5, or 4.5-5.0 kb. In some embodiments, the genetic element has a length between about 3.6-3.9 kb. In some embodiments, the genetic element has a length between about 2.8-2.9 kb. In some embodiments, the genetic element has a length between about 2.0-3.2 kb.

**[0380]** In some embodiments, the genetic element comprises one or more of the features described herein, e.g., a sequence encoding a substantially non-pathogenic protein, a protein binding sequence, one or more sequences encoding a regulatory nucleic acid, one or more regulatory sequences, one or more sequences encoding a replication protein, and other sequences.

**[0381]** In embodiments, the genetic element was produced from a double-stranded circular DNA (e.g., produced by transcription).

**[0382]** In some embodiments, the genetic element does not comprise one or more bacterial plasmid elements (e.g., a bacterial origin of replication or a selectable marker, e.g., a bacterial resistance gene). In some embodiments, the genetic element does not comprise a bacterial plasmid backbone.

**[0383]** In one embodiment, the disclosure provides a genetic element comprising a nucleic acid sequence (e.g., an RNA sequence) encoding (i) a substantially non-pathogenic exterior protein, (ii) an exterior protein binding sequence that binds the genetic element to the substantially non-pathogenic exterior protein, and (iii) a regulatory nucleic acid. In such an embodiment, the genetic element may comprise one or more sequences with at least about 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide sequences to a native viral sequence (e.g., a native Anellovirus sequence, e.g., as described herein).

#### Protein Binding Sequence

**[0384]** In some embodiments, the genetic element encodes a protein binding sequence that binds to the substantially non-pathogenic protein. In some embodiments, the protein binding sequence facilitates packaging the genetic element into the proteinaceous exterior. In some embodiments, the protein binding sequence specifically binds an arginine-rich region of the substantially non-pathogenic protein. In some embodiments, the genetic element comprises a protein binding sequence as described in Example 8 of PCT/US19/65995.

**[0385]** In some embodiments, the genetic element comprises a protein binding sequence having at least 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence

identity to a 5' UTR conserved domain or GC-rich domain of an Anellovirus sequence, e.g., as described herein.

[0386] In embodiments, the protein binding sequence has at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus 5' UTR conserved domain nucleotide sequence, e.g., as described herein.

#### 5' UTR Regions

[0387] In some embodiments, a nucleic acid molecule as described herein (e.g., a genetic element, genetic element construct, or genetic element region) comprises a 5' UTR sequence, e.g., a 5' UTR conserved domain sequence as described herein (e.g., in any of Tables A1, B1, or C1), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

[0388] In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAAC-CACCGAAGTCAGGGCAATTCTGGGCTAGGGX-iCAGTCT (SEQ ID NO: 951), or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGT-GAGTGAAACCACCGAAGTCAGGGCAAT-TCGGGCTAGGGXiCAGTCT (SEQ ID NO: 951), or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences (e.g., substitutions, deletions, or additions) relative thereto. In embodiments, X<sub>1</sub> is A. In embodiments, X<sub>1</sub> is absent.

[0389] In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR of an Alphatorquevirus (e.g., Ring1), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR conserved domain listed in Table A1, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 95% sequence identity to the 5' UTR conserved domain listed in Table A1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 95.775% sequence identity to the 5' UTR conserved domain listed in Table A1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97% sequence identity to the 5' UTR conserved domain listed in Table A1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97.183% sequence identity to the 5' UTR conserved domain listed in Table A1. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTTACACACCGCAGTCAGGGCAAT-TCGGGCTCGGGACTGGC (SEQ ID NO: 952), or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTTACACACCGCAGT-CAAGGGCAATTCTGGGCTCGGGACTGGC (SEQ ID NO: 952), or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences (e.g., substitutions, deletions, or additions) relative thereto.

[0390] In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR of an Betatorquevirus (e.g., Ring2), or a sequence having at least

75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR conserved domain listed in Table B1, or a sequence having at least 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 85% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 87% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 87.324% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 88% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 88.732% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 91% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 91.549% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 92% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 92.958% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 94% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 94.366% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 95% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 95.775% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97.183% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAAGT-CAAGGGCAATTCTGGGCTAGATCAGTCT (SEQ ID NO: 953), or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAAC-ACCGAAGTCAGGGCAATTCTGGGTA-GATCAGTCT (SEQ ID NO: 953), or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences (e.g., substitutions, deletions, or additions) relative thereto.

**[0391]** In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR of a Gammatorquevirus (e.g., Ring4), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR conserved domain listed in Table C1, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97% sequence identity to the 5' UTR conserved domain listed in Table C1. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAGGTCTAGGGGCAAT-TCGGGCTAGGGCAGTCT (SEQ ID NO: 954), or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAAC-CACCGAGGTCTAGGGGCAAT-TCGGGCTAGGGCAGTCT (SEQ ID NO: 954), or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences (e.g., substitutions, deletions, or additions) relative thereto.

**[0392]** In some embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to an Anellovirus 5' UTR sequence, e.g., a nucleic acid sequence shown in Table 38. In some embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence of the Consensus 5' UTR sequence shown in Table 38, wherein X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, and X<sub>5</sub> are each independently any nucleotide, e.g., wherein X<sub>1</sub>=G or T, X<sub>2</sub>=C or A, X<sub>3</sub>=G or A, X<sub>4</sub>=T or C, and X<sub>5</sub>=A, C, or T. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Consensus 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the exemplary TTV 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-CT30F 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-HD23a 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the

genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-JA20 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-TJN02 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-th8 5' UTR sequence shown in Table 38.

**[0393]** In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Consensus 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 1 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 2 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 3 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 4 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 5 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 6 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 7 5' UTR sequence shown in Table 38.

TABLE 38

Exemplary 5' UTR sequences from Anelloviruses		
Source	Sequence	SEQ ID NO:
Consensus	CGGGTGCAGX <sub>1</sub> , AGGTGAGTTACACACCGX <sub>2</sub> , AGT CAAGGGCAATTCTGGCTCX <sub>3</sub> , GGACTGCCGGG CX <sub>4</sub> X <sub>5</sub> TGGG X <sub>1</sub> = G or T X <sub>2</sub> = C or A X <sub>3</sub> = G or A X <sub>4</sub> = T or C X <sub>5</sub> = A, C, or T	105
Exemplary TTV Sequence	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT WTGGG	106
TTV-CT30F	CGGGTGCAGTAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT ATGGG	107
TTV-HD23a	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT CTGGG	108
TTV-JA20	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT TTGGG	109
TTV-TJN02	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT ATGGG	110
TTV-tth8	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCAGGACTGCCGGGCT TTGGG	111
Alphatorquevirus Consensus 5' UTR	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT X <sub>1</sub> X <sub>2</sub> TGGG; wherein X <sub>1</sub> comprises T or C, and wherein X <sub>2</sub> comprises A, C, or T.	112
Alphatorquevirus Clade 1 5' UTR (e.g., TTV-CT30F)	CGGGTGCAGTAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT ATGGG	113
Alphatorquevirus Clade 2 5' UTR (e.g., TTV-P13-1)	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT CGGG	114
Alphatorquevirus Clade 3 5' UTR (e.g., TTV-tth8)	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCAGGACTGCCGGGCT TTGGG	115
Alphatorquevirus Clade 4 5' UTR (e.g., TTV-HD20a)	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGAGGCCAT GGG	116
Alphatorquevirus Clade 5 5' UTR (e.g., TTV-16)	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT CCGGG	117
Alphatorquevirus Clade 6 5' UTR (e.g., TTV-TJN02)	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT ATGGG	118
Alphatorquevirus Clade 7 5' UTR (e.g., TTV-HD16d)	CGGGTGCAGGAGGTGAGTTACACACCCAGTC AAGGGCAATTCTGGCTCGGGACTGCCGGGCT ATGGG	119

## Identification of 5' UTR Sequences

[0394] In some embodiments, an Anellovirus 5' UTR sequence can be identified within the genome of an Anellovirus (e.g., a putative Anellovirus genome identified, for example, by nucleic acid sequencing techniques, e.g., deep

sequencing techniques). In some embodiments, an Anellovirus 5' UTR sequence is identified by one or both of the following steps:

[0395] (i) Identification of circularization junction point: In some embodiments, a 5' UTR will be positioned near a

circularization junction point of a full-length, circularized Anellovirus genome. A circularization junction point can be identified, for example, by identifying overlapping regions of the sequence. In some embodiments, a overlapping region of the sequence can be trimmed from the sequence to produce a full-length Anellovirus genome sequence that has been circularized. In some embodiments, a genome sequence is circularized in this manner using software. Without wishing to be bound by theory, computationally circularizing a genome may result in the start position for the sequence being oriented in a non-biological. Landmarks within the sequence can be used to re-orient sequences in the proper direction. For example, landmark sequence may include sequences having substantial homology to one or more elements within an Anellovirus genome as described herein (e.g., one or more of a TATA box, cap site, initiator element, transcriptional start site, 5' UTR conserved domain, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, three open-reading frame region, poly(A) signal, or GC-rich region of an Anellovirus, e.g., as described herein).

**[0396]** (ii) Identification of 5' UTR sequence: Once a putative Anellovirus genome sequence has been obtained, the sequence (or portions thereof, e.g., having a length between about 40-50, 50-60, 60-70, 70-80, 80-90, or 90-100 nucleotides) can be compared to one or more Anellovirus 5' UTR sequences (e.g., as described herein) to identify sequences having substantial homology thereto. In some embodiments, a putative Anellovirus 5' UTR region has at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus 5' UTR sequence as described herein.

#### GC-Rich Regions

**[0397]** In some embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a nucleic acid sequence shown in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a GC-rich sequence shown in Table 39.

**[0398]** In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a 36-nucleotide GC-rich sequence as shown in Table 39 (e.g., 36-nucleotide consensus GC-rich region sequence 1, 36-nucleotide consensus GC-rich region sequence 2, TTV Clade 1 36-nucleotide region, TTV Clade 3 36-nucleotide region, TTV Clade 3 isolate GH1 36-nucleotide region, TTV Clade 3 sle1932 36-nucleotide region, TTV Clade 4 ctdc002 36-nucleotide region, TTV Clade 5 36-nucleotide region, TTV Clade 6 36-nucleotide region, or TTV Clade 7 36-nucleotide region). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of a 36-nucleotide GC-rich sequence as shown in Table 39 (e.g., 36-nucleotide consensus GC-rich region sequence 1, 36-nucleotide consensus GC-rich region sequence 2, TTV Clade 1 36-nucleotide region, TTV Clade 3 36-nucleotide region, TTV Clade 3

isolate GH1 36-nucleotide region, TTV Clade 3 sle1932 36-nucleotide region, TTV Clade 4 ctdc002 36-nucleotide region, TTV Clade 5 36-nucleotide region, TTV Clade 6 36-nucleotide region, or TTV Clade 7 36-nucleotide region).

**[0399]** In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to an Alphatorquevirus GC-rich region sequence, e.g., selected from TTV-CT30F, TTV-P13-1, TTV-th8, TTV-HD20a, TTV-16, TTV-TJN02, or TTV-HD16d, e.g., as listed in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence comprising at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 104, 105, 108, 110, 111, 115, 120, 122, 130, 140, 145, 150, 155, or 156 consecutive nucleotides of an Alphatorquevirus GC-rich region sequence, e.g., selected from TTV-CT30F, TTV-P13-1, TTV-th8, TTV-HD20a, TTV-16, TTV-TJN02, or TTV-HD16d, e.g., as listed in Table 39.

**[0400]** In embodiments, the 36-nucleotide GC-rich sequence is selected from:

(i) (SEQ ID NO: 160)  
CGCGCTGCGCGCCGCCAGTAGGGGGAGCCATGC,

(ii) (SEQ ID NO: 164)  
CGCCTX<sub>1</sub>CGCGCGCGCCGGGGCTGCGCCCCCCC,

**[0401]** wherein X<sub>1</sub> is selected from T, G, or A;

(iii) (SEQ ID NO: 165)  
GCGCTTCGCGCGCCGCCACTAGGGGGCGTTGCGCG;

(iv) (SEQ ID NO: 166)  
GCGCTGCGCGCCGCCAGTAGGGGGCGCAATGCG;

(v) (SEQ ID NO: 167)  
GCGCTGCGCGCGCCGGCCCCGGGGAGGCATTGCCT;

(vi) (SEQ ID NO: 168)  
GCGCTGCGCGCGCGCCGGGGGGCGCCAGCGCCC;

(vii) (SEQ ID NO: 169)  
GCGCTTCGCGCGCGCGCCGGGGGCTCGCCCCCCC;

(viii) (SEQ ID NO: 170)  
GCGCTTCGCGCGCGCGCCGGGGGCTCGCCCCCCC;

(ix) (SEQ ID NO: 171)  
GCGCTACCGCGCGCGCGCCGGGGGCTCGCCCCCCC;  
or

(x) (SEQ ID NO: 172)  
GCGCTACCGCGCGCGCCGGGGGCTCTGCCCCCCC.

In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises the nucleic acid sequence CGCGCTGCGCGCGCCAGTAGGGGGAGCCATGC (SEQ ID NO: 160).

**[0402]** In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence of the Consensus GC-rich sequence shown in Table 39, wherein X<sub>1</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>7</sub>, X<sub>12</sub>, X<sub>13</sub>, X<sub>14</sub>, X<sub>15</sub>, X<sub>20</sub>, X<sub>21</sub>, X<sub>22</sub>, X<sub>26</sub>, X<sub>29</sub>, X<sub>30</sub>, and X<sub>33</sub> are each independently any nucleotide and wherein X<sub>2</sub>, X<sub>3</sub>, X<sub>8</sub>, X<sub>9</sub>, X<sub>10</sub>, X<sub>11</sub>, X<sub>16</sub>, X<sub>17</sub>, X<sub>18</sub>, X<sub>19</sub>, X<sub>23</sub>, X<sub>24</sub>, X<sub>25</sub>, X<sub>27</sub>, X<sub>28</sub>, X<sub>31</sub>, X<sub>32</sub>, and X<sub>34</sub> are each independently absent or any nucleotide. In some embodiments, one or more of (e.g., all of) X<sub>1</sub> through X<sub>34</sub> are each independently the nucleotide (or absent) specified in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to an exemplary TTV GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, or any combination thereof, e.g., Fragments 1-3 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a TTV-CT30F GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, Fragment 4, Fragment 5, Fragment 6, Fragment 7, Fragment 8, or any combination thereof, e.g., Fragments 1-7 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a TTV-HD23a GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, Fragment 4, Fragment 5, Fragment 6, or any combination thereof, e.g., Fragments 1-6 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises

a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a TTV-JA20 GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, or any combination thereof, e.g., Fragments 1 and 2 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a TTV-TJN02 GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, Fragment 4, Fragment 5, Fragment 6, Fragment 7, Fragment 8, or any combination thereof, e.g., Fragments 1-8 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a TTV-tth8 GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, Fragment 4, Fragment 5, Fragment 6, Fragment 7, Fragment 8, Fragment 9, or any combination thereof, e.g., Fragments 1-6 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to Fragment 7 shown in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to Fragment 8 shown in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to Fragment 9 shown in Table 39.

TABLE 39

Exemplary GC-rich sequences from Anelloviruses		
Source	Sequence	SEQ ID NO:
Consensus	CGCGGGX <sub>1</sub> GGX <sub>2</sub> GX <sub>3</sub> X <sub>4</sub> X <sub>5</sub> C GCGCTX <sub>6</sub> C GCGC GX <sub>7</sub> X <sub>8</sub> X <sub>9</sub> X <sub>10</sub> CX <sub>11</sub> X <sub>12</sub> X <sub>13</sub> X <sub>14</sub> GGGGX <sub>15</sub> X <sub>16</sub> X <sub>17</sub> X <sub>18</sub> X <sub>19</sub> X <sub>20</sub> X <sub>21</sub> GCX <sub>22</sub> X <sub>23</sub> X <sub>24</sub> X <sub>25</sub> CCCCCCCCX <sub>26</sub> C GCGC ATX <sub>27</sub> X <sub>28</sub> GCX <sub>29</sub> CGGGX <sub>30</sub> CCCCCCCCCX <sub>31</sub> X <sub>32</sub> X <sub>33</sub> GGGGGGCTCCGX <sub>34</sub> CCCCCGGGCCCC	120
	X <sub>1</sub> = G or C X <sub>2</sub> = G, C, or absent X <sub>3</sub> = C or absent X <sub>4</sub> = G or C X <sub>5</sub> = G or C X <sub>6</sub> = T, G, or A X <sub>7</sub> = G or C X <sub>8</sub> = G or absent X <sub>9</sub> = C or absent X <sub>10</sub> = C or absent X <sub>11</sub> = G, A, or absent X <sub>12</sub> = G or C X <sub>13</sub> = C or T X <sub>14</sub> = G or A X <sub>15</sub> = G or A X <sub>16</sub> = A, G, T, or absent X <sub>17</sub> = G, C, or absent X <sub>18</sub> = G, C, or absent X <sub>19</sub> = C, A, or absent X <sub>20</sub> = C or A X <sub>21</sub> = T or A X <sub>22</sub> = G or C	

TABLE 39-continued

Exemplary GC-rich sequences from Anelloviruses		
Source	Sequence	SEQ ID NO:
	X <sub>23</sub> = G, T, or absent X <sub>24</sub> = C or absent X <sub>25</sub> = G, C, or absent X <sub>26</sub> = G or C X <sub>27</sub> = G or absent X <sub>28</sub> = C or absent X <sub>29</sub> = G or A X <sub>30</sub> = G or T X <sub>31</sub> = C, T, or absent X <sub>32</sub> = G, C, A, or absent X <sub>33</sub> = G or C X <sub>34</sub> = C or absent	
Exemplary TTV Sequence	Full sequence      GCCGCCGCGCGGGGGGGNNNSGCGCGCT DCGGCGCSNNNCRCCRGGGGNNNNCWG CSNCNCNCNCNCGGCGCATGCGGGGKCC CCCCCCCNNNGGGGGCTCCGCCCCCCCGGC CCCCCCCCGTGCTAAACCCACCGCGCATGC GCGACCACGCCCGCCGCC Fragment 1      GCCGCCGCGCGGGGGGGNNNSGCGCGCT DCGGCGCSNNNCRCCRGGGGNNNNCWG CSNCNCNCNCNCGGCGCAT Fragment 2      GCGGGGKCCCCCCCCNNNGGGGGCTC CG Fragment 3      CCCCCCGGCCCCCCCCCGTGCCTAAACCCAC CGCGCATGCGCGACCACGCCCGCCGCC	121
TTV-CT30F	Full sequence      GCGGGGG-GGGGGCG-GCCGCG- TTCCGGCGCCGCCAACCAAGGGGTG-- CTGCG-CGCCCCCCCCCGCGCAT GCGGGGGCCCCCCCC GGGGGGGCTCCGCCCCCCCGGCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCGCGCC Fragment 1      GCGGGG Fragment 2      GGGGGCG Fragment 3      GCCCG Fragment 4      TTCCGGCGCCGCCAACCAAGGGGTG Fragment 5      CTGCG Fragment 6      CGCCCCCCCCCGCGCAT Fragment 7      GCGGGGGCCCCCCCC Fragment 8      GGGGGGGCTCCGCCCCCCCGGCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCGCGCC	125
TTV-HD23a	Full sequence      CGGGGGCGGGCGGCG- CGCCCGCTGCGCGCGC--- CGCGGGGGGGCGCCAGCG- CCCCCCCCCGCGCAT GCACGGGTCCCCCCCCCACGGGGGCTCC GCCCGCGCCCCCCCC Fragment 1      CGGGGGCGGGCG Fragment 2      CGCCCGCTGCGCGCG Fragment 3      CGCGGGGGGGCGCCAGCG Fragment 4      CCCCCCCCCCGCGCAT Fragment 5      GCACGGGTCCCCCCCCCACGGGGGCTCC G Fragment 6      GCCCGCGCCCCCCCC	134
TTV-JA20	Full sequence      CCGTCGGCGGGGGGGCGCGCTGCGCG CGCGGCC- CGGGGGGAGGCACAGCCTCCCCCCCCCGCG CGCATGCGCGGGTCCCCCCCCCTCCGG GGGCTCCGCCCGCCCCCGGCC Fragment 1      CCGTCGGCGGGGGGGCGCGCTGCGCG CGCGGCC Fragment 2      CGGGGGGAGGCACAGCCTCCCCCCCCCGCG CGCATGCGCGGGTCCCCCCCCCTCCGG GGGCTCCGCCCGCCCCCGGCC	141

TABLE 39-continued

Exemplary GC-rich sequences from Anelloviruses			
Source	Sequence	SEQ ID NO:	
TTV-TJN02	Full sequence	CGGCGGGGGCG-CGGCGCTACGGGGCGG- -CGCGGGGGG---CTGCCGC- CCCCCCCCCGCGCAT GCGGGGGGCCCCCCCC GCGGGGGGCTCCG CCCCCCGGCCCC CGGCGCGCG Fragment 1 CGCGCGCTACGCGCGCG 145 Fragment 2 CGCGCGCTACGCGCGCG 146 Fragment 3 CGCGGGGGG 147 Fragment 4 CTGCCGC 148 Fragment 5 CCCCCCCCCCGCGCAT 149 Fragment 6 GCGGGGGGCCCCCCCC 150 Fragment 7 GCGGGGGGCTCCG 151 Fragment 8 CCCCCCGGCCCC 152	144
	Fragment 1	CGGCGCGCG 145	
	Fragment 2	CGCGCGCTACGCGCGCG 146	
	Fragment 3	CGCGGGGGG 147	
	Fragment 4	CTGCCGC 148	
	Fragment 5	CCCCCCCCCGCGCAT 149	
	Fragment 6	GCGGGGGGCCCCCCCC 150	
	Fragment 7	GCGGGGGGCTCCG 151	
	Fragment 8	CCCCCGGCCCC 152	
TTV-tth8	Full sequence	GCCGCCGCGGGGGGGGGGG- GCGCGCGCTGCGCGCGCCAGTAGG GGGAGCCATGCG- - CCCCCCCCCCGCGCAT GCGCGGGGGCCCCCCCC GCGGGGGGCTCCG CCCCCGGCCCCCCCC GCGCCCGCGGGGGGGGG 154 GCGCGCGCTGCGCGCGCCAGTAGG 155	153
	Fragment 1	CCCCCCCCCGCGCAT 154	
	Fragment 2	GCGCGGGGGCCCCCCCC 155	
	Fragment 3	GGGAGCCATGCG 156	
	Fragment 4	CCCCCCCCCGCGCAT 156	
	Fragment 5	GCGGGGGGCTCCG 157	
	Fragment 6	CCCCCGGCCCCCCCC 158	
	Fragment 7	GCGCTGCGCGCGCCAGTAGGGGA 159	
	Fragment 8	GCGCTGCGCGCGCCAGTAGGGGA 160	
	Fragment 9	GCCATGC CCGCATCTTAAGTAGTTGAGGGGACGGT GGCTTGAGTCAAAGGTACCATCAGCAC ACCTACTCAAATGGTGG 161	
Additional GC-rich Sequences	36-nucleotide consensus GC- rich region sequence 1	GGCGCTGCGCGCGCCAGTAGGGGA GCCATGC 163	
	36-nucleotide region consensus sequence 2	GCGCTX1CGCGCGCGGCCGGGGGCTGCG CCCCCCC, wherein X <sub>i</sub> is selected from T, G, or A 164	
	TTV Clade 1 36-nucleotide region	GCGCTTCGCGCGCCGCCACTAGGGGGCGT TGCGCG 165	
	TTV Clade 3 36-nucleotide region	GCGCTGCGCGCGCCGCCAGTAGGGGGCG CAATGCG 166	
	TTV Clade 3 isolate GH1 36- nucleotide region	GCGCTGCGCGCGGCCGGCCCCGGGGAGGC ATTGCCT 167	
	TTV Clade 3 slel932 36- nucleotide region	GCGCTGCGCGCGCGCCGGGGGGCGCC AGCGCCC 168	
	TTV Clade 4 ctdc002 36- nucleotide region	GCGCTTCGCGCGCGCGCCGGGGCTCCGC CCCCCC 169	
	TTV Clade 5 36-nucleotide region	GCGCTTCGCGCGCGCGCCGGGGCTGCGC CCCCCC 170	
	TTV Clade 6 36-nucleotide region	GCGCTACGCGCGCGCGCCGGGGCTGCG CCCCCC 171	

TABLE 39-continued

## Effectors

**[0403]** In some embodiments, the genetic element may include one or more sequences that are or encode an effector, e.g., a functional effector, e.g., an endogenous effector or an exogenous effector, e.g., a therapeutic polypeptide or nucleic acid, e.g., cytotoxic or cytolytic RNA or protein. In some embodiments, the functional nucleic acid is a non-coding RNA. In some embodiments, the functional nucleic acid is a coding RNA. The effector may modulate a biological activity, for example increasing or decreasing enzymatic activity, gene expression, cell signaling, and cellular or organ function. Effector activities may also include binding regulatory proteins to modulate activity of the regulator, such as transcription or translation. Effector activities also may include activator or inhibitor functions. For example, the effector may induce enzymatic activity by triggering increased substrate affinity in an enzyme, e.g., fructose 2,6-bisphosphate activates phosphofructokinase 1 and increases the rate of glycolysis in response to the insulin. In another example, the effector may inhibit substrate binding to a receptor and inhibit its activation, e.g., naltrexone and naloxone bind opioid receptors without activating them and block the receptors' ability to bind opioids. Effector activities may also include modulating protein stability/degradation and/or transcript stability/degradation. For example, proteins may be targeted for degradation by the polypeptide co-factor, ubiquitin, onto proteins to mark them for degra-

dation. In another example, the effector inhibits enzymatic activity by blocking the enzyme's active site, e.g., methotrexate is a structural analog of tetrahydrofolate, a coenzyme for the enzyme dihydrofolate reductase that binds to dihydrofolate reductase 1000-fold more tightly than the natural substrate and inhibits nucleotide base synthesis.

**[0404]** In some embodiments, the sequence encoding an effector comprises 100-2000, 100-1000, 100-500, 100-200, 200-2000, 200-1000, 200-500, 500-1000, 500-2000, or 1000-2000 nucleotides. In some embodiments, the effector is a nucleic acid or protein payload, e.g., as described herein.

Regulatory Nucleic Acids

**[0405]** In some embodiments, the effector is a regulatory nucleic acid. Regulatory nucleic acids modify expression of an endogenous gene and/or an exogenous gene. In one embodiment, the regulatory nucleic acid targets a host gene. The regulatory nucleic acids may include, but are not limited to, a nucleic acid that hybridizes to an endogenous gene (e.g., miRNA, siRNA, mRNA, lncRNA, RNA, DNA, an antisense RNA, gRNA as described herein elsewhere), nucleic acid that hybridizes to an exogenous nucleic acid such as a viral DNA or RNA, nucleic acid that hybridizes to an RNA, nucleic acid that interferes with gene transcription, nucleic acid that interferes with RNA translation, nucleic acid that stabilizes RNA or destabilizes RNA such as through targeting for degradation, and nucleic acid that modulates a DNA or RNA binding factor. In embodiments, the regulatory nucleic acid encodes an miRNA. In some

embodiments, the regulatory nucleic acid is endogenous to a wild-type Anellovirus. In some embodiments, the regulatory nucleic acid is exogenous to a wild-type Anellovirus. [0406] In some embodiments, the regulatory nucleic acid comprises RNA or RNA-like structures typically containing 5-500 base pairs (depending on the specific RNA structure, e.g., miRNA 5-30 bps, lncRNA 200-500 bps) and may have a nucleobase sequence identical (or complementary) or nearly identical (or substantially complementary) to a coding sequence in an expressed target gene within the cell, or a sequence encoding an expressed target gene within the cell.

[0407] In some embodiments, the regulatory nucleic acid comprises a nucleic acid sequence, e.g., a guide RNA (gRNA). In some embodiments, the DNA targeting moiety comprises a guide RNA or nucleic acid encoding the guide RNA. A gRNA short synthetic RNA can be composed of a "scaffold" sequence necessary for binding to the incomplete effector moiety and a user-defined ~20 nucleotide targeting sequence for a genomic target. In practice, guide RNA sequences are generally designed to have a length of between 17-24 nucleotides (e.g., 19, 20, or 21 nucleotides) and complementary to the targeted nucleic acid sequence. Custom gRNA generators and algorithms are available commercially for use in the design of effective guide RNAs. Gene editing has also been achieved using a chimeric "single guide RNA" ("sgRNA"), an engineered (synthetic) single RNA molecule that mimics a naturally occurring crRNA-tracrRNA complex and contains both a tracrRNA (for binding the nuclease) and at least one crRNA (to guide the nuclease to the sequence targeted for editing). Chemically modified sgRNAs have also been demonstrated to be effective in genome editing; see, for example, Hendel et al. (2015) *Nature Biotechnol.*, 985-991.

[0408] The regulatory nucleic acid comprises a gRNA that recognizes specific DNA sequences (e.g., sequences adjacent to or within a promoter, enhancer, silencer, or repressor of a gene).

[0409] Certain regulatory nucleic acids can inhibit gene expression through the biological process of RNA interference (RNAi). RNAi molecules comprise RNA or RNA-like structures typically containing 15-50 base pairs (such as about 18-25 base pairs) and having a nucleobase sequence identical (complementary) or nearly identical (substantially complementary) to a coding sequence in an expressed target gene within the cell. RNAi molecules include, but are not limited to: short interfering RNAs (siRNAs), double-strand RNAs (dsRNA), micro RNAs (miRNAs), short hairpin RNAs (shRNA), meroduplexes, and dicer substrates (U.S. Pat. Nos. 8,084,599 8,349,809 and 8,513,207).

[0410] Long non-coding RNAs (lncRNA) are defined as non-protein coding transcripts longer than 100 nucleotides. This somewhat arbitrary limit distinguishes lncRNAs from small regulatory RNAs such as microRNAs (miRNAs), short interfering RNAs (siRNAs), and other short RNAs. In general, the majority (~78%) of lncRNAs are characterized as tissue-specific. Divergent lncRNAs that are transcribed in the opposite direction to nearby protein-coding genes (comprise a significant proportion ~20% of total lncRNAs in mammalian genomes) may possibly regulate the transcription of the nearby gene.

[0411] The genetic element may encode regulatory nucleic acids with a sequence substantially complementary, or fully complementary, to all or a fragment of an endogenous gene

or gene product (e.g., mRNA). The regulatory nucleic acids may complement sequences at the boundary between introns and exons to prevent the maturation of newly-generated nuclear RNA transcripts of specific genes into mRNA for transcription. The regulatory nucleic acids that are complementary to specific genes can hybridize with the mRNA for that gene and prevent its translation. The antisense regulatory nucleic acid can be DNA, RNA, or a derivative or hybrid thereof.

[0412] The length of the regulatory nucleic acid that hybridizes to the transcript of interest may be between 5 to 30 nucleotides, between about 10 to 30 nucleotides, or about 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more nucleotides. The degree of identity of the regulatory nucleic acid to the targeted transcript should be at least 75%, at least 80%, at least 85%, at least 90%, or at least 95%.

[0413] The genetic element may encode a regulatory nucleic acid, e.g., a micro RNA (miRNA) molecule identical to about 5 to about 25 contiguous nucleotides of a target gene. In some embodiments, the miRNA sequence targets a mRNA and commences with the dinucleotide AA, comprises a GC-content of about 30-70% (about 30-60%, about 40-60%, or about 45%-55%), and does not have a high percentage identity to any nucleotide sequence other than the target in the genome of the mammal in which it is to be introduced, for example as determined by standard BLAST search.

[0414] siRNAs and shRNAs resemble intermediates in the processing pathway of the endogenous microRNA (miRNA) genes (Bartel, *Cell* 116:281-297, 2004). In some embodiments, siRNAs can function as miRNAs and vice versa (Zeng et al., *Mol Cell* 9:1327-1333, 2002; Doench et al., *Genes Dev* 17:438-442, 2003). MicroRNAs, like siRNAs, use RISC to downregulate target genes, but unlike siRNAs, most animal miRNAs do not cleave the mRNA. Instead, miRNAs reduce protein output through translational suppression or polyA removal and mRNA degradation (Wu et al., *Proc Natl Acad Sci USA* 103:4034-4039, 2006). Known miRNA binding sites are within mRNA 3' UTRs; miRNAs seem to target sites with near-perfect complementarity to nucleotides 2-8 from the miRNA's 5' end (Rajewsky, *Nat Genet* 38 Suppl:S8-13, 2006; Lim et al., *Nature* 433:769-773, 2005). This region is known as the seed region. Because siRNAs and miRNAs are interchangeable, exogenous siRNAs downregulate mRNAs with seed complementarity to the siRNA (Birmingham et al., *Nat Methods* 3:199-204, 2006. Multiple target sites within a 3' UTR give stronger downregulation (Doench et al., *Genes Dev* 17:438-442, 2003).

[0415] Lists of known miRNA sequences can be found in databases maintained by research organizations, such as Wellcome Trust Sanger Institute, Penn Center for Bioinformatics, Memorial Sloan Kettering Cancer Center, and European Molecule Biology Laboratory, among others. Known effective siRNA sequences and cognate binding sites are also well represented in the relevant literature. RNAi molecules are readily designed and produced by technologies known in the art. In addition, there are computational tools that increase the chance of finding effective and specific sequence motifs (Lagana et al., *Methods Mol. Bio.*, 2015, 1269:393-412).

[0416] The regulatory nucleic acid may modulate expression of RNA encoded by a gene. Because multiple genes can

share some degree of sequence homology with each other, in some embodiments, the regulatory nucleic acid can be designed to target a class of genes with sufficient sequence homology. In some embodiments, the regulatory nucleic acid can contain a sequence that has complementarity to sequences that are shared amongst different gene targets or are unique for a specific gene target. In some embodiments, the regulatory nucleic acid can be designed to target conserved regions of an RNA sequence having homology between several genes thereby targeting several genes in a gene family (e.g., different gene isoforms, splice variants, mutant genes, etc.). In some embodiments, the regulatory nucleic acid can be designed to target a sequence that is unique to a specific RNA sequence of a single gene.

**[0417]** In some embodiments, the genetic element may include one or more sequences that encode regulatory nucleic acids that modulate expression of one or more genes.

**[0418]** In one embodiment, the gRNA described elsewhere herein are used as part of a CRISPR system for gene editing. For the purposes of gene editing, the anellovector may be designed to include one or multiple guide RNA sequences corresponding to a desired target DNA sequence; see, for example, Cong et al. (2013) *Science*, 339:819-823; Ran et al. (2013) *Nature Protocols*, 8:2281-2308. At least about 16 or 17 nucleotides of gRNA sequence generally allow for Cas9-mediated DNA cleavage to occur; for Cpf1 at least about 16 nucleotides of gRNA sequence is needed to achieve detectable DNA cleavage.

#### Therapeutic Effectors (e.g., Peptides or Polypeptides)

**[0419]** In some embodiments, the genetic element comprises a therapeutic expression sequence, e.g., a sequence that encodes a therapeutic peptide or polypeptide, e.g., an intracellular peptide or intracellular polypeptide, a secreted polypeptide, or a protein replacement therapeutic. In some embodiments, the genetic element includes a sequence encoding a protein e.g., a therapeutic protein. Some examples of therapeutic proteins may include, but are not limited to, a hormone, a cytokine, an enzyme, an antibody (e.g., one or a plurality of polypeptides encoding at least a heavy chain or a light chain), a transcription factor, a receptor (e.g., a membrane receptor), a ligand, a membrane transporter, a secreted protein, a peptide, a carrier protein, a structural protein, a nuclease, or a component thereof.

**[0420]** In some embodiments, the genetic element includes a sequence encoding a peptide e.g., a therapeutic peptide. The peptides may be linear or branched. The peptide has a length from about 5 to about 500 amino acids, about 15 to about 400 amino acids, about 20 to about 325 amino acids, about 25 to about 250 amino acids, about 50 to about 200 amino acids, or any range there between.

**[0421]** In some embodiments, the polypeptide encoded by the therapeutic expression sequence may be a functional variant or fragment thereof of any of the above, e.g., a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% identity to a protein sequence which disclosed in a table herein by reference to its UniProt ID.

**[0422]** In some embodiments, the therapeutic expression sequence may encode an antibody or antibody fragment that binds any of the above, e.g., an antibody against a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% identity to a protein sequence which disclosed in a table herein by reference to its UniProt ID. The term "antibody" herein is used in the broadest sense and encompasses various antibody structures, including but not limited to monoclonal antibodies, polyclonal antibodies, multispecific antibodies (e.g., bispecific antibodies), and antibody fragments so long as they exhibit the desired antigen-binding activity. An "antibody fragment" refers to a molecule that includes at least one heavy chain or light chain and binds an antigen. Examples of antibody fragments include but are not limited to Fv, Fab, Fab', Fab'-SH, F(ab')<sub>2</sub>; diabodies; linear antibodies; single-chain antibody molecules (e.g. scFv); and multispecific antibodies formed from antibody fragments.

**Exemplary Intracellular Polypeptide Effectors** In some embodiments, the effector comprises a cytosolic polypeptide or cytosolic peptide. In some embodiments, the effector comprises cytosolic peptide is a DPP-4 inhibitor, an activator of GLP-1 signaling, or an inhibitor of neutrophil elastase. In some embodiments, the effector increases the level or activity of a growth factor or receptor thereof (e.g., an FGF receptor, e.g., FGFR3). In some embodiments, the effector comprises an inhibitor of n-myc interacting protein activity (e.g., an n-myc interacting protein inhibitor); an inhibitor of EGFR activity (e.g., an EGFR inhibitor); an inhibitor of IDH1 and/or IDH2 activity (e.g., an IDH1 inhibitor and/or an IDH2 inhibitor); an inhibitor of LRP5 and/or DKK2 activity (e.g., an LRP5 and/or DKK2 inhibitor); an inhibitor of KRAS activity; an activator of HTT activity; or inhibitor of DPP-4 activity (e.g., a DPP-4 inhibitor).

**[0423]** In some embodiments, the effector comprises a regulatory intracellular polypeptide. In some embodiments, the regulatory intracellular polypeptide binds one or more molecule (e.g., protein or nucleic acid) endogenous to the target cell. In some embodiments, the regulatory intracellular polypeptide increases the level or activity of one or more molecule (e.g., protein or nucleic acid) endogenous to the target cell. In some embodiments, the regulatory intracellular polypeptide decreases the level or activity of one or more molecule (e.g., protein or nucleic acid) endogenous to the target cell.

#### Exemplary Secreted Polypeptide Effectors

**[0424]** Exemplary secreted therapeutics are described herein, e.g., in the tables below.

TABLE 50

Exemplary cytokines and cytokine receptors			
Cytokine	Cytokine receptor(s)	Entrez Gene ID	UniProt ID
IL-1 $\alpha$ , IL-1 $\beta$ , or a heterodimer thereof IL-1Ra	IL-1 type 1 receptor, IL-1 type 2 receptor IL-1 type 1 receptor, IL-1 type 2 receptor	3552, 3553 3454, 3455	P01583, P01584 P17181, P48551

TABLE 50-continued

Cytokine	Cytokine receptor(s)	Entrez Gene ID	UniProt ID
IL-2	IL-2R	3558	P60568
IL-3	IL-3 receptor $\alpha + \beta + \gamma$ (CD 131)	3562	P08700
IL-4	IL-4R type I, IL-4R type II	3565	P05112
IL-5	IL-5R	3567	P05113
IL-6	IL-6R (sIL-6R) gp130	3569	P05231
IL-7	IL-7R and sIL-7R	3574	P13232
IL-8	CXCR1 and CXCR2	3576	P10145
IL-9	IL-9R	3578	P15248
IL-10	IL-10R1/IL-10R2 complex	3586	P22301
IL-11	IL-11R $\alpha$ 1 gp130	3589	P20809
IL-12 (e.g., p35, p40, or a heterodimer thereof)	IL-12R $\beta$ 1 and IL-12R $\beta$ 2	3593, 3592	P29459, P29460
IL-13	IL-13R $\alpha$ 1 and IL-13R $\alpha$ 2	3596	P35225
IL-14	IL-14R	30685	P40222
IL-15	IL-15R	3600	P40933
IL-16	CD4	3603	Q14005
IL-17A	IL-17RA	3605	Q16552
IL-17B	IL-17RB	27190	Q9UHF5
IL-17C	IL-17RA to IL-17RE	27189	Q9P0M4
e	SEF	53342	Q8TAD2
IL-17F	IL-17RA, IL-17RC	112744	Q96PD4
IL-18	IL-18 receptor	3606	Q14116
IL-19	IL-20R1/IL-20R2	29949	Q9UHD0
IL-20	L-20R1/IL-20R2 and IL-22R1/IL-20R2	50604	Q9NYY1
IL-21	IL-21R	59067	Q9HBE4
IL-22	IL-22R	50616	Q9GZX6
IL-23 (e.g., p19, p40, or a heterodimer thereof)	IL-23R	51561	Q9NPF7
IL-24	IL-20R1/IL-20R2 and IL-22R1/IL-20R2	11009	Q13007
IL-25	IL-17RA and IL-17RB	64806	Q9H293
IL-26	IL-10R2 chain and IL-20R1 chain	55801	Q9NPH9
IL-27 (e.g., p28, EBI3, or a heterodimer thereof)	WSX-1 and gp130	246778	Q8NEV9
IL-28A, IL-28B, and IL-29	IL-28R1/IL-10R2	282617, 282618	Q8IZI9, Q8IU54
IL-30	IL-6R/gp130	246778	Q8NEV9
IL-31	IL-31RA/OSMR $\beta$	386653	Q6EBC2
IL-32		9235	P24001
IL-33	ST2	90865	O95760
IL-34	Colony-stimulating factor 1 receptor	146433	Q6ZMJ4
IL-35 (e.g., p35, EBI3, or a heterodimer thereof)	IL-12R $\beta$ 2/gp130; IL-12R $\beta$ 2/IL-12R $\beta$ 2; gp130/gp130	10148	Q14213
IL-36	IL-36Ra	27179	Q9UHA7
IL-37	IL-18R $\alpha$ and IL-18BP	27178	Q9NZH6
IL-38	IL-1R1, IL-36R	84639	Q8WWZ1
IFN- $\alpha$	IFNAR	3454	P17181
IFN- $\beta$	IFNAR	3454	P17181
IFN- $\gamma$	IFNGR1/IFNGR2	3459	P15260
TGF- $\beta$	T $\beta$ R-I and T $\beta$ R-II	7046, 7048	P36897, P37173
TNF- $\alpha$	TNFR1, TNFR2	7132, 7133	P19438, P20333

**[0425]** In some embodiments, an effector described herein comprises a cytokine of Table 50, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 96%, 98%, 99% sequence identity to an amino acid sequence listed in Table 50 by reference to its UniProt ID. In some embodiments, the functional variant binds to the corresponding cytokine receptor with a Kd of no more than 10%, 20%, 30%, 40%, or 50% higher or lower than the Kd of the corresponding wild-type cytokine for the same receptor under the same conditions. In some embodiments, the effector comprises a fusion protein comprising a first region

(e.g., a cytokine polypeptide of Table 50 or a functional variant or fragment thereof) and a second, heterologous region. In some embodiments, the first region is a first cytokine polypeptide of Table 50. In some embodiments, the second region is a second cytokine polypeptide of Table 50, wherein the first and second cytokine polypeptides form a cytokine heterodimer with each other in a wild-type cell. In some embodiments, the polypeptide of Table 50 or functional variant thereof comprises a signal sequence, e.g., a signal sequence that is endogenous to the effector, or a heterologous signal sequence. In some embodiments, an anellovector encoding a cytokine of Table 50, or a functional variant thereof, is used for the treatment of a disease or disorder described herein.

**[0426]** In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a cytokine of Table 50. In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a cytokine receptor of Table 50. In some embodiments, the antibody molecule comprises a signal sequence.

**[0427]** Exemplary cytokines and cytokine receptors are described, e.g., in Akdis et al., "Interleukins (from IL-1 to IL-38), interferons, transforming growth factor  $\beta$ , and TNF- $\alpha$ : Receptors, functions, and roles in diseases" October 2016 Volume 138, Issue 4, Pages 984-1010, which is herein incorporated by reference in its entirety, including Table I therein.

TABLE 51

Exemplary polypeptide hormones and receptors			
Hormone	Receptor	Entrez Gene ID	UniProt ID
Natriuretic Peptide, e.g., Atrial Natriuretic Peptide (ANP)	NPRA, NPRB, NRPC	4878	P01160
Brain Natriuretic Peptide (BNP)	NPRA, NPRB	4879	P16860
C-type natriuretic peptide (CNP)	NPRB	4880	P23582
Growth hormone (GH)	GHR	2690	P10912
Human growth hormone (hGH)	hGH receptor (human GHR)	2690	P10912
Prolactin (PRL)	PRLR	5617	P01236
Thyroid-stimulating hormone (TSH)	TSH receptor	7253	P16473
Adrenocorticotrophic hormone (ACTH)	ACTH receptor	5443	P01189
Follicle-stimulating hormone (FSH)	FSHR	2492	P23945
Luteinizing hormone (LH)	LHR	3973	P22888
Antidiuretic hormone (ADH)	Vasopressin receptors, e.g., V2; AVPR1A; AVPR1B; AVPR3; AVPR2	554	P30518
Oxytocin	OXTR	5020	P01178
Calcitonin	Calcitonin receptor (CT)	796	P01258
Parathyroid hormone (PTH)	PTHR1 and PTH2R	5741	P01270
Insulin	Insulin receptor (IR)	3630	P01308
Glucagon	Glucagon receptor	2641	P01275

**[0428]** In some embodiments, an effector described herein comprises a hormone of Table 51, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid

sequence listed in Table 51 by reference to its UniProt ID. In some embodiments, the functional variant binds to the corresponding receptor with a Kd of no more than 10%, 20%, 30%, 40%, or 50% higher than the Kd of the corresponding wild-type hormone for the same receptor under the same conditions. In some embodiments, the polypeptide of Table 51 or functional variant thereof comprises a signal sequence, e.g., a signal sequence that is endogenous to the effector, or a heterologous signal sequence. In some embodiments, an anellovector encoding a hormone of Table 51, or a functional variant thereof, is used for the treatment of a disease or disorder described herein.

TABLE 52

Exemplary growth factors			
Growth Factor	Entrez Gene ID	UniProt ID	
<u>PDGF family</u>			
PDGF (e.g., PDGF-1, PDGF-2, or a heterodimer thereof)	PDGF receptor, e.g., PDGFR $\alpha$ , PDGFR $\beta$	5156	P16234
CSF-1	CSF1R	1435	P09603
SCF	CD117	3815	P10721
<u>VEGF family</u>			
VEGF (e.g., isoforms VEGF 121, VEGF 165,	VEGFR-1, VEGFR-2	2321	P17948

**[0429]** In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a hormone of Table 51. In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a hormone receptor of Table 51. In some embodiments, the antibody molecule comprises a signal sequence.

TABLE 52-continued

Exemplary growth factors			
Growth Factor		Entrez Gene ID	UniProt ID
VEGF 189, and VEGF 206)			
VEGF-B	VEGFR-1	2321	P17949
VEGF-C	VEGFR-2 and VEGFR-3	2324	P35916
P1GF	VEGFR-1	5281	Q07326
EGF family			
EGF	EGFR	1950	P01133
TGF- $\alpha$	EGFR	7039	P01135
amphiregulin	EGFR	374	P15514
HB-EGF	EGFR	1839	Q99075
betacellulin	EGFR, ErbB-4	685	P35070
epiregulin	EGFR, ErbB-4	2069	O14944
Heregulin	EGFR, ErbB-4	3084	Q02297
FGF family			
FGF-1, FGF-2, FGF-3, FGF-4, FGF-5, FGF-6, FGF-7, FGF-8, FGF-9	FGFR1, FGFR2, FGFR3, and FGFR4	2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254	P05230, P09038, P11487, P08620, P12034, P10767, P21781, P55075, P31371
Insulin family			
Insulin	IR	3630	P01308
IGF-I	IGF-I receptor, IGF-II receptor	3479	P05019
IGF-II	IGF-II receptor	3481	P01344
HGF family			
HGF	MET receptor	3082	P14210
MSP	RON	4485	P26927
Neurotrophin family			
NGF	LNGFR, trkA	4803	P01138
BDNF	trkB	627	P23560
NT-3	trkA, trkB, trkC	4908	P20783
NT-4	trkA, trkB	4909	P34130
NT-5	trkA, trkB	4909	P34130
Angiopoietin family			
ANGPT1	HPK-6/TEK	284	Q15389
ANGPT2	HPK-6/TEK	285	O15123
ANGPT3	HPK-6/TEK	9068	O95841
ANGPT4	HPK-6/TEK	51378	Q9Y264

[0430] In some embodiments, an effector described herein comprises a growth factor of Table 52, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 52 by reference to its UniProt ID. In some embodiments, the functional variant binds to the corresponding receptor with a Kd of no more than 10%, 20%, 30%, 40%, or 50% higher than the Kd of the corresponding wild-type growth factor for the same receptor under the same conditions. In some embodiments, the polypeptide of Table 52 or functional variant thereof comprises a signal sequence, e.g., a signal sequence that is endogenous to the effector, or a heterologous signal sequence. In some embodiments, an anellovector encoding a growth factor of Table 52, or a functional variant thereof, is used for the treatment of a disease or disorder described herein.

[0431] In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a growth factor of Table 52. In some embodiments, an effector described herein comprises an antibody molecule (e.g., an

scFv) that binds a growth factor receptor of Table 52. In some embodiments, the antibody molecule comprises a signal sequence.

[0432] Exemplary growth factors and growth factor receptors are described, e.g., in Bafico et al., "Classification of Growth Factors and Their Receptors" Holland-Frei Cancer Medicine, 6th edition, which is herein incorporated by reference in its entirety.

TABLE 53

Clotting-associated factors			
Effector	Indication	Entrez Gene ID	UniProt ID
Factor I (fibrinogen)	Afibrinogenemia	2243, 2266, 2244	P02671, P02679, P02675
Factor II	Factor II Deficiency	2147	P00734
Factor IX	Hemophilia B	2158	P00740
Factor V	Owren's disease	2153	P12259
Factor VIII	Hemophilia A	2157	P00451

TABLE 53-continued

Clotting-associated factors			
Effector	Indication	Entrez Gene ID	UniProt ID
Factor X	Stuart-Prower Factor Deficiency	2159	P00742
Factor XI	Hemophilia C	2160	P03951
Factor XIII	Fibrin Stabilizing factor deficiency	2162, 2165	P00488, P05160
vWF	von Willebrand disease	7450	P04275

[0433] In some embodiments, an effector described herein comprises a polypeptide of Table 53, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%,

95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 53 by reference to its UniProt ID. In some embodiments, the functional variant catalyzes the same reaction as the corresponding wild-type protein, e.g., at a rate no less than 10%, 20%, 30%, 40%, or 50% lower than the wild-type protein. In some embodiments, the polypeptide of Table 53 or functional variant thereof comprises a signal sequence, e.g., a signal sequence that is endogenous to the effector, or a heterologous signal sequence. In some embodiments, an anellovector encoding a polypeptide of Table 53, or a functional variant thereof is used for the treatment of a disease or disorder of Table 53.

#### Exemplary Protein Replacement Therapeutics

[0434] Exemplary protein replacement therapeutics are described herein, e.g., in the tables below.

TABLE 54

Exemplary enzymatic effectors and corresponding indications			
Effector	deficiency	Entrez Gene ID	UniProt ID
3-methylcrotonyl-CoA carboxylase	3-methylcrotonyl-CoA carboxylase deficiency	56922, 64087	Q96RQ3, Q9HCC0
Acetyl-CoA-glucosaminide N-acetyltransferase	Mucopolysaccharidosis MPS III (Sanfilippo's syndrome)	138050	Q68CP4
ADAMTS13	Type III-C Thrombotic Thrombocytopenic Purpura	11093	Q76LX8
adenine phosphoribosyltransferase	Adenine phosphoribosyltransferase deficiency	353	P07741
Adenosine deaminase	Adenosine deaminase deficiency	100	P00813
ADP-ribose protein hydrolase	Glutamyl ribose-5-phosphate storage disease	26119, 54936	Q5SW96, Q9NX46
alpha glucosidase	Glycogen storage disease type 2 (Pompe's disease)	2548	P10253
Arginase	Familial hyperarginemia	383, 384	P05089, P78540
Arylsulfatase A	Metachromatic leukodystrophy	410	P15289
Cathepsin K	Pycnodysostosis	1513	P43235
Ceramidase	Farber's disease (lipogranulomatosis)	125981, 340485, 55331	Q8TDN7, Q5QJU3, Q9NUN7
Cystathione B synthase	Homocystinuria	875	P35520
Dolichol-P-mannose synthase	Congenital disorders of N-glycosylation CDG Ie	8813, 54344	O60762, Q9P2X0
Dolicho-P-Glc:Man9GlcNAc2-PP-dolichol glucosyltransferase	Congenital disorders of N-glycosylation CDG Ic	84920	Q5BKT4
Dolicho-P-Man:Man5GlcNAc2-PP-dolichol mannosyltransferase	Congenital disorders of N-glycosylation CDG Id	10195	Q92685
Dolichyl-P-glucose:Glc-1-Man-9-GlcNAc-2-PP-dolichyl- $\alpha$ -3-glucosyltransferase	Congenital disorders of N-glycosylation CDG Ih	79053	Q9BVK2
Dolichyl-P-mannose:Man-7-GlcNAc-2-PP-dolichyl- $\alpha$ -6-mannosyltransferase	Congenital disorders of N-glycosylation CDG Ig	79087	Q9BV10
Factor II	Factor II Deficiency	2147	P00734
Factor IX	Hemophilia B	2158	P00740
Factor V	Owren's disease	2153	P12259
Factor VIII	Hemophilia A	2157	P00451
Factor X	Stuart-Prower Factor Deficiency	2159	P00742

TABLE 54-continued

Exemplary enzymatic effectors and corresponding indications			
Effector	deficiency	Entrez Gene ID	UniProt ID
Factor XI	Hemophilia C	2160	P03951
Factor XIII	Fibrin Stabilizing factor deficiency	2162, 2165	P00488, P05160
Galactosamine-6-sulfate sulfatase	Mucopolysaccharidosis MPS IV (Morquio's syndrome) Type IV-A	2588	P34059
Galactosylceramide $\beta$ -galactosidase	Krabbe's disease	2581	P54803
Ganglioside $\beta$ -galactosidase	GM1 gangliosidosis, generalized	2720	P16278
Ganglioside $\beta$ -galactosidase	GM2 gangliosidosis	2720	P16278
Ganglioside $\beta$ -galactosidase	Sphingolipidosis Type I	2720	P16278
Ganglioside $\beta$ -galactosidase	Sphingolipidosis Type II (juvenile type)	2720	P16278
Ganglioside $\beta$ -galactosidase	Sphingolipidosis Type III (adult type)	2720	P16278
Glucosidase I	Congenital disorders of N-glycosylation CDG IIb	2548	P10253
Glucosylceramide $\beta$ -glucosidase	Gaucher's disease	2629	P04062
Heparan-S-sulfate sulfamidase	Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-A	6448	P51688
homogentisate oxidase	Alkaptonuria	3081	Q93099
Hyaluronidase	Mucopolysaccharidosis MPS IX (hyaluronidase deficiency)	3373, 8692, 8372, 23553	Q12794, Q12891, O43820, Q2M3T9
Iduronate sulfate sulfatase	Mucopolysaccharidosis MPS II (Hunter's syndrome)	3423	P22304
Lecithin-cholesterol acyltransferase (LCAT)	Complete LCAT deficiency, Fish-eye disease, atherosclerosis, hypercholesterolemia	3931	606967
Lysine oxidase	Glutaric acidemia type I	4015	P28300
Lysosomal acid lipase	Cholesteryl ester storage disease (CESD)	3988	P38571
Lysosomal acid lipase	Lysosomal acid lipase deficiency	3988	P38571
lysosomal acid lipase	Wolman's disease	3988	P38571
Lysosomal pepstatin-insensitive peptidase	Ceroid lipofuscinosiis Late infantile form (CLN2, Jansky-Bielschowsky disease)	1200	O14773
Mannose (Man) phosphate (P) isomerase	Congenital disorders of N-glycosylation CDG Ib	4351	P34949
Mannosyl- $\alpha$ -1,6-glycoprotein- $\beta$ -1,2-N-acetylglucosaminyl-transferase	Congenital disorders of N-glycosylation CDG IIa	4247	Q10469
Metalloproteinase-2 methylmalonyl-CoA mutase	Winchester syndrome	4313	P08253
N-Acetyl galactosamine $\alpha$ -4-sulfate sulfatase (arylsulfatase B)	Methylmalonic acidemia (vitamin b12 non-responsive)	4594	P22033
N-Acetyl-D-glucosaminidase	Mucopolysaccharidosis MPS VI (Maroteaux-Lamy syndrome)	411	P15848
N-Acetyl-galactosaminidase	Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-B	4669	P54802
N-Acetyl-galactosaminidase	Schindler's disease Type I (infantile severe form)	4668	P17050
N-Acetyl-galactosaminidase	Schindler's disease Type II (Kanzaki disease, adult-onset form)	4668	P17050
N-Acetyl-galactosaminidase	Schindler's disease Type III (intermediate form)	4668	P17050
N-acetyl-glucosaminine-6-sulfatase	Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-D	2799	P15586

TABLE 54-continued

Exemplary enzymatic effectors and corresponding indications			
Effector	deficiency	Entrez Gene ID	UniProt ID
N-acetylglucosaminyl-1-phosphotransferase	Mucolipidosis ML III (pseudo-Hurler's polydystrophy)	79158	Q3T906
N-Acetylglucosaminyl-1-phosphotransferase catalytic subunit	Mucolipidosis ML II (I-cell disease)	79158	Q3T906
N-acetylglucosaminyl-1-phosphotransferase, substrate-recognition subunit	Mucolipidosis ML III (pseudo-Hurler's polydystrophy) Type III-C	84572	Q9UJJ9
N-Aspartylglucosaminidase	Aspartylglucosaminuria	175	P20933
Neuramidase 1 (sialidase)	Sialidosis	4758	Q99519
Palmitoyl-protein thioesterase-1	Ceroid lipofuscinosis Adult form (CLN4, Kufs' disease)	5538	P50897
Palmitoyl-protein thioesterase-1	Ceroid lipofuscinosis Infantile form (CLN1, Santavuori-Haltia disease)	5538	P50897
Phenylalanine hydroxylase	Phenylketonuria	5053	P00439
Phosphomannomutase-2	Congenital disorders of N-glycosylation CDG Ia (solely neurologic and neurologic-multivisceral forms)	5373	O15305
Porphobilinogen deaminase	Acute Intermittent Porphyria	3145	P08397
Purine nucleoside phosphorylase	Purine nucleoside phosphorylase deficiency	4860	P00491
pyrimidine 5' nucleotidase	Hemolytic anemia and/or pyrimidine 5' nucleotidase deficiency	51251	Q9H0P0
Sphingomyelinase	Niemann-Pick disease type A	6609	P17405
Sphingomyelinase	Niemann-Pick disease type B	6609	P17405
Sterol 27-hydroxylase	Cerebrotendinous xanthomatosis (cholestanol lipidosis)	1593	Q02318
Thymidine phosphorylase	Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE)	1890	P19971
Trihexosylceramide $\alpha$ -galactosidase	Fabry's disease	2717	P06280
tyrosinase, e.g., OCA1	albinism, e.g., ocular albinism	7299	P14679
UDP-GlcNAc:dolichyl-P NAcGlc phosphotransferase	Congenital disorders of N-glycosylation CDG Ij	1798	Q9H3H5
UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase, sialin	Sialuria French type	10020	Q9Y223
Uricase	Lesch-Nyhan syndrome, gout	391051	No protein
uridine diphosphate glucuronaryl-transferase (e.g., UGT1A1)	Crigler-Najjar syndrome	54658	P22309
$\alpha$ -1,2-Mannosyltransferase	Congenital disorders of N-glycosylation CDG II (608776)	79796	Q9H6U8
$\alpha$ -1,2-Mannosyltransferase	Congenital disorders of N-glycosylation, type I (pre-Golgi glycosylation defects)	79796	Q9H6U8
$\alpha$ -1,3-Mannosyltransferase	Congenital disorders of N-glycosylation CDG II	440138	Q2TAAS
$\alpha$ -D-Mannosidase	$\alpha$ -Mannosidosis, type I (severe) or II (mild)	10195	Q92685
$\alpha$ -L-Fucosidase	Fucosidosis	4123	Q9NTJ4
$\alpha$ -1-Iduronidase	Mucopolysaccharidoses MPS I H/S (Hurler-Scheie syndrome)	2517	P04066
$\alpha$ -1-Iduronidase	Mucopolysaccharidoses MPS I-H (Hurler's syndrome)	3425	P35475

TABLE 54-continued

Exemplary enzymatic effectors and corresponding indications			
Effector	deficiency	Entrez Gene ID	UniProt ID
α-1-Iduronidase	Mucopolysaccharidosis MPS I-S (Scheie's syndrome)	3425	P35475
β-1,4-Galactosyltransferase	Congenital disorders of N-glycosylation CDG IIc	3425	P35475
β-1,4-Mannosyltransferase	Congenital disorders of N-glycosylation CDG Ick	2683	P15291
β-D-Mannosidase	β-Mannosidosis	56052	Q9BT22
β-Galactosidase	Mucopolysaccharidosis MPS IV (Morquio's syndrome) Type IV-B	4126	O00462
β-Glucuronidase	Mucopolysaccharidosis MPS VII (Sly's syndrome)	2720	P16278
β-Hexosaminidase A	Tay-Sachs disease	2990	P08236
β-Hexosaminidase B	Sandhoff's disease	3073	P06865

[0435] In some embodiments, an effector described herein comprises an enzyme of Table 54, or a functional variant thereof e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 54 by reference to its UniProt ID. In some embodiments, the functional variant catalyzes the same reaction as the corresponding wild-type protein, e.g., at a rate no less than 10%, 20%, 30%, 40%, or 50% lower than

the wild-type protein. In some embodiments, an anellovector encoding an enzyme of Table 54, or a functional variant thereof is used for the treatment of a disease or disorder of Table 54. In some embodiments, an anellovector is used to deliver uridine diphosphate glucuronyl-transferase or a functional variant thereof to a target cell, e.g., a liver cell. In some embodiments, an anellovector is used to deliver OCA1 or a functional variant thereof to a target cell, e.g., a retinal cell.

TABLE 55

Exemplary non-enzymatic effectors and corresponding indications			
Effector	Indication	Entrez Gene ID	UniProt ID
Survival motor neuron protein (SMN)	spinal muscular atrophy	6606	Q16637
Dystrophin or micro-dystrophin	muscular dystrophy (e.g., Duchenne muscular dystrophy or Becker muscular dystrophy)	1756	P11532
Complement protein, e.g., Complement factor C1	Complement Factor I deficiency	3426	P05156
Complement factor H	Atypical hemolytic uremic syndrome	3075	P08603
Cystinosin (lysosomal cystine transporter)	Cystinosis	1497	O60931
Epididymal secretory protein 1 (HE1; NPC2 protein)	Niemann-Pick disease Type C2	10577	P61916
GDP-fucose transporter-1	Congenital disorders of N-glycosylation CDG IIc (Rambam-Hasharon syndrome)	55343	Q96A29
GM2 activator protein	GM2 activator protein deficiency (Tay-Sachs disease AB variant, GM2A)	2760	Q17900
Lysosomal transmembrane CLN3 protein	Ceroid lipofuscinosis Juvenile form (CLN3, Batten disease, Vogt-Spielmeyer disease)	1207	Q13286
Lysosomal transmembrane CLN5 protein	Ceroid lipofuscinosis Variant late infantile form, Finnish type (CLN5)	1203	O75503
Na phosphate cotransporter, sialin	Infantile sialic acid storage disorder	26503	Q9NRA2

TABLE 55-continued

Exemplary non-enzymatic effectors and corresponding indications			
Effector	Indication	Entrez Gene ID	UniProt ID
Na phosphate cotransporter, sialin	Sialuria Finnish type (Salla disease)	26503	Q9NRA2
NPC1 protein	Niemann-Pick disease Type C1/Type D	4864	O15118
Oligomeric Golgi complex-7	Congenital disorders of N-glycosylation CDG IIe	91949	P83436
Prosaposin Protective protein/cathepsin A (PPCA)	Prosaposin deficiency Galactosialidosis (Goldberg's syndrome, combined neuraminidase and $\beta$ -galactosidase deficiency)	5660 5476	P07602 P10619
Protein involved in mannose-P-dolichol utilization	Congenital disorders of N-glycosylation CDG If	9526	O75352
Saposin B	Saposin B deficiency (sulfatide activator deficiency)	5660	P07602
Saposin C	Saposin C deficiency (Gaucher's activator deficiency)	5660	P07602
Sulfatase-modifying factor-1	Mucosulfatidosis (multiple sulfatase deficiency)	285362	Q8NBK3
Transmembrane CLN6 protein	Ceroid lipofuscinosis Variant late infantile form (CLN6)	54982	Q9NWW5
Transmembrane CLN8 protein	Ceroid lipofuscinosis Progressive epilepsy with intellectual disability	2055	Q9UBY8
vWF Factor I (fibrinogen)	von Willebrand disease Afibrinogenemia	7450 2243, 2244, 2266	P04275 P02671, P02675, P02679
erythropoietin (hEPO)			

**[0436]** In some embodiments, an effector described herein comprises an erythropoietin (EPO), e.g., a human erythropoietin (hEPO), or a functional variant thereof. In some embodiments, an anellovector encoding an erythropoietin, or a functional variant thereof is used for stimulating erythropoiesis. In some embodiments, an anellovector encoding an erythropoietin, or a functional variant thereof is used for the treatment of a disease or disorder, e.g., anemia. In some embodiments, an anellovector is used to deliver EPO or a functional variant thereof to a target cell, e.g., a red blood cell.

**[0437]** In some embodiments, an effector described herein comprises a polypeptide of Table 55, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 55 by reference to its UniProt ID. In some embodiments, an anellovector encoding a polypeptide of Table 55, or a functional variant thereof is used for the treatment of a disease or disorder of Table 55. In some embodiments, an anellovector is used to deliver SMN or a functional variant thereof to a target cell, e.g., a cell of the spinal cord and/or a motor neuron. In some embodiments, an anellovector is used to deliver a micro-dystrophin to a target cell, e.g., a myocyte.

**[0438]** Exemplary micro-dystrophins are described in Duan, "Systemic AAV Micro-dystrophin Gene Therapy for Duchenne Muscular Dystrophy." Mol Ther. 2018 Oct. 3; 26(10):2337-2356. doi: 10.1016/j.ymthe.2018.07.011. Epub 2018 Jul. 17.

**[0439]** In some embodiments, an effector described herein comprises a clotting factor, e.g., a clotting factor listed in Table 54 or Table 55 herein. In some embodiments, an effector described herein comprises a protein that, when mutated, causes a lysosomal storage disorder, e.g., a protein listed in Table 54 or Table 55 herein. In some embodiments, an effector described herein comprises a transporter protein, e.g., a transporter protein listed in Table 55 herein.

**[0440]** In some embodiments, a functional variant of a wild-type protein comprises a protein that has one or more activities of the wild-type protein, e.g., the functional variant catalyzes the same reaction as the corresponding wild-type protein, e.g., at a rate no less than 10%, 20%, 30%, 40%, or 50% lower than the wild-type protein. In some embodiments, the functional variant binds to the same binding partner that is bound by the wild-type protein, e.g., with a Kd of no more than 10%, 20%, 30%, 40%, or 50% higher than the Kd of the corresponding wild-type protein for the same binding partner under the same conditions. In some embodiments, the functional variant has at a polypeptide sequence at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%,

or 99% identical to that of the wild-type polypeptide. In some embodiments, the functional variant comprises a homolog (e.g., ortholog or paralog) of the corresponding wild-type protein. In some embodiments, the functional variant is a fusion protein. In some embodiments, the fusion comprises a first region with at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to the corresponding wild-type protein, and a second, heterologous region. In some embodiments, the functional variant comprises or consists of a fragment of the corresponding wild-type protein.

#### Regeneration, Repair, and Fibrosis Factors

**[0441]** Therapeutic polypeptides described herein also include growth factors, e.g., as disclosed in Table 56, or functional variants thereof, e.g., a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% identity to a protein sequence disclosed in Table 56 by reference to its UniProt ID. Also included are antibodies or fragments thereof against such growth factors, or miRNAs that promote regeneration and repair.

TABLE 56

Exemplary regeneration, repair, and fibrosis factors		
Target	Gene accession #	Protein accession #
VEGF-A	NG_008732	NP_001165094
NRG-1	NG_012005	NP_001153471

TABLE 56-continued

Exemplary regeneration, repair, and fibrosis factors		
Target	Gene accession #	Protein accession #
FGF2	NG_029067	NP_001348594
FGF1	Gene ID: 2246	NP_001341882
miR-199-3p	MIMAT0000232	
miR-590-3p	MIMAT0004801	
mi-17-92	MI0000071	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2732113/figure/F1/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2732113/figure/F1/</a>
miR-222	MI0000299	
miR-302-367	MIR302A And MIR367	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400607/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400607/</a>

#### Transformation Factors

**[0442]** Therapeutic polypeptides described herein also include transformation factors, e.g., protein factors that transform fibroblasts into differentiated cell e.g., factors disclosed in Table 57 or functional variants thereof, e.g., a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% identity to a protein sequence disclosed in Table 57 by reference to its UniProt ID.

TABLE 57

Exemplary transformation factors			
Target	Indication	Gene accession #	Protein accession #
MESP1	Organ Repair by transforming fibroblasts	Gene ID: 55897	EAX02066
ETS2	Organ Repair by transforming fibroblasts	GeneID: 2114	NP_005230
HAND2	Organ Repair by transforming fibroblasts	GeneID: 9464	NP_068808
MYOCARDIN	Organ Repair by transforming fibroblasts	GeneID: 93649	NP_001139784
ESRRA	Organ Repair by transforming fibroblasts	Gene ID: 2101	AAH92470
miR-1	Organ Repair by transforming fibroblasts	MI0000651	n/a
miR-133	Organ Repair by transforming fibroblasts	MI0000450	n/a
TGFb	Organ Repair by transforming fibroblasts	GeneID: 7040	NP_000651.3
WNT	Organ Repair by transforming fibroblasts	Gene ID: 7471	NP_005421
JAK	Organ Repair by transforming fibroblasts	Gene ID: 3716	NP_001308784
NOTCH	Organ Repair by transforming fibroblasts	GeneID: 4851	XP_011517019

## Proteins that Stimulate Cellular Regeneration

[0443] Therapeutic polypeptides described herein also include proteins that stimulate cellular regeneration e.g., proteins disclosed in Table 58 or functional variants thereof, e.g., a protein having at least 80%, 85%, 90%, 95%, 96.7%, 98%, 99% identity to a protein sequence disclosed in Table 58 by reference to its UniProt ID.

TABLE 58

Exemplary proteins that stimulate cellular regeneration		
Target	Gene accession #	Protein accession #
MST1	NG_016454	NP_066278
STK30	Gene ID: 26448	NP_036103
MST2	Gene ID: 6788	NP_006272
SAV1	Gene ID: 60485	NP_068590
LATS1	Gene ID: 9113	NP_004681
LATS2	Gene ID: 26524	NP_055387
YAP1	NG_029530	NP_001123617
CDKN2b	NG_023297	NP_004927
CDKN2a	NG_007485	NP_478102

## STING Modulator Effectors

[0444] In some embodiments, a secreted effector described herein modulates STING/cGAS signaling. In some embodiments, the STING modulator is a polypeptide, e.g., a viral polypeptide or a functional variant thereof. For instance, the effector may comprise a STING modulator (e.g., inhibitor) described in Maringer et al. "Message in a bottle: lessons learned from antagonism of STING signalling during RNA virus infection" Cytokine & Growth Factor Reviews Volume 25, Issue 6, December 2014, Pages 669-679, which is incorporated herein by reference in its entirety. Additional STING modulators (e.g., activators) are described, e.g., in Wang et al. "STING activator c-di-GMP enhances the anti-tumor effects of peptide vaccines in melanoma-bearing mice." Cancer Immunol Immunother. 2015 August; 64(8):1057-66. doi: 10.1007/s00262-015-1713-5. Epub 2015 May 19; Bose "cGAS/STING Pathway in Cancer: Jekyll and Hyde Story of Cancer Immune Response" Int J Mol Sci. 2017 November; 18(11): 2456; and Fu et al. "STING agonist formulated cancer vaccines can cure established tumors resistant to PD-1 blockade" Sci Transl Med. 2015 Apr. 15; 7(283): 283ra52, each of which is incorporated herein by reference in its entirety.

[0445] Some examples of peptides include, but are not limited to, fluorescent tag or marker, antigen, peptide therapeutic, synthetic or analog peptide from naturally-bioactive peptide, agonist or antagonist peptide, anti-microbial peptide, a targeting or cytotoxic peptide, a degradation or self-destruction peptide, and degradation or self-destruction peptides. Peptides useful in the invention described herein also include antigen-binding peptides, e.g., antigen binding antibody or antibody-like fragments, such as single chain antibodies, nanobodies (see, e.g., Steeland et al. 2016. Nanobodies as therapeutics: big opportunities for small antibodies. Drug Discov Today: 21(7):1076-113). Such antigen binding peptides may bind a cytosolic antigen, a nuclear antigen, or an intra-organelar antigen.

[0446] In some embodiments, the genetic element comprises a sequence that encodes small peptides, peptidomimetics (e.g., peptoids), amino acids, and amino acid analogs.

Such therapeutics generally have a molecular weight less than about 5,000 grams per mole, a molecular weight less than about 2,000 grams per mole, a molecular weight less than about 1,000 grams per mole, a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. Such therapeutics may include, but are not limited to, a neurotransmitter, a hormone, a drug, a toxin, a viral or microbial particle, a synthetic molecule, and agonists or antagonists thereof.

[0447] In some embodiments, the composition or anellovector described herein includes a polypeptide linked to a ligand that is capable of targeting a specific location, tissue, or cell.

## Gene Editing Components

[0448] The genetic element of the anellovector may include one or more genes that encode a component of a gene editing system. Exemplary gene editing systems include the clustered regulatory interspaced short palindromic repeat (CRISPR) system, zinc finger nucleases (ZFNs), and Transcription Activator-Like Effector-based Nucleases (TALEN). ZFNs, TALENs, and CRISPR-based methods are described, e.g., in Gaj et al. Trends Biotechnol. 31.7(2013): 397-405; CRISPR methods of gene editing are described, e.g., in Guan et al., Application of CRISPR-Cas system in gene therapy: Pre-clinical progress in animal model. DNA Repair 2016 October; 46:1-8. doi: 10.1016/j.dnarep.2016.07.004; Zheng et al., Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. BioTechniques, Vol. 57, No. 3, September 2014, pp. 115-124.

[0449] CRISPR systems are adaptive defense systems originally discovered in bacteria and archaea. CRISPR systems use RNA-guided nucleases termed CRISPR-associated or "Cas" endonucleases (e. g., Cas9 or Cpf1) to cleave foreign DNA. In a typical CRISPR/Cas system, an endonuclease is directed to a target nucleotide sequence (e. g., a site in the genome that is to be sequence-edited) by sequence-specific, non-coding "guide RNAs" that target single- or double-stranded DNA sequences. Three classes (I-III) of CRISPR systems have been identified. The class II CRISPR systems use a single Cas endonuclease (rather than multiple Cas proteins). One class II CRISPR system includes a type II Cas endonuclease such as Cas9, a CRISPR RNA ("crRNA"), and a trans-activating crRNA ("tracrRNA"). The crRNA contains a "guide RNA", typically about 20-nucleotide RNA sequence that corresponds to a target DNA sequence. The crRNA also contains a region that binds to the tracrRNA to form a partially double-stranded structure which is cleaved by RNase III, resulting in a crRNA/tracrRNA hybrid. The crRNA/tracrRNA hybrid then directs the Cas9 endonuclease to recognize and cleave the target DNA sequence. The target DNA sequence must generally be adjacent to a "protospacer adjacent motif" ("PAM") that is specific for a given Cas endonuclease; however, PAM sequences appear throughout a given genome.

[0450] In some embodiments, the anellovector includes a gene for a CRISPR endonuclease. For example, some CRISPR endonucleases identified from various prokaryotic species have unique PAM sequence requirements; examples of PAM sequences include 5'-NGG (*Streptococcus pyogenes*), 5'-NNAGAA (*Streptococcus thermophilus* CRISPR1), 5'-NGGNG (*Streptococcus thermophilus* CRISPR3), and 5'-NNNGATT (*Neisseria meningitidis*).

Some endonucleases, e. g., Cas9 endonucleases, are associated with G-rich PAM sites, e. g., 5'-NGG, and perform blunt-end cleaving of the target DNA at a location 3 nucleotides upstream from (5' from) the PAM site. Another class II CRISPR system includes the type V endonuclease Cpf1, which is smaller than Cas9; examples include AsCpf1 (from Acidaminococcus sp.) and LbCpf1 (from Lachnospiraceae sp.). Cpf1 endonucleases, are associated with T-rich PAM sites, e. g., 5'-TTN. Cpf1 can also recognize a 5'-CTA PAM motif. Cpf1 cleaves the target DNA by introducing an offset or staggered double-strand break with a 4- or 5-nucleotide 5' overhang, for example, cleaving a target DNA with a 5-nucleotide offset or staggered cut located 18 nucleotides downstream from (3' from) the PAM site on the coding strand and 23 nucleotides downstream from the PAM site on the complimentary strand; the 5-nucleotide overhang that results from such offset cleavage allows more precise genome editing by DNA insertion by homologous recombination than by insertion at blunt-end cleaved DNA. See, e. g., Zetsche et al. (2015) Cell, 163:759-771.

[0451] A variety of CRISPR associated (Cas) genes may be included in the anellovector. Specific examples of genes are those that encode Cas proteins from class II systems including Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cpf1, C2C1, or C2C3. In some embodiments, the anellovector includes a gene encoding a Cas protein, e.g., a Cas9 protein, may be from any of a variety of prokaryotic species. In some embodiments, the anellovector includes a gene encoding a particular Cas protein, e.g., a particular Cas9 protein, is selected to recognize a particular protospacer-adjacent motif (PAM) sequence. In some embodiments, the anellovector includes nucleic acids encoding two or more different Cas proteins, or two or more Cas proteins, may be introduced into a cell, zygote, embryo, or animal, e.g., to allow for recognition and modification of sites comprising the same, similar or different PAM motifs. In some embodiments, the anellovector includes a gene encoding a modified Cas protein with a deactivated nuclease, e.g., nuclease-deficient Cas9.

[0452] Whereas wild-type Cas9 protein generates double-strand breaks (DSBs) at specific DNA sequences targeted by a gRNA, a number of CRISPR endonucleases having modified functionalities are known, for example: a “nickase” version of Cas endonuclease (e.g., Cas9) generates only a single-strand break; a catalytically inactive Cas endonuclease, e.g., Cas9 (“dCas9”) does not cut the target DNA. A gene encoding a dCas9 can be fused with a gene encoding an effector domain to repress (CRISPRi) or activate (CRISPRa) expression of a target gene. For example, the gene may encode a Cas9 fusion with a transcriptional silencer (e.g., a KRAB domain) or a transcriptional activator (e.g., a dCas9-VP64 fusion). A gene encoding a catalytically inactive Cas9 (dCas9) fused to FokI nuclease (“dCas9-FokI”) can be included to generate DSBs at target sequences homologous to two gRNAs. See, e. g., the numerous CRISPR/Cas9 plasmids disclosed in and publicly available from the Addgene repository (Addgene, 75 Sidney St., Suite 550A, Cambridge, MA 02139; addgene.org/crispr/). A “double nickase” Cas9 that introduces two separate double-strand breaks, each directed by a separate guide RNA, is described as achieving more accurate genome editing by Ran et al. (2013) Cell, 154:1380-1389.

[0453] CRISPR technology for editing the genes of eukaryotes is disclosed in US Patent Application Publica-

tions 2016/0138008A1 and US2015/0344912A1, and in U.S. Pat. Nos. 8,697,359, 8,771,945, 8,945,839, 8,999,641, 8,993,233, 8,895,308, 8,865,406, 8,889,418, 8,871,445, 8,889,356, 8,932,814, 8,795,965, and 8,906,616. Cpf1 endonuclease and corresponding guide RNAs and PAM sites are disclosed in US Patent Application Publication 2016/0208243 A1.

[0454] In some embodiments, the anellovector comprises a gene encoding a polypeptide described herein, e.g., a targeted nuclease, e.g., a Cas9, e.g., a wild type Cas9, a nickase Cas9 (e.g., Cas9 D10A), a dead Cas9 (dCas9), eSpCas9, Cpf1, C2C1, or C2C3, and a gRNA. The choice of genes encoding the nuclease and gRNA(s) is determined by whether the targeted mutation is a deletion, substitution, or addition of nucleotides, e.g., a deletion, substitution, or addition of nucleotides to a targeted sequence. Genes that encode a catalytically inactive endonuclease e.g., a dead Cas9 (dCas9, e.g., D10A; H840A) tethered with all or a portion of (e.g., biologically active portion of) an (one or more) effector domain (e.g., VP64) create chimeric proteins that can modulate activity and/or expression of one or more target nucleic acids sequences.

[0455] In some embodiments, the anellovector includes a gene encoding a fusion of a dCas9 with all or a portion of one or more effector domains (e.g., a full-length wild-type effector domain, or a fragment or variant thereof, e.g., a biologically active portion thereof) to create a chimeric protein useful in the methods described herein. Accordingly, in some embodiments, the anellovector includes a gene encoding a dCas9-methylase fusion. In other some embodiments, the anellovector includes a gene encoding a dCas9-enzyme fusion with a site-specific gRNA to target an endogenous gene.

[0456] In other aspects, the anellovector includes a gene encoding 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more effector domains (all or a biologically active portion) fused with dCas9.

#### Regulatory Sequences

[0457] In some embodiments, the genetic element comprises a regulatory sequence, e.g., a promoter or an enhancer, operably linked to the sequence encoding the effector. In some embodiments, e.g., wherein the genetic element is an mRNA, a promoter may be absent from the genetic element. In some embodiments, a genetic element construct comprises a promoter that is used to drive production of the RNA genetic element.

[0458] In some embodiments, a promoter includes a DNA sequence that is located adjacent to a DNA sequence that encodes an expression product. A promoter may be linked operatively to the adjacent DNA sequence. A promoter typically increases an amount of product expressed from the DNA sequence as compared to an amount of the expressed product when no promoter exists. A promoter from one organism can be utilized to enhance product expression from the DNA sequence that originates from another organism. For example, a vertebrate promoter may be used for the expression of jellyfish GFP in vertebrates. Hence, one promoter element can enhance the expression of one or more products. Multiple promoter elements are well-known to persons of ordinary skill in the art.

[0459] In one embodiment, high-level constitutive expression is desired. Examples of such promoters include, without limitation, the retroviral Rous sarcoma virus (RSV) long

terminal repeat (LTR) promoter/enhancer, the cytomegalovirus (CMV) immediate early promoter/enhancer (see, e.g., Boshart et al, Cell, 41:521-530 (1985)), the SV40 promoter, the dihydrofolate reductase promoter, the cytoplasmic .beta.-actin promoter and the phosphoglycerol kinase (PGK) promoter.

[0460] In another embodiment, inducible promoters may be desired. Inducible promoters are those which are regulated by exogenously supplied compounds, e.g., provided either in cis or in trans, including without limitation, the zinc-inducible sheep metallothioneine (MT) promoter; the dexamethasone (Dex)-inducible mouse mammary tumor virus (MMTV) promoter; the T7 polymerase promoter system (WO 98/10088); the tetracycline-repressible system (Gossen et al, Proc. Natl. Acad. Sci. USA, 89:5547-5551 (1992)); the tetracycline-inducible system (Gossen et al., Science, 268:1766-1769 (1995); see also Harvey et al., Curr. Opin. Chem. Biol., 2:512-518 (1998)); the RU486-inducible system (Wang et al., Nat. Biotech., 15:239-243 (1997) and Wang et al., Gene Ther., 4:432-441 (1997)]; and the rapamycin-inducible system (Magari et al., J. Clin. Invest., 100: 2865-2872 (1997); Rivera et al., Nat. Medicine, 2:1028-1032 (1996)). Other types of inducible promoters which may be useful in this context are those which are regulated by a specific physiological state, e.g., temperature, acute phase, or in replicating cells only.

[0461] In some embodiments, a native promoter for a gene or nucleic acid sequence of interest is used. The native promoter may be used when it is desired that expression of the gene or the nucleic acid sequence should mimic the native expression. The native promoter may be used when expression of the gene or other nucleic acid sequence must be regulated temporally or developmentally, or in a tissue-specific manner, or in response to specific transcriptional stimuli. In a further embodiment, other native expression control elements, such as enhancer elements, polyadenylation sites or Kozak consensus sequences may also be used to mimic the native expression.

[0462] In some embodiments, the genetic element comprises a gene operably linked to a tissue-specific promoter. For instance, if expression in skeletal muscle is desired, a promoter active in muscle may be used. These include the promoters from genes encoding skeletal a-actin, myosin light chain 2A, dystrophin, muscle creatine kinase, as well as synthetic muscle promoters with activities higher than naturally-occurring promoters. See Li et al., Nat. Biotech., 17:241-245 (1999). Examples of promoters that are tissue-specific are known for liver albumin, Miyatake et al. J. Virol., 71:5124-32 (1997); hepatitis B virus core promoter, Sandig et al., Gene Ther. 3:1002-9 (1996); alpha-fetoprotein (AFP), Arbuthnot et al., Hum. Gene Ther., 7:1503-14 (1996)], bone (osteocalcin, Stein et al., Mol. Biol. Rep., 24:185-96 (1997); bone sialoprotein, Chen et al., J. Bone Miner. Res. 11:654-64 (1996)), lymphocytes (CD2, Hansal et al., J. Immunol., 161:1063-8 (1998); immunoglobulin heavy chain; T cell receptor a chain), neuronal (neuron-specific enolase (NSE) promoter, Andersen et al. Cell. Mol. Neurobiol., 13:503-15 (1993); neurofilament light-chain gene, Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991); the neuron-specific vgf gene, Piccioli et al., Neuron, 15:373-84 (1995)]; among others.

[0463] The genetic element construct may include an enhancer, e.g., a DNA sequence that is located adjacent to the DNA sequence that encodes a gene. Enhancer elements

are typically located upstream of a promoter element or can be located downstream of or within a coding DNA sequence (e.g., a DNA sequence transcribed or translated into a product or products). Hence, an enhancer element can be located 100 base pairs, 200 base pairs, or 300 or more base pairs upstream or downstream of a DNA sequence that encodes the product. Enhancer elements can increase an amount of recombinant product expressed from a DNA sequence above increased expression afforded by a promoter element. Multiple enhancer elements are readily available to persons of ordinary skill in the art.

[0464] In some embodiments, the genetic element comprises one or more inverted terminal repeats (ITR) flanking the sequences encoding the expression products described herein. In some embodiments, the genetic element comprises one or more long terminal repeats (LTR) flanking the sequence encoding the expression products described herein. Examples of promoter sequences that may be used, include, but are not limited to, the simian virus 40 (SV40) early promoter, mouse mammary tumor virus (MMTV), human immunodeficiency virus (HIV) long terminal repeat (LTR) promoter, MoMuLV promoter, an avian leukemia virus promoter, an Epstein-Barr virus immediate early promoter, and a Rous sarcoma virus promoter.

#### Other Sequences

[0465] In some embodiments, the genetic element further includes a nucleic acid encoding a product (e.g., a ribozyme, a therapeutic mRNA encoding a protein, an exogenous gene).

[0466] In some embodiments, the genetic element includes one or more sequences that affect species and/or tissue and/or cell tropism (e.g. capsid protein sequences), infectivity (e.g. capsid protein sequences), immunosuppression/activation (e.g. regulatory nucleic acids), viral genome binding and/or packaging, immune evasion (non-immunogenicity and/or tolerance), pharmacokinetics, endocytosis and/or cell attachment, nuclear entry, intracellular modulation and localization, exocytosis modulation, propagation, and nucleic acid protection of the anellovector in a host or host cell.

[0467] In some embodiments, the genetic element may comprise other sequences that include DNA, RNA, or artificial nucleic acids. The other sequences may include, but are not limited to, genomic DNA, cDNA, or sequences that encode tRNA, mRNA, rRNA, miRNA, gRNA, siRNA, or other RNAi molecules. In one embodiment, the genetic element includes a sequence encoding an siRNA to target a different loci of the same gene expression product as the regulatory nucleic acid. In one embodiment, the genetic element includes a sequence encoding an siRNA to target a different gene expression product as the regulatory nucleic acid.

[0468] In some embodiments, the genetic element further comprises one or more of the following sequences: a sequence that encodes one or more miRNAs, a sequence that encodes one or more replication proteins, a sequence that encodes an exogenous gene, a sequence that encodes a therapeutic, a regulatory sequence (e.g., a promoter, enhancer), a sequence that encodes one or more regulatory sequences that targets endogenous genes (siRNA, lncRNAs, shRNA), and a sequence that encodes a therapeutic mRNA or protein.

**[0469]** The other sequences may have a length from about 2 to about 5000 nts, about 10 to about 100 nts, about 50 to about 150 nts, about 100 to about 200 nts, about 150 to about 250 nts, about 200 to about 300 nts, about 250 to about 350 nts, about 300 to about 500 nts, about 10 to about 1000 nts, about 50 to about 1000 nts, about 100 to about 1000 nts, about 1000 to about 2000 nts, about 2000 to about 3000 nts, about 3000 to about 4000 nts, about 4000 to about 5000 nts, or any range therebetween.

#### Encoded Genes

**[0470]** For example, the genetic element may include a gene associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples include a disease associated gene or polynucleotide. A “disease-associated” gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease.

**[0471]** Examples of disease-associated genes and polynucleotides are available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.). Examples of disease-associated genes and polynucleotides are listed in Tables A and B of U.S. Pat. No. 8,697,359, which are herein incorporated by reference in their entirety. Disease specific information is available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.). Examples of signaling biochemical pathway-associated genes and polynucleotides are listed in Tables A-C of U.S. Pat. No. 8,697,359, which are herein incorporated by reference in their entirety.

**[0472]** Moreover, the genetic elements can encode targeting moieties, as described elsewhere herein. This can be achieved, e.g., by inserting a polynucleotide encoding a sugar, a glycolipid, or a protein, such as an antibody. Those skilled in the art know additional methods for generating targeting moieties.

#### Viral Sequence

**[0473]** In some embodiments, the genetic element comprises at least one viral sequence. In some embodiments, the sequence has homology or identity to one or more sequence from a single stranded DNA virus, e.g., Anellovirus, Bidnavirus, Circovirus, Geminivirus, Genomovirus, Inovirus, Microvirus, Nanovirus, Parvovirus, and Spiravirus. In some embodiments, the sequence has homology or identity to one or more sequence from a double stranded DNA virus, e.g., Adenovirus, Ampullavirus, Ascovirus, Asfarvirus, Baculovirus, Fusellovirus, Globulovirus, Guttavirus, Hytrosavirus, Herpesvirus, Iridovirus, Lipothrixvirus, Nimavirus, and Poxvirus. In some embodiments, the sequence has homology or identity to one or more sequence from an RNA virus, e.g., Alphavirus, Furovirus, Hepatitis virus, Hordeivirus, Tobamovirus, Tobravirus, Tricornavirus, Rubivirus, Birnavirus, Cystovirus, Partitivirus, and Reovirus.

**[0474]** In some embodiments, the genetic element may comprise one or more sequences from a non-pathogenic virus, e.g., a symbiotic virus, e.g., a commensal virus, e.g., a native virus, e.g., an Anellovirus. Recent changes in nomenclature have classified the three Anelloviruses able to infect human cells into Alphatorquevirus (TT), Betatorquevirus (TTM), and Gammatorquevirus (TTMD) Genera of the Anelloviridae family of viruses. In some embodiments, the genetic element may comprise a sequence with homology or identity to a Torque Teno Virus (TT), a non-enveloped, single-stranded DNA virus with a circular, negative-sense genome. In some embodiments, the genetic element may comprise a sequence with homology or identity to a SEN virus, a Sentinel virus, a TTV-like mini virus, and a TT virus. Different types of TT viruses have been described including TT virus genotype 6, TT virus group, TTV-like virus DXL1, and TTV-like virus DXL2. In some embodiments, the genetic element may comprise a sequence with homology or identity to a smaller virus, Torque Teno-like Mini Virus (TTM), or a third virus with a genomic size in between that of TTV and TTMV, named Torque Teno-like Midi Virus (TTMD). In some embodiments, the genetic element may comprise one or more sequences or a fragment of a sequence from a non-pathogenic virus having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide sequences described herein.

**[0475]** In some embodiments, the genetic element may comprise one or more sequences or a fragment of a sequence from a substantially non-pathogenic virus having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide sequences described herein, e.g., Table 41.

TABLE 41

Examples of Anelloviruses and their sequences. Accessions numbers and related sequence information may be obtained at [www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/), as referenced on Dec. 11, 2018.

Accession #	Description
AB017613.1	Torque teno virus 16 DNA, complete genome, isolate: TUS01
AB026345.1	TT virus genes for ORF1 and ORF2, complete cds, isolate: TRM1
AB026346.1	TT virus genes for ORF1 and ORF2, complete cds, isolate: TK16
AB026347.1	TT virus genes for ORF1 and ORF2, complete cds, isolate: TP1-3
AB028669.1	TT virus gene for ORF1 and ORF2, complete genome, isolate: TJJN02
AB030487.1	TT virus gene for pORF2a, pORF2b, pOrf1, complete cds, clone: JaCHCTC19
AB030488.1	TT virus gene for pORF2a, pORF2b, pOrf1, complete cds, clone: JaBD89

TABLE 41-continued

Examples of Anelloviruses and their sequences. Accessions numbers and related sequence information may be obtained at [www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/), as referenced on Dec. 11, 2018.

Accession #	Description
AB030489.1	TT virus gene for pORF2a, pORF2b, pOrf1, complete cds, clone: JaBD98
AB038340.1	TT virus genes for ORF2s, ORF1, ORF3, complete cds
AB038622.1	TT virus genes for ORF2, ORF1, ORF3, complete cds, isolate: TTVyon-LC011
AB038623.1	TT virus genes for ORF2, ORF1, ORF3, complete cds, isolate: TTVyon-KC186
AB038624.1	TT virus genes for ORF2, ORF1, ORF3, complete cds, isolate: TTVyon-KC197
AB041821.1	TT virus mRNA for VP1, complete cds
AB050448.1	Torque teno virus genes for ORF1, ORF2, ORF3, ORF4, complete cds, isolate: TYM9
AB060592.1	Torque teno virus gene for ORF1, ORF2, ORF3, ORF4, clone: SAa-39
AB060593.1	Torque teno virus gene for ORF1, ORF2, ORF3, ORF4, complete cds, clone: SAa-38
AB060595.1	TT virus gene for ORF1, ORF2, ORF3, ORF4, complete cds, clone: SAj-30
AB060596.1	TT virus gene for ORF1, ORF2, ORF3, ORF4, complete cds, clone: SAF-09
AB064596.1	Torque teno virus DNA, complete genome, isolate: CT25F
AB064597.1	Torque teno virus DNA, complete genome, isolate: CT30F
AB064599.1	Torque teno virus DNA, complete genome, isolate: JT03F
AB064600.1	Torque teno virus DNA, complete genome, isolate: JT05F
AB064601.1	Torque teno virus DNA, complete genome, isolate: JT14F
AB064602.1	Torque teno virus DNA, complete genome, isolate: JT19F
AB064603.1	Torque teno virus DNA, complete genome, isolate: JT41F
AB064604.1	Torque teno virus DNA, complete genome, isolate: CT39F
AB064606.1	Torque teno virus DNA, complete genome, isolate: JT33F
AB290918.1	Torque teno midi virus 1 DNA, complete genome, isolate: MD1-073
AF079173.1	TT virus strain TTVCHN1, complete genome
AF116842.1	TT virus strain BDH1, complete genome
AF122914.3	TT virus isolate JA20, complete genome
AF122917.1	TT virus isolate JA4, complete genome
AF122919.1	TT virus isolate JA10 unknown genes
AF129887.1	TT virus TTVCHN2, complete genome
AF247137.1	TT virus isolate TUPB, complete genome
AF254410.1	TT virus ORF2 protein and ORF1 protein genes, complete cds
AF298585.1	TT virus Polish isolate P/1C1, complete genome
AF315076.1	TTV-like virus DXL1 unknown genes
AF315077.1	TTV-like virus DXL2 unknown genes
AF345521.1	TT virus isolate TCHN-G1 Orf2 and Orf1 genes, complete cds
AF345522.1	TT virus isolate TCHN-E Orf2 and Orf1 genes, complete cds
AF345523.1	TT virus isolate TCHN-D2 Orf2 and Orf1 genes, complete cds
AF345527.1	TT virus isolate TCHN-C2 Orf2 and Orf1 genes, complete cds
AF345528.1	TT virus isolate TCHN-F Orf2 and Orf1 genes, complete cds
AF345529.1	TT virus isolate TCHN-G2 Orf2 and Orf1 genes, complete cds
AF371370.1	TT virus ORF1, ORF3, and ORF2 genes, complete cds
AJ620212.1	Torque teno virus, isolate th6, complete genome
AJ620213.1	Torque teno virus, isolate th10, complete genome
AJ620214.1	Torque teno virus, isolate th11g2, complete genome
AJ620215.1	Torque teno virus, isolate th18, complete genome
AJ620216.1	Torque teno virus, isolate th20, complete genome
AJ620217.1	Torque teno virus, isolate th21, complete genome
AJ620218.1	Torque teno virus, isolate th3, complete genome
AJ620219.1	Torque teno virus, isolate th9, complete genome
AJ620220.1	Torque teno virus, isolate th16, complete genome
AJ620221.1	Torque teno virus, isolate th17, complete genome
AJ620222.1	Torque teno virus, isolate th25, complete genome
AJ620223.1	Torque teno virus, isolate th26, complete genome
AJ620224.1	Torque teno virus, isolate th27, complete genome
AJ620225.1	Torque teno virus, isolate th31, complete genome
AJ620226.1	Torque teno virus, isolate th4, complete genome
AJ620227.1	Torque teno virus, isolate th5, complete genome
AJ620228.1	Torque teno virus, isolate th14, complete genome
AJ620229.1	Torque teno virus, isolate th29, complete genome
AJ620230.1	Torque teno virus, isolate th7, complete genome
AJ620231.1	Torque teno virus, isolate th8, complete genome
AJ620232.1	Torque teno virus, isolate th13, complete genome
AJ620233.1	Torque teno virus, isolate th19, complete genome
AJ620234.1	Torque teno virus, isolate th22g4, complete genome
AJ620235.1	Torque teno virus, isolate th23, complete genome
AM711976.1	TT virus sle1957 complete genome
AM712003.1	TT virus sle1931 complete genome
AM712004.1	TT virus sle1932 complete genome
AM712030.1	TT virus sle2057 complete genome
AM712031.1	TT virus sle2058 complete genome
AM712032.1	TT virus sle2072 complete genome

TABLE 41-continued

Examples of Anelloviruses and their sequences. Accessions numbers and related sequence information may be obtained at [www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/), as referenced on Dec. 11, 2018.

Accession #	Description
AM712033.1	TT virus sle2061 complete genome
AM712034.1	TT virus sle2065 complete genome
AY026465.1	TT virus isolate L01 ORF2 and ORF1 genes, complete cds
AY026466.1	TT virus isolate L02 ORF2 and ORF1 genes, complete cds
DQ003341.1	Torque teno virus clone P2-9-02 ORF2 (ORF2), ORF1A (ORF1A), and ORF1B (ORF1B) genes, complete cds
DQ003342.1	Torque teno virus clone P2-9-07 ORF2 (ORF2), ORF1A (ORF1A), and ORF1B (ORF1B) genes, complete cds
DQ003343.1	Torque teno virus clone P2-9-08 ORF2 (ORF2), ORF1A (ORF1A), and ORF1B (ORF1B) genes, complete cds
DQ003344.1	Torque teno virus clone P2-9-16 ORF2 (ORF2), ORF1A (ORF1A), and ORF1B (ORF1B) genes, complete cds
DQ186994.1	Torque teno virus clone P601 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ186995.1	Torque teno virus clone P605 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ186996.1	Torque teno virus clone BM1A-02 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ186997.1	Torque teno virus clone BM1A-09 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ186998.1	Torque teno virus clone BM1A-13 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ186999.1	Torque teno virus clone BM1B-05 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ187000.1	Torque teno virus clone BM1B-07 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ187001.1	Torque teno virus clone BM1B-11 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ187002.1	Torque teno virus clone BM1B-14 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ187003.1	Torque teno virus clone BM1B-08 ORF2 (ORF2) gene, complete cds; and nonfunctional ORF1 (ORF1) gene, complete sequence
DQ187004.1	Torque teno virus clone BM1C-16 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ187005.1	Torque teno virus clone BM1C-10 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds
DQ187007.1	Torque teno virus clone BM2C-25 ORF2 (ORF2) gene, complete cds; and nonfunctional ORF1 (ORF1) gene, complete sequence
DQ361268.1	Torque teno virus isolate ViPi04 ORF1 gene, complete cds
EF538879.1	Torque teno virus isolate CSC5 ORF2 and ORF1 genes, complete cds
EU305675.1	Torque teno virus isolate LTT7 ORF1 gene, complete cds
EU305676.1	Torque teno virus isolate LTT10 ORF1 gene, complete cds
EU889253.1	Torque teno virus isolate ViPi08 nonfunctional ORF1 gene, complete sequence
FJ392105.1	Torque teno virus isolate TW53A25 ORF2 gene, partial cds; and ORF1 gene, complete cds
FJ392107.1	Torque teno virus isolate TW53A27 ORF2 gene, partial cds; and ORF1 gene, complete cds
FJ392108.1	Torque teno virus isolate TW53A29 ORF2 gene, partial cds; and ORF1 gene, complete cds
FJ392111.1	Torque teno virus isolate TW53A35 ORF2 gene, partial cds; and ORF1 gene, complete cds
FJ392112.1	Torque teno virus isolate TW53A39 ORF2 gene, partial cds; and ORF1 gene, complete cds
FJ392113.1	Torque teno virus isolate TW53A26 ORF2 gene, complete cds; and nonfunctional ORF1 gene, complete sequence
FJ392114.1	Torque teno virus isolate TW53A30 ORF2 and ORF1 genes, complete cds
FJ392115.1	Torque teno virus isolate TW53A31 ORF2 and ORF1 genes, complete cds
FJ392117.1	Torque teno virus isolate TW53A37 ORF1 gene, complete cds
FJ426280.1	Torque teno virus strain SIA109, complete genome
FR751500.1	Torque teno virus complete genome, isolate TTV-HD23a (rheu215)
GU797360.1	Torque teno virus clone 8-17, complete genome
HC742700.1	Sequence 7 from Patent WO2010044889
HC742710.1	Sequence 17 from Patent WO2010044889
JX134044.1	TTV-like mini virus isolate TTMV LY1, complete genome
JX134045.1	TTV-like mini virus isolate TTMV LY2, complete genome
KU243129.1	TTV-like mini virus isolate TTMV-204, complete genome
KY856742.1	TTV-like mini virus isolate zhenjiang, complete genome
LC381845.1	Torque teno virus Human/Japan/KS025/2016 DNA, complete genome
MH648892.1	<i>Anelloviridae</i> sp. isolate ctde048, complete genome
MH648893.1	<i>Anelloviridae</i> sp. isolate ctih007, complete genome
MH648897.1	<i>Anelloviridae</i> sp. isolate ctcb038, complete genome

TABLE 41-continued

Examples of Anelloviruses and their sequences. Accessions numbers and related sequence information may be obtained at [www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/), as referenced on Dec. 11, 2018.

Accession #	Description
MH648900.1	<i>Anelloviridae</i> sp. isolate ctfc019, complete genome
MH648901.1	<i>Anelloviridae</i> sp. isolate ctbb022, complete genome
MH648907.1	<i>Anelloviridae</i> sp. isolate ctcf040, complete genome
MH648911.1	<i>Anelloviridae</i> sp. isolate cthi018, complete genome
MH648912.1	<i>Anelloviridae</i> sp. isolate ctea38, complete genome
MH648913.1	<i>Anelloviridae</i> sp. isolate ctbg006, complete genome
MH648916.1	<i>Anelloviridae</i> sp. isolate ctbg020, complete genome
MH648925.1	<i>Anelloviridae</i> sp. isolate ctci019, complete genome
MH648932.1	<i>Anelloviridae</i> sp. isolate ctid031, complete genome
MH648946.1	<i>Anelloviridae</i> sp. isolate ctbd017, complete genome
MH648957.1	<i>Anelloviridae</i> sp. isolate ctch017, complete genome
MH648958.1	<i>Anelloviridae</i> sp. isolate ctbh011, complete genome
MH648959.1	<i>Anelloviridae</i> sp. isolate ctbe020, complete genome
MH648962.1	<i>Anelloviridae</i> sp. isolate ctif015, complete genome
MH648966.1	<i>Anelloviridae</i> sp. isolate ctei055, complete genome
MH648969.1	<i>Anelloviridae</i> sp. isolate ctjg000, complete genome
MH648976.1	<i>Anelloviridae</i> sp. isolate ct ej064, complete genome
MH648977.1	<i>Anelloviridae</i> sp. isolate ctbj022, complete genome
MH648982.1	<i>Anelloviridae</i> sp. isolate ctbf014, complete genome
MH648983.1	<i>Anelloviridae</i> sp. isolate ctbd027, complete genome
MH648985.1	<i>Anelloviridae</i> sp. isolate ctch016, complete genome
MH648986.1	<i>Anelloviridae</i> sp. isolate ctbd020, complete genome
MH648989.1	<i>Anelloviridae</i> sp. isolate ctga035, complete genome
MH648990.1	<i>Anelloviridae</i> sp. isolate ct hf001, complete genome
MH648995.1	<i>Anelloviridae</i> sp. isolate ctbd067, complete genome
MH648997.1	<i>Anelloviridae</i> sp. isolate ctce026, complete genome
MH648999.1	<i>Anelloviridae</i> sp. isolate ctfb058, complete genome
MH649002.1	<i>Anelloviridae</i> sp. isolate ctij046, complete genome
MH649006.1	<i>Anelloviridae</i> sp. isolate ctcf030, complete genome
MH649008.1	<i>Anelloviridae</i> sp. isolate ctbg025, complete genome
MH649011.1	<i>Anelloviridae</i> sp. isolate ctbh052, complete genome
MH649014.1	<i>Anelloviridae</i> sp. isolate ctba003, complete genome
MH649017.1	<i>Anelloviridae</i> sp. isolate ctbb016, complete genome
MH649022.1	<i>Anelloviridae</i> sp. isolate ctch023, complete genome
MH649023.1	<i>Anelloviridae</i> sp. isolate ctbd051, complete genome
MH649028.1	<i>Anelloviridae</i> sp. isolate ctbf9, complete genome
MH649038.1	<i>Anelloviridae</i> sp. isolate ctbi030, complete genome
MH649039.1	<i>Anelloviridae</i> sp. isolate ctca057, complete genome
MH649040.1	<i>Anelloviridae</i> sp. isolate ctch033, complete genome
MH649042.1	<i>Anelloviridae</i> sp. isolate ctjd005, complete genome
MH649045.1	<i>Anelloviridae</i> sp. isolate ctde021, complete genome
MH649051.1	<i>Anelloviridae</i> sp. isolate ct dg044, complete genome
MH649056.1	<i>Anelloviridae</i> sp. isolate ctcc062, complete genome
MH649061.1	<i>Anelloviridae</i> sp. isolate ctid009, complete genome
MH649062.1	<i>Anelloviridae</i> sp. isolate ctde018, complete genome
MH649063.1	<i>Anelloviridae</i> sp. isolate ctbf012, complete genome
MH649068.1	<i>Anelloviridae</i> sp. isolate ctcc066, complete genome
MH649070.1	<i>Anelloviridae</i> sp. isolate ct da011, complete genome
MH649077.1	<i>Anelloviridae</i> sp. isolate ctbh034, complete genome
MH649083.1	<i>Anelloviridae</i> sp. isolate ct dg028, complete genome
MH649084.1	<i>Anelloviridae</i> sp. isolate ctii061, complete genome
MH649085.1	<i>Anelloviridae</i> sp. isolate ct eh021, complete genome
MH649092.1	<i>Anelloviridae</i> sp. isolate ctbg012, complete genome
MH649101.1	<i>Anelloviridae</i> sp. isolate ctif053, complete genome
MH649104.1	<i>Anelloviridae</i> sp. isolate ct ei657, complete genome
MH649106.1	<i>Anelloviridae</i> sp. isolate ct ea015, complete genome
MH649114.1	<i>Anelloviridae</i> sp. isolate ctbf050, complete genome
MH649122.1	<i>Anelloviridae</i> sp. isolate ctde002, complete genome
MH649125.1	<i>Anelloviridae</i> sp. isolate ct bb15, complete genome
MH649127.1	<i>Anelloviridae</i> sp. isolate ct ba013, complete genome
MH649137.1	<i>Anelloviridae</i> sp. isolate ct bb000, complete genome
MH649141.1	<i>Anelloviridae</i> sp. isolate ct be019, complete genome
MH649142.1	<i>Anelloviridae</i> sp. isolate ct id026, complete genome
MH649144.1	<i>Anelloviridae</i> sp. isolate ct f004, complete genome
MH649152.1	<i>Anelloviridae</i> sp. isolate ct cj13, complete genome
MH649156.1	<i>Anelloviridae</i> sp. isolate ct ci006, complete genome
MH649157.1	<i>Anelloviridae</i> sp. isolate ct bd025, complete genome
MH649158.1	<i>Anelloviridae</i> sp. isolate ct bf005, complete genome
MH649161.1	<i>Anelloviridae</i> sp. isolate ct cf045, complete genome
MH649165.1	<i>Anelloviridae</i> sp. isolate ct cc29, complete genome
MH649169.1	<i>Anelloviridae</i> sp. isolate ct ib021, complete genome

TABLE 41-continued

Examples of Anelloviruses and their sequences. Accessions numbers and related sequence information may be obtained at [www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/), as referenced on Dec. 11, 2018.

Accession #	Description
MH649172.1	<i>Anelloviridae</i> sp. isolate ctbh857, complete genome
MH649174.1	<i>Anelloviridae</i> sp. isolate ctbj049, complete genome
MH649178.1	<i>Anelloviridae</i> sp. isolate ctfc006, complete genome
MH649179.1	<i>Anelloviridae</i> sp. isolate ctbe000, complete genome
MH649183.1	<i>Anelloviridae</i> sp. isolate ctbb031, complete genome
MH649186.1	<i>Anelloviridae</i> sp. isolate ctcb33, complete genome
MH649189.1	<i>Anelloviridae</i> sp. isolate ctce12, complete genome
MH649196.1	<i>Anelloviridae</i> sp. isolate ctci060, complete genome
MH649199.1	<i>Anelloviridae</i> sp. isolate ctbb017, complete genome
MH649203.1	<i>Anelloviridae</i> sp. isolate ctch018, complete genome
MH649204.1	<i>Anelloviridae</i> sp. isolate ctbj003, complete genome
MH649206.1	<i>Anelloviridae</i> sp. isolate ctbg010, complete genome
MH649208.1	<i>Anelloviridae</i> sp. isolate ctid008, complete genome
MH649209.1	<i>Anelloviridae</i> sp. isolate ctbg056, complete genome
MH649210.1	<i>Anelloviridae</i> sp. isolate ctta001, complete genome
MH649212.1	<i>Anelloviridae</i> sp. isolate ctcf004, complete genome
MH649217.1	<i>Anelloviridae</i> sp. isolate ctbe029, complete genome
MH649223.1	<i>Anelloviridae</i> sp. isolate ctci016, complete genome
MH649224.1	<i>Anelloviridae</i> sp. isolate ctce11, complete genome
MH649228.1	<i>Anelloviridae</i> sp. isolate ctcf013, complete genome
MH649229.1	<i>Anelloviridae</i> sp. isolate ctcb036, complete genome
MH649241.1	<i>Anelloviridae</i> sp. isolate ctta027, complete genome
MH649242.1	<i>Anelloviridae</i> sp. isolate ctbf003, complete genome
MH649254.1	<i>Anelloviridae</i> sp. isolate ctjb007, complete genome
MH649255.1	<i>Anelloviridae</i> sp. isolate ctbb023, complete genome
MH649256.1	<i>Anelloviridae</i> sp. isolate ctca002, complete genome
MH649258.1	<i>Anelloviridae</i> sp. isolate ctcg010, complete genome
MH649263.1	<i>Anelloviridae</i> sp. isolate ctgh3, complete genome
MK012439.1	<i>Anelloviridae</i> sp. isolate ctte000, complete genome
MK012440.1	<i>Anelloviridae</i> sp. isolate ctjd008, complete genome
MK012448.1	<i>Anelloviridae</i> sp. isolate ctch012, complete genome
MK012457.1	<i>Anelloviridae</i> sp. isolate ctta009, complete genome
MK012458.1	<i>Anelloviridae</i> sp. isolate ctcd015, complete genome
MK012485.1	<i>Anelloviridae</i> sp. isolate ctfd011, complete genome
MK012489.1	<i>Anelloviridae</i> sp. isolate ctba003, complete genome
MK012492.1	<i>Anelloviridae</i> sp. isolate ctbb005, complete genome
MK012493.1	<i>Anelloviridae</i> sp. isolate ctcj014, complete genome
MK012500.1	<i>Anelloviridae</i> sp. isolate ctcb001, complete genome
MK012504.1	<i>Anelloviridae</i> sp. isolate ctcj010, complete genome
MK012516.1	<i>Anelloviridae</i> sp. isolate ctcf003, complete genome
NC_038336.1	Torque teno virus 5 isolate TCHN-C1 Orf2 and Orf1 genes, complete cds
NC_038338.1	Torque teno virus 11 isolate TCHN-D1 Orf2 and Orf1 genes, complete cds
NC_038339.1	Torque teno virus 13 isolate TCHN-A Orf2 and Orf1 genes, complete cds
NC_038340.1	Torque teno virus 20 ORF4, ORF3, ORF2, ORF1 genes, complete cds, clone: SAa-10
NC_038341.1	Torque teno virus 21 isolate TCHN-B ORF2 and ORF1 genes, complete cds
NC_038342.1	Torque teno virus 23 ORF2, ORF1 genes, complete cds, isolate: s-TTV CH65-2
NC_038343.1	Torque teno virus 24 ORF4, ORF3, ORF2, ORF1 genes, complete cds, clone: SAa-01
NC_038344.1	Torque teno virus 29 ORF2, ORF1, ORF3 genes, complete cds, isolate: TTVyon-KC009
NC_038345.1	Torque teno mini virus 10 isolate LIL-y1 ORF2, ORF1, ORF3, and ORF4 genes, complete cds
NC_038346.1	Torque teno mini virus 11 isolate LIL-y2 ORF2, ORF1, and ORF3 genes, complete cds
NC_038347.1	Torque teno mini virus 12 isolate LIL-y3 ORF2, ORF1, ORF3, and ORF4 genes, complete cds
NC_038350.1	Torque teno midi virus 3 isolate 2PoSMA ORF2 and ORF1 genes, complete cds
NC_038351.1	Torque teno midi virus 4 isolate 6PoSMA ORF2, ORF1, and ORF3 genes, complete cds
NC_038352.1	Torque teno midi virus 5 DNA, complete genome, isolate: MDJHem2
NC_038353.1	Torque teno midi virus 6 DNA, complete genome, isolate: MDJHem3-1
NC_038354.1	Torque teno midi virus 7 DNA, complete genome, isolate: MDJHem3-2
NC_038355.1	Torque teno midi virus 8 DNA, complete genome, isolate: MDJN1
NC_038356.1	Torque teno midi virus 9 DNA, complete genome, isolate: MDJN2
NC_038357.1	Torque teno midi virus 10 DNA, complete genome, isolate: MDJN14
NC_038358.1	Torque teno midi virus 11 DNA, complete genome, isolate: MDJN47
NC_038359.1	Torque teno midi virus 12 DNA, complete genome, isolate: MDJN51
NC_038360.1	Torque teno midi virus 13 DNA, complete genome, isolate: MDJN69

TABLE 41-continued

Examples of Anelloviruses and their sequences. Accessions numbers and related sequence information may be obtained at [www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/), as referenced on Dec. 11, 2018.

Accession #	Description
NC_038361.1	Torque teno midi virus 14 DNA, complete genome, isolate: MDJN97
NC_038362.1	Torque teno midi virus 15 DNA, complete genome, isolate: Pt-TTMDV210

[0476] In some embodiments, the genetic element comprises one or more sequences with homology or identity to one or more sequences from one or more non-Anelloviruses, e.g., adenovirus, herpes virus, pox virus, vaccinia virus, SV40, papilloma virus, an RNA virus such as a retrovirus, e.g., lentivirus, a single-stranded RNA virus, e.g., hepatitis virus, or a double-stranded RNA virus e.g., rotavirus. Since, in some embodiments, recombinant retroviruses are defective, assistance may be provided order to produce infectious particles. Such assistance can be provided, e.g., by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences within the LTR. Suitable cell lines for replicating the anellovectors described herein include cell lines known in the art, e.g., A549 cells, which can be modified as described herein. Said genetic element can additionally contain a gene encoding a selectable marker so that the desired genetic elements can be identified.

[0477] In some embodiments, the genetic element includes non-silent mutations, e.g., base substitutions, deletions, or additions resulting in amino acid differences in the encoded polypeptide, so long as the sequence remains at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the polypeptide encoded by the first nucleotide sequence or otherwise is useful for practicing the present invention. In this regard, certain conservative amino acid substitutions may be made which are generally recognized not to inactivate overall protein function: such as in regard of positively charged amino acids (and vice versa), lysine, arginine and histidine; in regard of negatively charged amino acids (and vice versa), aspartic acid and glutamic acid; and in regard of certain groups of neutrally charged amino acids (and in all cases, also vice versa), (1) alanine and serine, (2) asparagine, glutamine, and histidine, (3) cysteine and serine, (4) glycine and proline, (5) isoleucine, leucine and valine, (6) methionine, leucine and isoleucine, (7) phenylalanine, methionine, leucine, and tyrosine, (8) serine and threonine, (9) tryptophan and tyrosine, (10) and for example tyrosine, tryptophan and phenylalanine. Amino acids can be classified according to physical properties and contribution to secondary and tertiary protein structure. A conservative substitution is recognized in the art as a substitution of one amino acid for another amino acid that has similar properties.

[0478] Identity of two or more nucleic acid or polypeptide sequences having the same or a specified percentage of nucleotides or amino acid residues that are the same (e.g., about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) may be measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and

visual inspection (see, e.g., NCBI web site [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)or the like). Identity may also refer to, or may be applied to, the compliment of a test sequence. Identity also includes sequences that have deletions and/or additions, as well as those that have substitutions. As described herein, the algorithms account for gaps and the like. Identity may exist over a region that is at least about 10 amino acids or nucleotides in length, about 15 amino acids or nucleotides in length, about 20 amino acids or nucleotides in length, about 25 amino acids or nucleotides in length, about 30 amino acids or nucleotides in length, about 35 amino acids or nucleotides in length, about 40 amino acids or nucleotides in length, about 45 amino acids or nucleotides in length, about 50 amino acids or nucleotides in length, or more. Since the genetic code is degenerate, a homologous nucleotide sequence can include any number of silent base changes, i.e., nucleotide substitutions that nonetheless encode the same amino acid.

#### Proteinaceous Exterior

[0479] In some embodiments, the anellovector, e.g., synthetic anellovector, comprises a proteinaceous exterior that encloses the genetic element. The proteinaceous exterior can comprise a substantially non-pathogenic exterior protein that fails to elicit an unwanted immune response in a mammal. The proteinaceous exterior of the anellovectors typically comprises a substantially non-pathogenic protein that may self-assemble into an icosahedral formation that makes up the proteinaceous exterior.

[0480] In some embodiments, the proteinaceous exterior protein is encoded by a sequence of the genetic element of the anellovector (e.g., is in cis with the genetic element). In other embodiments, the proteinaceous exterior protein is encoded by a nucleic acid separate from the genetic element of the anellovector (e.g., is in trans with the genetic element).

[0481] In some embodiments, the protein, e.g., substantially non-pathogenic protein and/or proteinaceous exterior protein, comprises one or more glycosylated amino acids, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, or more.

[0482] In some embodiments, the protein, e.g., substantially non-pathogenic protein and/or proteinaceous exterior protein comprises at least one hydrophilic DNA-binding region, an arginine-rich region, a threonine-rich region, a glutamine-rich region, a N-terminal polyarginine sequence, a variable region, a C-terminal polyglutamine/glutamate sequence, and one or more disulfide bridges.

[0483] In some embodiments, the protein is a capsid protein, e.g., has a sequence having at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a protein encoded by any one of the nucleotide sequences encoding a capsid protein described herein, e.g., an Anellovirus ORF1 molecule and/or

capsid protein sequence, e.g., as described herein. In some embodiments, the protein or a functional fragment of a capsid protein is encoded by a nucleotide sequence having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF1 nucleic acid, e.g., as described herein.

[0484] In some embodiments, the anellovector comprises a nucleotide sequence encoding a capsid protein or a functional fragment of a capsid protein or a sequence having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF1 molecule as described herein.

[0485] In some embodiments, the ranges of amino acids with less sequence identity may provide one or more of the properties described herein and differences in cell/tissue/species specificity (e.g. tropism).

[0486] In some embodiments, the anellovector lacks lipids in the proteinaceous exterior. In some embodiments, the anellovector lacks a lipid bilayer, e.g., a viral envelope. In some embodiments, the interior of the anellovector is entirely covered (e.g., 100% coverage) by a proteinaceous exterior. In some embodiments, the interior of the anellovector is less than 100% covered by the proteinaceous exterior, e.g., 95%, 90%, 85%, 80%, 70%, 60%, 50% or less coverage. In some embodiments, the proteinaceous exterior comprises gaps or discontinuities, e.g., permitting permeability to water, ions, peptides, or small molecules, so long as the genetic element is retained in the anellovector.

[0487] In some embodiments, the proteinaceous exterior comprises one or more proteins or polypeptides that specifically recognize and/or bind a host cell, e.g., a complementary protein or polypeptide, to mediate entry of the genetic element into the host cell.

[0488] In some embodiments, the proteinaceous exterior comprises one or more of the following: an arginine-rich region, jelly-roll region, N22 domain, hypervariable region, and/or C-terminal domain, e.g., of an ORF1 molecule, e.g., as described herein. In some embodiments, the proteinaceous exterior comprises one or more of the following: one or more glycosylated proteins, a hydrophilic DNA-binding region, an arginine-rich region, a threonine-rich region, a glutamine-rich region, a N-terminal polyarginine sequence, a variable region, a C-terminal polyglutamine/glutamate sequence, and one or more disulfide bridges. For example, the proteinaceous exterior comprises a protein encoded by an Anellovirus ORF1 nucleic acid, e.g., as described herein.

[0489] In some embodiments, the proteinaceous exterior comprises one or more of the following characteristics: an icosahedral symmetry, recognizes and/or binds a molecule that interacts with one or more host cell molecules to mediate entry into the host cell, lacks lipid molecules, lacks carbohydrates, is pH and temperature stable, is detergent resistant, and is substantially non-immunogenic or non-pathogenic in a host.

### III. Methods of Use

[0490] The anellovectors and compositions comprising anellovectors described herein may be used in methods of treating a disease, disorder, or condition, e.g., in a subject (e.g., a mammalian subject, e.g., a human subject) in need thereof. Administration of a pharmaceutical composition described herein may be, for example, by way of parenteral (including intravenous, intratumoral, intraperitoneal, intramuscular, intracavity, and subcutaneous) administration.

The anellovectors may be administered alone or formulated as a pharmaceutical composition.

[0491] The anellovectors may be administered in the form of a unit-dose composition, such as a unit dose parenteral composition. Such compositions are generally prepared by admixture and can be suitably adapted for parenteral administration. Such compositions may be, for example, in the form of injectable and infusible solutions or suspensions or suppositories or aerosols.

[0492] In some embodiments, administration of an anellovector or composition comprising same, e.g., as described herein, may result in delivery of a genetic element comprised by the anellovector to a target cell, e.g., in a subject.

[0493] An anellovector or composition thereof described herein, e.g., comprising an effector (e.g., an endogenous or exogenous effector), may be used to deliver the effector to a cell, tissue, or subject. In some embodiments, the anellovector or composition thereof is used to deliver the effector to bone marrow, blood, heart, GI or skin. Delivery of an effector by administration of an anellovector composition described herein may modulate (e.g., increase or decrease) expression levels of a noncoding RNA or polypeptide in the cell, tissue, or subject. Modulation of expression level in this fashion may result in alteration of a functional activity in the cell to which the effector is delivered. In some embodiments, the modulated functional activity may be enzymatic, structural, or regulatory in nature.

[0494] In some embodiments, the anellovector, or copies thereof, are detectable in a cell 24 hours (e.g., 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 30 days, or 1 month) after delivery into a cell. In embodiments, a anellovector or composition thereof mediates an effect on a target cell, and the effect lasts for at least 1, 2, 3, 4, 5, 6, or 7 days, 2, 3, or 4 weeks, or 1, 2, 3, 6, or 12 months.

[0495] In some embodiments (e.g., wherein the anellovector or composition thereof comprises a genetic element encoding an exogenous protein), the effect lasts for less than 1, 2, 3, 4, 5, 6, or 7 days, 2, 3, or 4 weeks, or 1, 2, 3, 6, or 12 months.

[0496] Examples of diseases, disorders, and conditions that can be treated with the anellovector described herein, or a composition comprising the anellovector, include, without limitation: immune disorders, interferonopathies (e.g., Type I interferonopathies), infectious diseases, inflammatory disorders, autoimmune conditions, cancer (e.g., a solid tumor, e.g., lung cancer, non-small cell lung cancer, e.g., a tumor that expresses a gene responsive to miR-625, e.g., caspase-3), and gastrointestinal disorders. In some embodiments, the anellovector modulates (e.g., increases or decreases) an activity or function in a cell with which the anellovector is contacted. In some embodiments, the anellovector modulates (e.g., increases or decreases) the level or activity of a molecule (e.g., a nucleic acid or a protein) in a cell with which the anellovector is contacted. In some embodiments, the anellovector decreases viability of a cell, e.g., a cancer cell, with which the anellovector is contacted, e.g., by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more. In some embodiments, the anellovector comprises an effector, e.g., a miRNA, e.g., miR-625, that decreases viability of a cell, e.g., a cancer cell, with which the anellovector is contacted, e.g., by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more. In some embodiments, the anellovector

increases apoptosis of a cell, e.g., a cancer cell, e.g., by increasing caspase-3 activity, with which the anellovector is contacted, e.g., by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more. In some embodiments, the anellovector comprises an effector, e.g., an miRNA, e.g., miR-625, that increases apoptosis of a cell, e.g., a cancer cell, e.g., by increasing caspase-3 activity, with which the anellovector is contacted, e.g., by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more.

#### IV. Administration/Delivery

**[0497]** The composition (e.g., a pharmaceutical composition comprising an anellovector as described herein) may be formulated to include a pharmaceutically acceptable excipient. Pharmaceutical compositions may optionally comprise one or more additional active substances, e.g. therapeutically and/or prophylactically active substances. Pharmaceutical compositions of the present invention may be sterile and/or pyrogen-free. General considerations in the formulation and/or manufacture of pharmaceutical agents may be found, for example, in Remington: The Science and Practice of Pharmacy 21st ed., Lippincott Williams & Wilkins, 2005 (incorporated herein by reference).

**[0498]** Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to any other animal, e.g., to non-human animals, e.g. non-human mammals. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, including commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as poultry, chickens, ducks, geese, and/or turkeys.

**[0499]** Formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, dividing, shaping and/or packaging the product.

**[0500]** In one aspect, the invention features a method of delivering an anellovector to a subject. The method includes administering a pharmaceutical composition comprising an anellovector as described herein to the subject. In some embodiments, the administered anellovector replicates in the subject (e.g., becomes a part of the virome of the subject).

**[0501]** The pharmaceutical composition may include wild-type or native viral elements and/or modified viral elements. The anellovector may include one or more Anellovirus sequences (e.g., nucleic acid sequences or nucleic acid sequences encoding amino acid sequences thereof) or a sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide

sequence identity thereto. The anellovector may comprise a nucleic acid molecule comprising a nucleic acid sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% sequence identity to one or more Anellovirus sequences (e.g., an Anellovirus ORF1 nucleic acid sequence). The anellovector may comprise a nucleic acid molecule encoding an amino acid sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% sequence identity to an Anellovirus amino acid sequence (e.g., the amino acid sequence of an Anellovirus ORF1 molecule). The anellovector may comprise a polypeptide comprising an amino acid sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% sequence identity to an Anellovirus amino acid sequence (e.g., the amino acid sequence of an Anellovirus ORF1 molecule).

**[0502]** In some embodiments, the anellovector is sufficient to increase (stimulate) endogenous gene and protein expression, e.g., at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more as compared to a reference, e.g., a healthy control. In certain embodiments, the anellovector is sufficient to decrease (inhibit) endogenous gene and protein expression, e.g., at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more as compared to a reference, e.g., a healthy control.

**[0503]** In some embodiments, the anellovector inhibits/enhances one or more viral properties, e.g., tropism, infectivity, immunosuppression/activation, in a host or host cell, e.g., at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more as compared to a reference, e.g., a healthy control.

**[0504]** In some embodiments, the subject is administered the pharmaceutical composition further comprising one or more viral strains that are not represented in the viral genetic information.

**[0505]** In some embodiments, the pharmaceutical composition comprising an anellovector described herein is administered in a dose and time sufficient to modulate a viral infection. Some non-limiting examples of viral infections include adeno-associated virus, Aichi virus, Australian bat lyssavirus, BK polyomavirus, Banna virus, Barmah forest virus, Bunyamwera virus, Bunyavirus La Crosse, Bunyavirus snowshoe hare, Cercopithecine herpesvirus, Chandipura virus, Chikungunya virus, Cosavirus A, Cowpox virus, Coxsackievirus, Crimean-Congo hemorrhagic fever virus, Dengue virus, Dhori virus, Dugbe virus, Duvenhage virus, Eastern equine encephalitis virus, Ebolavirus, Echovirus, Encephalomyocarditis virus, Epstein-Barr virus, European bat lyssavirus, GB virus C/Hepatitis G virus, Hantaan virus, Hendra virus, Hepatitis A virus, Hepatitis B virus, Hepatitis C virus, Hepatitis E virus, Hepatitis delta virus, Horsepox virus, Human adenovirus, Human astrovirus, Human coronavirus, Human cytomegalovirus, Human enterovirus 68, Human enterovirus 70, Human herpesvirus 1, Human herpesvirus 2, Human herpesvirus 6, Human herpesvirus 7, Human herpesvirus 8, Human immunodeficiency virus, Human papillomavirus 1, Human papillomavirus 2, Human papillomavirus 16, Human papillomavirus 18, Human parainfluenza, Human parvovirus B19, Human respiratory syncytial virus, Human rhinovirus, Human SARS coronavirus, Human spumaretrovirus, Human T-lymphotropic virus, Human torovirus, Influenza A virus, Influenza B virus, Influenza C virus, Isfahan virus, JC polyomavirus, Japanese encephalitis virus, Junin arenavirus, KI Polyomavirus, Kun-

jin virus, Lagos bat virus, Lake Victoria marburgvirus, Langat virus, Lassa virus, Lordsdale virus, Louping ill virus, Lymphocytic choriomeningitis virus, Machupo virus, Mayaro virus, MERS coronavirus, Measles virus, Mengo encephalomyocarditis virus, Merkel cell polyomavirus, Mokola virus, Molluscum contagiosum virus, Monkeypox virus, Mumps virus, Murray valley encephalitis virus, New York virus, Nipah virus, Norwalk virus, O'nyong-nyong virus, Orf virus, Oropouche virus, Pichinde virus, Poliovirus, Punta toro phlebovirus, Puumala virus, Rabies virus, Rift valley fever virus, Rossavirus A, Ross river virus, Rotavirus A, Rotavirus B, Rotavirus C, Rubella virus, Sagi-yama virus, Salivirus A, Sandfly fever sicilian virus, Sapporo virus, Semliki forest virus, Seoul virus, Simian foamy virus, Simian virus 5, Sindbis virus, Southampton virus, St. louis encephalitis virus, Tick-borne powassan virus, Torque teno virus, Toscana virus, Uukuniemi virus, Vaccinia virus, Varicella-zoster virus, Variola virus, Venezuelan equine encephalitis virus, Vesicular stomatitis virus, Western equine encephalitis virus, WU polyomavirus, West Nile virus, Yaba monkey tumor virus, Yaba-like disease virus, Yellow fever virus, and Zika Virus. In certain embodiments, the anellovector is sufficient to outcompete and/or displace a virus already present in the subject, e.g., at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more as compared to a reference. In certain embodiments, the anellovector is sufficient to compete with chronic or acute viral infection. In certain embodiments, the anellovector may be administered prophylactically to protect from viral infections (e.g. a provirotic). In some embodiments, the anellovector is in an amount sufficient to modulate (e.g., phenotype, virus levels, gene expression, compete with other viruses, disease state, etc. at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more). In some embodiments, treatment, treating, and cognates thereof comprise medical management of a subject (e.g., by administering an anellovector, e.g., an anellovector made as described herein), e.g., with the intent to improve, ameliorate, stabilize, prevent or cure a disease, pathological condition, or disorder. In some embodiments, treatment comprises active treatment (treatment directed to improve the disease, pathological condition, or disorder), causal treatment (treatment directed to the cause of the associated disease, pathological condition, or disorder), palliative treatment (treatment designed for the relief of symptoms), preventative treatment (treatment directed to preventing, minimizing or partially or completely inhibiting the development of the associated disease, pathological condition, or disorder), and/or supportive treatment (treatment employed to supplement another therapy).

**[0506]** All references and publications cited herein are hereby incorporated by reference. The following examples are provided to further illustrate some embodiments of the present invention, but are not intended to limit the scope of the invention; it will be understood by their exemplary nature that other procedures, methodologies, or techniques known to those skilled in the art may alternatively be used.

## EXAMPLES

### Table of Contents

**[0507]** Example 1: In vitro assembly of anellovectors using ORF1 produced using baculovirus

**[0508]** Example 2: In vitro assembly of Ring2 ORF1-based anellovectors encapsidating mRNA

**[0509]** Example 3: In vitro assembly of an mRNA-encapsidating anellovector using a modified ORF1 protein

**[0510]** Example 4: In vitro assembly of an mRNA-encapsidating anellovector using a modified mRNA

**[0511]** Example 5: Structural analysis of anellovirus ORF1 capsid proteins

### Example 1: In Vitro Assembly of Anellovectors Using ORF1 Produced Using Baculovirus

**[0512]** In this example, baculovirus constructs suitable for expression of Anellovirus proteins (e.g., ORF1) were generated by in vitro assembly.

**[0513]** In a first example, DNA encoding Ring2 ORF1 fused to an N-terminal HIS<sub>6</sub>-tag (SEQ ID NO: 955) (HIS-ORF1) was codon optimized for insect expression and cloned into the baculovirus expression vector pFASTbac system according to manufacturer instructions (ThermoFisher Scientific). 10 liters of insect cell culture (Sf9) was infected with Ring2 HIS-ORF1 baculovirus and the cells were harvested 3-days post-infection by centrifugation. The cells were lysed and the lysis was purified using a chelating resin column using standard methods in the field. The elution fraction containing HIS-ORF1 was dialyzed and treated with DNase to digest host cell DNA. The resulting material was purified again using a chelating resin column and fractions containing ORF1 were retained for nucleic acid encapsidation and viral vector purification. ORF1-containing fractions were also analyzed by negative staining electron microscopy.

**[0514]** In a second example, DNA encoding Ring10 ORF1 fused to an N-terminal HIS<sub>6</sub>-tag (SEQ ID NO: 955) (HIS-ORF1) was codon optimized for insect expression and cloned into the baculovirus expression vector pFASTbac system according to manufacturer instructions (ThermoFisher Scientific). Insect cells (Sf9) were infected with Ring HIS-ORF1 baculovirus and the cells were harvested 3-days post-infection by centrifugation. The cells were lysed and the protein was purified using a chelating resin affinity column (HisTrap, GE Healthcare) using standard methods in the field. The resulting material was purified again using a heparin affinity column (Heparin HiTrap, GE Healthcare) and fractions containing ORF1 were analyzed by negative staining electron microscopy.

**[0515]** In a third example, DNA encoding chicken anemia virus (CAV) capsid protein (Vp1) fused to an N-terminal HIS<sub>6</sub>-Flag-tag (HIS-Flag-Vp1) ("HIS<sub>6</sub>" is disclosed as SEQ ID NO: 955) and helper protein (Vp2) were codon optimized for mammalian expression and cloned into a mammalian expression vector using a CMV promoter. Mammalian cells (293Expi) were transfected with CAV Vp1 and Vp2 expression vectors. The cells were harvested 3-days post-infection by centrifugation. The cells were lysed and the lysis was purified using a the chelation and heparin purification process. The elution fraction containing Chicken Anemia Virus (CAV) Vp1 were analyzed by negative staining electron microscopy.

**[0516]** As shown in FIG. 1, both Ring 2 ORF1 and Ring 10 ORF1 showed a propensity to form ~35 nm virus-like particles.

**[0517]** Nucleic acid encapsidation and viral vector purification: Ring ORF1 (wildtype protein, chimeric protein, or

fragments thereof) will be treated with conditions sufficient to dissociate VLPs or viral capsids to enable reassembly with nucleic acid cargo. Nucleic acid cargo can be defined, for example, RNA which encodes a gene of interest that one wants to deliver as a therapeutic agent. Nucleic acid cargo of defined concentration will be combined with Ring ORF1 of defined concentration and treated with conditions sufficient to permit nucleic acid encapsidation and the resulting particle, defined as viral vector, will be subsequently purified using standard viral purification procedures.

**Example 2: In Vitro Assembly of Ring2 ORF1-Based Anellovectors Encapsidating mRNA**

[0518] Ring2 ORF1 is purified by size exclusion chromatography (SEC) with mobile phases including Tris pH 8.0 with 500 mM NaCl, Tris pH 8.0 with 500 mM NaCl with 0.1% SDS, CAPS buffer pH 10.5 with 150 mM NaCl, CAPS buffer pH 10.5 with 500 mM NaCl or CAPS buffer pH 10.5 with 500 mM NaCl with 0.1% SDS to dissociate viral particles or VLPs into dispersed protein or capsomers.

[0519] In a first example, the ORF1 is mixed with mRNA, a fluorescently labeled mRNA or an mRNA transgene chemically conjugated to a segment of ssDNA shown in Example 1 to be competent for inducing vector formation. Viral vectors are formed through dialysis and SEC purification using Tris pH 8.0 buffer to isolate the anellovector encapsidating RNA (e.g., as measured by retained fluorescent absorption). Anellovector assembly is further evaluated by biophysical assessment such as DLS or electron microscopy.

[0520] In a second example, purified ORF1 is treated with 1 M NaCl with 0.1% SDS dissociate oligomers or VLPs into dispersed protein or capsomers. ORF1 is then mixed with mRNA, such as an mRNA that translate a gene of interest (e.g., a reporter gene, e.g., GFP, mCherry; or an effector of interest, e.g., EPO), and dialyzed against Tris pH 8.0 with 150 mM NaCl to permit VLP formation. The subsequent complex is purified by SEC using Tris pH 8.0 buffer to isolate the AV vector encapsidating mRNA. Anellovector vector assembly may be further evaluated by in vitro or in vivo readout, for example, by transducing cells and observing the expression of the reporter gene (e.g., mCherry or GFP) or through expression of an effector of interest (e.g., using an ELISA to detect the expression of a gene, such as EPO).

**In Vitro Assembly of Ring2 ORF1-Based Anellovectors Encapsidating GFP mRNA**

[0521] In a further example, Ring2 ORF1 protein was expressed as a full-length protein in insect cells and assembled VLPs were purified by a heparin affinity column followed by size exclusion chromatography (SEC) using a Tris buffer mobile phase. VLPs formed from the isolated Ring2 ORF1 proteins were observed with negative staining electron microscopy (EM) and had an estimated particle titer of  $10^{10}$  particles/ml (pts/ml; FIG. 4A). The VLPs were treated with 2 molar (2M) urea to disassemble the VLPs. Reimaging by EM showed no VLPs observed (FIG. 4B). Urea-treated VLPs were then dialyzed to remove urea either in the absence of mRNA (FIG. 4C) or in the presence of ~10x excess mRNA encoding GFP (FIGS. 4D and 4E). For VLP samples treated with urea and dialyzed in absence of mRNA, few particles (less than  $10^1$  particles/ml; FIG. 4C) were observed by EM. In contrast, dialysis in the presence of excess mRNA resulted in the observation of substantially

higher titers of particles ( $\sim 10^9$ - $10^{10}$  particles/ml; FIGS. 4D and 4E) by EM. These data demonstrate that disassembly and reassembly of VLPs is more efficient in the presence of mRNA, and indicate that Anellovirus ORF1 protein can be used to encapsidate mRNA in vitro to form anellovectors.

**Example 3: In Vitro Assembly of an mRNA-Encapsidating Anellovector Using a Modified ORF1 Protein**

[0522] In this example, packaging of an mRNA genetic element is improved by modifying the ORF1 protein to harbor contact residues that bind mRNA. In this example, ssDNA contact residues and/or jellyroll beta strands that contact ssDNA and/or the N-terminal arginine-rich motif (ARM) can be replaced with components of an mRNA binding viral protein or other mRNA-binding protein to permit efficient binding and packaging of mRNA. This mRNA-binding chimeric ORF1 is then treated with 1 M NaCl with 0.1% SDS to dissociate oligomers or VLPs into dispersed protein or capsomers. The chimeric ORF1 is then mixed with mRNA, such as an mRNA that encodes a gene of interest (e.g., a reporter gene, e.g., GFP, mCherry; or an effector of interest, e.g., EPO), and dialyzed against Tris pH 8.0 with 150 mM NaCl to permit VLP formation. The subsequent complex is purified by SEC using Tris pH 8.0 buffer to isolate the anellovector encapsidating mRNA. Anellovector assembly can be further evaluated by in vitro or in vivo readout, for example, by transducing cells and observing the expression of the reporter gene (e.g., mCherry or GFP) or through expression of an effector of interest (such as using an ELISA to detect the expression of a gene such as EPO).

[0523] Exemplary modifications to ORF1 molecules: Ring ORF1 molecules that may be used in the methods described herein include, for example, several wildtype Anellovirus ORF1 proteins; CAV capsid protein (VP1) variants; Anellovirus ORF1 proteins harboring mutations to improve assembly efficiency, yield or stability; and chimeric ORF1 strains or functional fragments thereof. In some instances, affinity tags are attached to the ORF1 molecule, e.g., at the N-terminus (SEQ ID NOs: 1-2). In some instances, the ORF1 molecules are untagged proteins. Ring ORF1 molecules may be expressed alone or in combination with any number of helper proteins including, but not limited to, Anellovirus ORF2 and/or ORF3 proteins.

[0524] Ring ORF1 proteins harboring mutations to improve assembly efficiency may include, but are not limited to, ORF1 proteins that harbor mutations introduced into the N-terminal arginine-rich motif (ARM), for example, to alter the pI of the ARG arm, which may permit pH sensitive nucleic acid binding to trigger particle assembly (SEQ ID NOs: 3-5). ORF1 mutations that improve stability may include, for example, mutations to the interprotomer contacting beta strands F and G of the canonical jellyroll beta-barrel (F and G beta strands), e.g., to alter the hydrophobic state of the protomer surface and/or to make capsid formation more thermodynamically favored.

[0525] Chimeric ORF1 proteins may include, but are not limited to, ORF1 proteins which have a portion or portions of their sequence replaced with comparable portions from another capsid protein, such as BFDV capsid protein, Hepatitis E capsid protein (e.g., the ARG arm and/or F and G beta strands, or comparable components thereof). Chimeric ORF1 proteins may also include ORF1 proteins which have

a portion or portions of their sequence replaced with comparable portions of another Anellovirus ORF1 protein (such as jelly roll fragments or the C-terminal portion of Ring 2 ORF1 replaced with comparable portions of Ring 9 ORF1; see, e.g., SEQ ID NOs: 8-15).

**[0526]** Generally, ORF1 molecules can be purified using purification techniques including, but not limited to, chelating purification, heparin purification, gradient sedimentation purification and/or SEC purification.

**Example 4: In Vitro Assembly of an mRNA-Encapsidating Anellovector Using a Modified mRNA**

**[0527]** In this example, encapsidation of an mRNA-based genetic element is optimized by binding the mRNA molecule to ssDNA or by modifying the mRNA transgene in such a way that that a section of the backbone would permit binding to the ssDNA contact residues of wildtype ORF1. The mRNA generally encodes a gene of interest, such as a reporter gene (e.g., GFP or mCherry), and/or an effector gene (e.g., EPO).

**[0528]** In one example, modified ssDNA that can bind ORF1 by virtue of its sugar-chain backbone, but which can also pair with mRNA non-covalently, is mixed with an mRNA of interest to produce an mRNA/DNA complex. This mRNA/DNA complex can then be encapsidated using a Ring ORF1 to form an anellovector, for example, as described below.

**[0529]** In another example, an mRNA molecule is synthesized with a section or sections of the mRNA molecule harboring a DNA backbone permitting binding and encapsidation with ORF1, while retaining the portion of the mRNA that encodes a gene (e.g., a reporter gene or an effector gene) to be delivered. This mRNA/DNA hybrid molecule can then be encapsidated using a Ring ORF1 to form an anellovector, for example, as described below.

**[0530]** Encapsulation by in vitro assembly: The mRNA/DNA genetic elements described above are then encapsidated by in vitro assembly. Briefly, anellovector ORF1 is then treated with 1 M NaCl with 0.1% SDS to dissociate oligomers or VLPs into dispersed protein or

capsomers. The ORF1 is then mixed with the synthetic mRNA complexes or hybrid molecules and dialyzed against Tris pH 8.0 with 150 mM NaCl to permit VLP formation. The subsequent particle is purified by SEC using Tris pH 8.0 buffer to isolate the anellovector encapsidating mRNA. Anellovector assembly could be further evaluated by in vitro or in vivo readout by transducing cells and observing the expression of the reporter gene or effector gene, e.g., as described herein.

**Example 5: Structural Analysis of Anellovirus ORF1 Capsid Proteins**

**[0531]** Anelloviruses share predicted structural features to other well-characterized viruses such as the avian pathogens Beak and Feather Disease Virus (BFDV) or Chicken Anemia Virus (CAV).

**[0532]** Anellovirus ORF1 capsid proteins contain an N-terminal ARM sequence similar to that of BFDV. Secondary structure prediction showed that the first ~250 residues of ORF1, dependent on strain, includes 8 predicted beta strands (FIG. 3). When the 8 predicted beta strands of ORF1, named B through I following jelly roll domain naming conventions, are aligned guided by the secondary structure prediction to the capsid proteins of BFDV and Hep E, conserved lysine and arginine residues in ORF1 align with the known ssDNA contact residues of BFDV and Hep E capsid proteins (FIG. 3, denoted by asterisk).

**Sequences**

**[0533]** The sequences listed below are annotated as follows. Bolded and underlined text indicates a sequence comprising a His6 tag (HHHHHH) used for chelating purification and a Flag tag (DYKDDDDK), a strong epitope used, e.g., for Western blot detection of low-expressing proteins. Bolded and italicized sequences indicate Ring9 ORF1 sequence or portions thereof. Unbolded, non-underlined sequences are Ring2 sequences or portions thereof. Unbolded, underlined sequences are from Beak and Feather Disease Virus (BFDV). Gray highlighting indicates the positions of lysine-to-histidine mutations, e.g., in the arginine-rich region and the first beta strand of Ring 9 ORF1.

---

SEQ ID NO: 1:  
Ring 2 N-terminal HIS-FLAG-3CProtease-ORF1:  
**MGSSHHHHHHGSDYKDDDKSGSLEVLFQGP**SGMPYYRRRYYRPRWYGRGWIRPFRRFRKRVRPTYT

TIPLKQWQPYPYKRTCYIKGQDCLYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKLCRNWWSTTN  
QDLPLVRYKGCKITFYQSTFTDYIVRIHTELAPSNSKLTPYNTHPLMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPQFE  
NKWYFATDLYKIPLLQIHCTACNLQNPFVKPDCLSNNVTLWSLNTISIQNRNMSVDQGQSWPKILGTQSFYFYTG  
NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTYEKNKFTNYWGPNFNKHIEHDMILYSLKSPEAIKNEW  
TENMKWNQLNNAGTMALTPFNEPIFTQI~~Q~~YNPDRDTGEDTQLYLSSNATGTWDPPGIPELILEGFPPLWL~~I~~YWGFAD  
FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWPQVQYQLEAQNKL~~L~~QTGPFT  
PNIQGQLSDNI~~S~~MFYKFYFKWGGSPPKAINVENPAHQI~~Q~~YPIPRNEHETTSLQSPGEAPESILYS~~P~~DYRHGNYTTTALSRI  
SQDWALKDTVSKITEPDRQQLKQALECLQISETQEKEKEVQQLISNLRQQQQLYRERIISLLKDQ

SEQ ID NO: 2:  
Ring 9 N-terminal HIS-FLAG-3CProtease-ORF1:  
**MGSSHHHHHHGSDYKDDDKSGSLEVLFQGP**SGMPYYWRQKYYRRRYYRPF~~S~~W~~R~~TRR~~T~~IQR~~R~~KRWYRKPRKTY

WRRKLRVRKRFYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQGS~~P~~ERANNNYI~~Q~~TTYSVPDKEP~~G~~GGGWTLITESLSSL

---

- continued

---

*WEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSAYTDYIAQVENCYPMTDKYTHADSAPNRMLKKH VIRVPS  
 RETRKKRKPYKRVRGPPSQMNKWFQRDICEIPLIMIAATAVDFRYPFCASDCASNLLTCLNPPLL FQNQDFDH  
 PSDTQGYFPKPGVYLYSTQRSNKPSSDCIYLGNTRDNQEKSASSLMLKTQKITDWGNPFWHYIDGSKKIFS YFK  
 PPSQLDSSDFEHMTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPSSDNLKLDGFPLYDMCWGFIDWI  
 EKVHETENLLTNYCFCIRSSAFNEKKTVDIPVDHSFLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLT GPGCARSP  
 YGNYMQAKMSYKFHVWKGGCPKTYKEPVDPSCSQPNWTIPHNLNETIQIQNPNTCPQTELQEW DWRRDIVTKKAI  
 ERIRQHTEPHETLQISTGSKHNPVHRQTSWTDSETDSEEKDQTOEIQIOLNKLHKQOHLKQQLKQYLPQNIE*

SEQ ID NO: 3:  
 Ring 2 ORF1 with ARG arm of Ring 9 (Ring 291)  
MGSSHHHHHHGSDYKDDDKSGSLEVLFGQPSGMPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTY  
*WRRKLRVRKRRVRPTYTTIPLKQWQPPYKRTCYIKGQDCLIYYSNRLGMNSTMYEKSIVPVHWPGGGSFSVSM LTDA  
 ALYDIHKLCRNWWTSTNQDPLVRYKGCKITFYQSTFTDYIVRIHELPANSNKLTYPNTHPLMMMSKYKHIIPS RQ  
 RRKKPYTKIFVKPPPQFENKWFATDLYKIPLLQIHCTACNLQNPVFKPDKLSNNVTLWSLNTISIQNRNMSV DQGS  
 WPKILGTQSFYFYFTGANLPGDTTQIPVADLLPLTNPRINRPGQLNEAKITDHTFTEYKNKFTNYWGNP FNKHIQE  
 HLDMILYSLKSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGT GDPPG  
 IPELILEGFPLWLIYWGFAFDQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPL PEDNI KWYP  
 QVQYQLEAQNKLLQGTGPFTPNIQGQLSDNISMFYKFWGGSPPKAINVENPAHQIQYPIPRNEHETTS LQSPGEAPE  
 SILYSFDYRHGNYTTTALSRI SDWALKDTVS KITEPDRQQLKQALECLQISEETQEKEKEVQQLISNL RQQQOLYRERI  
 ISLLKDQ*

SEQ ID NO: 4:  
 Ring 2 ORF1 with ARG arm and Beta strands 1 + 2 722 epitope of Ring 9 (Ring 292)  
MGSSHHHHHHGSDYKDDDKSGSLEVLFGQPSGMPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTY  
*WRRKLRVRKRFYKRKLKIVLKQFQPKIIRRCTIFGTICLFQGSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSM LTDA  
 LYDIHKLCRNWWTSTNQDPLVRYKGCKITFYQSTFTDYIVRIHELPANSNKLTYPNTHPLMMMSKYKHIIPS RQTR  
 RKKPYTKIFVKPPPQFENKWFATDLYKIPLLQIHCTACNLQNPVFKPDKLSNNVTLWSLNTISIQNRNMSV DQGSW  
 PFKILGTQSFYFYFTGANLPGDTTQIPVADLLPLTNPRINRPGQLNEAKITDHTFTEYKNKFTNYWGNP FNKHIQEHL  
 DMILYSLKSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGT GDPPGIP  
 EILILEGFPLWLIYWGFAFDQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPL PEDNI KWYPQV  
 QYQLEAQNKLLQGTGPFTPNIQGQLSDNISMFYKFWGGSPPKAINVENPAHQIQYPIPRNEHETTS LQSPGEAPE S  
 YSF DYRHGNYTTTALSRI SDWALKDTVS KITEPDRQQLKQALECLQISEETQEKEKEVQQLISNL RQQQOLYRERI ISL  
 LKDQ*

SEQ ID NO: 5:  
 Ring 9 with LYS/HIS mutations in ARG arm and first beta strand  
MGSSHHHHHHGSDYKDDDKSGSLEVLFGQPSGMPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTY   
*WRRKLRVRHRYHRLHHHVLQFQPKIIRRCTIFGTICLFQGS PERANNNYIQTIIYSPVDPKEPGGGWTLITESLSSL  
 WEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSAYTDYIAQVENCYPMTDKYTHADSAPNRMLKKH VIRVPS  
 RETRKKRKPYKRVRGPPSQMNKWFQRDICEIPLIMIAATAVDFRYPFCASDCASNLLTCLNPPLL FQNQDFDH  
 PSDTQGYFPKPGVYLYSTQRSNKPSSDCIYLGNTRDNQEKSASSLMLKTQKITDWGNPFWHYIDGSKKIFS YFK  
 PPSQLDSSDFEHMTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPSSDNLKLDGFPLYDMCWGFIDWI  
 EKVHETENLLTNYCFCIRSSAFNEKKTVDIPVDHSFLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLT GPGCARSP*

---

- continued

---

**YGNYMQAKMSYKFHVWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAI**

**ERIRQHTEPHETLQISTGSKHNPVHRQTSPWTDSETDSEEKDQTQEIQIQLNKLKRHQQHLKQQLKQYLKPQNIE**

SEQ ID NO: 6:

Ring 9 with ARG arm of BFDV:

**MGSSHHHHHGSDYKDDDKSGSLEVLFQGPSPGMWGTNSCACAKFQIRRRYARPYRRR<sup>35</sup>HIRRYYRRRRHFR<sup>36</sup>R  
RFRTTNRRKRFYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQGS PERANNNYIQT<sup>37</sup>IYSYVPDKEPGGGWTLITESLSSL**

**WEDWEHLKNVVWTQSNAGLPLVRYGGVTLYFYQSA<sup>38</sup>YTDYIAQVENCYPMTDTKYTHAD<sup>39</sup>SAPNRMLKKH<sup>40</sup>VIRVPS**

**RETRKKRK<sup>41</sup>PYKRVRVGPPSQMNK<sup>42</sup>WYFQRDICEIPLIMIAATAVDFRYPFCASDCASNNLTCLNPLL<sup>43</sup>FQ<sup>44</sup>NQDFDH**

**PSDTQGYFPKPGVYLYSTQRSNK<sup>45</sup>PS<sup>46</sup>SSDCIYL<sup>47</sup>GNTKD<sup>48</sup>NQEGKSASSL<sup>49</sup>MTL<sup>50</sup>K<sup>51</sup>Q<sup>52</sup>K<sup>53</sup>ITDWGNPFWHYYIDGSKKIFS<sup>54</sup>YFK**

**PPSQLDSSDFEHMTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPSSDNLKLDGFPLYDMCWGFIDWI**

**EKVHETENLLTNYCFCIRSSAFNEKKTVFIPVDHSFLTGFS<sup>55</sup>YETPVKSSDQAH<sup>56</sup>WHPQIRFQTKSINDICLTGPGCARSP**

**YGNYMQAKMSYKFHVWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAI**

**ERIRQHTEPHETLQISTGSKHNPVHRQTSPWTDSETDSEEKDQTQEIQIQLNKLKRHQQHLKQQLKQYLKPQNIE**

SEQ ID NO: 7:

Ring 9 with beta strands F and G of BFDV capsid protein:

**MGSSHHHHHGSDYKDDDKSGSLEVLFQGPSPGM<sup>57</sup>PPWRQ<sup>58</sup>YRR<sup>59</sup>YRPF<sup>60</sup>SWTR<sup>61</sup>RI<sup>62</sup>IQRRKRWRYRKPRKT<sup>63</sup>**

**WRRKL<sup>64</sup>LRV<sup>65</sup>RKRFYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQGS PERANNNYIQT<sup>66</sup>IYSYVPDKEPGGGWTLITESLSSL**

**WEDWEHLKNVVWTQSNAGLPLVRYGGVTLYFYQSA<sup>67</sup>YTDYIAQVENCYPMTDTKYTHAD<sup>68</sup>SAPNRMLKKH<sup>69</sup>AKKWF<sup>70</sup>S**

**RETRKKRK<sup>71</sup>P<sup>72</sup>G<sup>73</sup>F<sup>74</sup>K<sup>75</sup>R<sup>76</sup>L<sup>77</sup>G<sup>78</sup>P<sup>79</sup>S<sup>80</sup>QM<sup>81</sup>MN<sup>82</sup>K<sup>83</sup>WYFQRDICEIPLIMIAATAVDFRYPFCASDCASNNLTCLNPLL<sup>84</sup>FQ<sup>85</sup>NQDFDH<sup>86</sup>P**

**SDTQGYFPKPGVYLYSTQRSNK<sup>87</sup>PS<sup>88</sup>SSDCIYL<sup>89</sup>GNTKD<sup>90</sup>NQEGKSASSL<sup>91</sup>MTL<sup>92</sup>K<sup>93</sup>T<sup>94</sup>Q<sup>95</sup>ITDWGNPFWHYYIDGSKKIFS<sup>96</sup>YFKP**

**PSQLDSSDFEHMTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPSSDNLKLDGFPLYDMCWGFIDWIE**

**KVHETENLLTNYCFCIRSSAFNEKKTVFIPVDHSFLTGFS<sup>97</sup>YETPVKSSDQAH<sup>98</sup>WHPQIRFQTKSINDICLTGPGCARSPY**

**GNYMQAKMSYKFHVWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAI**

**RIRQHTEPHETLQISTGSKHNPVHRQTSPWTDSETDSEEKDQTQEIQIQLNKLKRHQQHLKQQLKQYLKPQNIE**

SEQ ID NO: 8:

Ring 2 with beta C of Ring 9:

**MGSSHHHHHGSDYKDDDKSGSLEVLFQGPSPGM<sup>99</sup>MPYYRR<sup>100</sup>RYRR<sup>101</sup>PRWYGRGWIRRPFRRKRRVRPTYT**

**TIPLKQWQPPYKRC<sup>102</sup>TIFGTICLFQGSNLRLGMNSTMYEKSIVPVHWP<sup>103</sup>GGGSFSVSMLTLDALYDIHKLCRNWW<sup>104</sup>STN**

**QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPL<sup>105</sup>MMMSKYKHI<sup>106</sup>PSRQTRKKPYTKIFVK<sup>107</sup>PPPQFE**

**NK<sup>108</sup>WYFATDLYKIPLLQI<sup>109</sup>HCTACNLQNP<sup>110</sup>FVKPD<sup>111</sup>KLSNNV<sup>112</sup>TLWSLN<sup>113</sup>TISIQNRNMSVDQGQSWPF<sup>114</sup>KILGTQS<sup>115</sup>FYFY<sup>116</sup>TGA**

**NLPGDTTQIPVADLLPLTNPRINRGQSLNEAKITDHTFTEYKNKFTNYWG<sup>117</sup>NP<sup>118</sup>FN<sup>119</sup>KHI<sup>120</sup>EHLD<sup>121</sup>MILYSLK<sup>122</sup>SPEAI<sup>123</sup>NEW<sup>124</sup>T**

**TENMKWNQLNNAGT<sup>125</sup>MLTPFNEPIFTQI<sup>126</sup>QYNPDRDTG<sup>127</sup>EDTQ<sup>128</sup>LYLLSNATGTGW<sup>129</sup>D<sup>130</sup>PPG<sup>131</sup>I<sup>132</sup>PEL<sup>133</sup>I<sup>134</sup>LEG<sup>135</sup>FPL<sup>136</sup>WL<sup>137</sup>I<sup>138</sup>YWG<sup>139</sup>FAD**

**FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLV<sup>140</sup>LN<sup>141</sup>DTFVEG<sup>142</sup>NS<sup>143</sup>YE<sup>144</sup>KQPL<sup>145</sup>PED<sup>146</sup>NI<sup>147</sup>KWY<sup>148</sup>PQ<sup>149</sup>YQ<sup>150</sup>LEA<sup>151</sup>QNKL<sup>152</sup>LTGP<sup>153</sup>F**

**PNIQGQLSDNISM<sup>154</sup>MFYK<sup>155</sup>FWGG<sup>156</sup>SP<sup>157</sup>KAIN<sup>158</sup>VENPAHQ<sup>159</sup>I<sup>160</sup>QY<sup>161</sup>P<sup>162</sup>IR<sup>163</sup>E<sup>164</sup>H<sup>165</sup>T<sup>166</sup>TS<sup>167</sup>LQ<sup>168</sup>S<sup>169</sup>P<sup>170</sup>G<sup>171</sup>E<sup>172</sup>A<sup>173</sup>P<sup>174</sup>E<sup>175</sup>I<sup>176</sup>S<sup>177</sup>F<sup>178</sup>D<sup>179</sup>Y<sup>180</sup>R<sup>181</sup>H<sup>182</sup>G<sup>183</sup>N<sup>184</sup>TT<sup>185</sup>T<sup>186</sup>AL<sup>187</sup>S<sup>188</sup>R<sup>189</sup>I**

**SQDWALKDTVSKITEPD<sup>190</sup>RQ<sup>191</sup>QL<sup>192</sup>L<sup>193</sup>QALEC<sup>194</sup>LQ<sup>195</sup>I<sup>196</sup>SE<sup>197</sup>T<sup>198</sup>Q<sup>199</sup>E<sup>200</sup>K<sup>201</sup>E<sup>202</sup>K<sup>203</sup>E<sup>204</sup>V<sup>205</sup>Q<sup>206</sup>Q<sup>207</sup>Q<sup>208</sup>LYR<sup>209</sup>R<sup>210</sup>E<sup>211</sup>R<sup>212</sup>I<sup>213</sup>IS<sup>214</sup>LL<sup>215</sup>KD<sup>216</sup>Q**

SEQ ID NO: 9:

Ring 2 with linker 1 of Ring 9:

**MGSSHHHHHGSDYKDDDKSGSLEVLFQGPSPGM<sup>217</sup>MPYYRR<sup>218</sup>RYRR<sup>219</sup>PRWYGRGWIRRPFRRKRRVRPTYT**

**TIPLKQWQPPYKRC<sup>220</sup>YI<sup>221</sup>KGQD<sup>222</sup>C<sup>223</sup>I<sup>224</sup>Y<sup>225</sup>S<sup>226</sup>PER<sup>227</sup>AN<sup>228</sup>NY<sup>229</sup>I<sup>230</sup>Q<sup>231</sup>T<sup>232</sup>I<sup>233</sup>Y<sup>234</sup>S<sup>235</sup>Y<sup>236</sup>VPD<sup>237</sup>K<sup>238</sup>E<sup>239</sup>PG<sup>240</sup>GG<sup>241</sup>W<sup>242</sup>T<sup>243</sup>L<sup>244</sup>I<sup>245</sup>T<sup>246</sup>S<sup>247</sup>L<sup>248</sup>S<sup>249</sup>L<sup>250</sup>W<sup>251</sup>E<sup>252</sup>D<sup>253</sup>WE<sup>254</sup>H<sup>255</sup>L<sup>256</sup>K<sup>257</sup>N<sup>258</sup>V<sup>259</sup>WT<sup>260</sup>QS<sup>261</sup>N**

**AGLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPL<sup>262</sup>MMMSKYKHI<sup>263</sup>IPSRQTRKKPYTKIFVK<sup>264</sup>PPPQFE**

**NK<sup>265</sup>WYFATDLYKIPLLQI<sup>266</sup>HCTACNLQNP<sup>267</sup>FVKPD<sup>268</sup>KLSNNV<sup>269</sup>TLWSLN<sup>270</sup>TISIQNRNMSVDQGQSWPF<sup>271</sup>KILGTQS<sup>272</sup>FYFY<sup>273</sup>TGA**

**NLPGDTTQIPVADLLPLTNPRINRGQSLNEAKITDHTFTEYKNKFTNYWG<sup>274</sup>NP<sup>275</sup>FN<sup>276</sup>KHI<sup>277</sup>EHLD<sup>278</sup>MILYSLK<sup>279</sup>SPEAI<sup>280</sup>NEW<sup>281</sup>T**

- continued

---

TENMKWNQLNNAAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLWLYWGFAD  
 FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT  
 PNIQGQLSDNISMFYKFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLSQSPGEAPEISLYSFYDVRHGNYTTTALSRI  
 SQDWALKDTVS KITEPDROQQLKQALECLQISEETQEKEKEVEQQLISNLRQQQQLYRERIIISLLKDQ

SEQ ID NO: 10:  
 Ring 2 with beta strand D of Ring 9:  
MGSSSHHHHHHGSDYKDDDKSGSLEVLFGPGSGMPYYRRRYNRRPRWYGRGWIRRPFRRKRRVRPTYT  
 TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMTLTDALYDIHKLCRNWWSTN  
 QDLPLVRYGGVTLYFYQSTFTDYIVRIHTELPANSNLTYPNTHPLMMMSKYKHIIPSRQTRKKPYTKIFVKPPPQF  
 ENKWYFATDLYKIPLLQIHCTACNLQNPFWKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYTG  
 ANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTYEKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNE  
 WTTEENMKWNQLNNAAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLWLYWGF  
 ADFOQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTG  
 PFTPNIQGQLSDNISMFYKFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLSQSPGEAPEISLYSFYDVRHGNYTTTAL  
 SRISQDWALKDTVS KITEPDROQQLKQALECLQISEETQEKEKEVEQQLISNLRQQQQLYRERIIISLLKDQ

SEQ ID NO: 11:  
 Ring 2 with linker 2 of Ring 9:  
MGSSSHHHHHHGSDYKDDDKSGSLEVLFGPGSGMPYYRRRYNRRPRWYGRGWIRRPFRRKRRVRPTYT  
 TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMTLTDALYDIHKLCRNWWSTN  
 QDLPLVRYKGCKITFYQSTFTDYIVRIHCYPMTDKYTHADSAPNRMLKKHKKHIIPSRQTRKKPYTKIFVKPPPQF  
 NKWYFATDLYKIPLLQIHCTACNLQNPFWKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYTG  
 NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTYEKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNEW  
 TENMKWNQLNNAAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLWLYWGFAD  
 FOKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT  
 PNIQGQLSDNISMFYKFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLSQSPGEAPEISLYSFYDVRHGNYTTTALSRI  
 SQDWALKDTVS KITEPDROQQLKQALECLQISEETQEKEKEVEQQLISNLRQQQQLYRERIIISLLKDQ

SEQ ID NO: 12:  
 Ring 2 with beta strand G DNA binding of Ring 9:  
MGSSSHHHHHHGSDYKDDDKSGSLEVLFGPGSGMPYYRRRYNRRPRWYGRGWIRRPFRRKRRVRPTYT  
 TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMTLTDALYDIHKLCRNWWSTN  
 QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNLTYPNTHPLMMMSKYKHIIPSRQTRKKPYKRVVKPPPQF  
 ENKWYFATDLYKIPLLQIHCTACNLQNPFWKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYTG  
 ANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTYEKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNE  
 WTTEENMKWNQLNNAAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLWLYWGF  
 ADFOQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTG  
 PFTPNIQGQLSDNISMFYKFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLSQSPGEAPEISLYSFYDVRHGNYTTTAL  
 SRISQDWALKDTVS KITEPDROQQLKQALECLQISEETQEKEKEVEQQLISNLRQQQQLYRERIIISLLKDQ

SEQ ID NO: 13:  
 Ring 2 with beta strand F interprotomer contact of Ring 9:  
MGSSSHHHHHHGSDYKDDDKSGSLEVLFGPGSGMPYYRRRYNRRPRWYGRGWIRRPFRRKRRVRPTYT  
 TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMTLTDALYDIHKLCRNWWSTN  
 QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNLTYPNTHPLMMMSKHVIRVPSRQTRKKPYTKIFVKPPPQF

- continued

---

ENKWYFATDLYKIPPLQLIHCTACNLQNPVVKPDKLSNNVTLWSLNTISIQNRNMSVDQQSWPFKILGTQSFYFYTG  
ANLPGDTTQIPVADLLPLTNPRINRPGQLSNEAKITDHITPTEYKNKFTNYWGNPNKHIQEHLDMILYSLKSPEAIKNE  
WTENMKWNQLNAGTMALETPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPIELILEGFPLWLIYWGF  
ADFQKNLKKVTNIDTNYMLVAKTKFTQPGTFYLVLNDTFVEGNSPYEKQPLPEDNIKWYPQVQYOLEAQNKLLQTG  
PFTPNIQGQLSDNISMFYKFWGGSPPKAINVENPAHQIQCYPIPRNEHETTSQSPGEAPESILYSFDYRHGNYTTAL  
SRISQDWALKDTVS KITEPDRQQLLKQALECLQISEETQEKEKEVQQLISNLRQQQOLYRERIIISLLKDQ

SEQ ID NO: 14:  
Ring 2 ORF1 with strand H and I with C-terminal fragment of Ring 9:  
MGSSHHHHHHGSDYKDDDKSGSLEVLFGPGSGMPYYRRRYNRRPRWYGRGWIRRPFRRKRRVRPTYT

TIPLKQWQPPYKRTCYIKGQDCLIIYNSNLRLGMNSTMYEKSIIVPVHWPGGSFSVSMLTLDALYDIHKLCRNWWSTSTN  
QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNLTYPNTHPLMMMSKYKHIIPSQRTRKKPYTKIFVKPPPQFE  
NKWYFATDLYKIPPLIMAATAVDFRYPFCASDNNTLTLCLNPLLFQNQDFDHPSDTQGYFPKPGVYLYSTQRSNK  
PSSSDCIYLGNTKDNQEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSFKPPSQLDSSDFEHMTELAEPMFIQ  
VRYNPERDTGQGNLIIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWIEKVHETENLLTNYCFCIRSSAFNEK  
KTVFIPVDHSFLTGFSPIETPVKSSDQAHWHFQIRFQTKSINDICLTGPGCARSPYGNIMQAKMSYKFHVWKWGCP  
KTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWWDWRDIVTKKAIERIRQHTEPHETLQISTGSKHNPPV  
HRQTSPWTDSETDSEEKDQTQEIQIQLNKLKRHQQHLKQQLKQYLPQNI

SEQ ID NO: 15:  
Ring 2 ORF1 with strand I and C-terminal fragment of Ring 9:  
MGSSHHHHHHGSDYKDDDKSGSLEVLFGPGSGMPYYRRRYNRRPRWYGRGWIRRPFRRKRRVRPTYT

TIPLKQWQPPYKRTCYIKGQDCLIIYNSNLRLGMNSTMYEKSIIVPVHWPGGSFSVSMLTLDALYDIHKLCRNWWSTSTN  
QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNLTYPNTHPLMMMSKYKHIIPSQRTRKKPYTKIFVKPPPQFE  
NKWYFATDLYKIPPLQIHCTACNLQNPVVKPDKLSNNLTLCNPLLFQNQDFDHPSDTQGYFPKPGVYLYSTQRSNK  
SSSDCIYLGNTKDNQEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSFKPPSQLDSSDFEHMTELAEPMFIQ  
RYNPERDTGQGNLIIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWIEKVHETENLLTNYCFCIRSSAFNEK  
TVFIPVDHSFLTGFSPIETPVKSSDQAHWHFQIRFQTKSINDICLTGPGCARSPYGNIMQAKMSYKFHVWKWGCP  
TYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWWDWRDIVTKKAIERIRQHTEPHETLQISTGSKHNPPV  
HRQTSPWTDSETDSEEKDQTQEIQIQLNKLKRHQQHLKQQLKQYLPQNI

---

#### SEQUENCE LISTING

Sequence total quantity: 968  
SEQ ID NO: 1 moltype = AA length = 699  
FEATURE Location/Qualifiers  
REGION 1..699 note = Description of Artificial Sequence: Synthetic  
polypeptide  
source 1..699 mol\_type = protein  
organism = synthetic construct  
SEQUENCE: 1  
MGSSHHHHH HGSDYKDDDK SGSLEVLFQG PSGMPYYR RRYNRYRPRW YGRGWIRRPF 60  
RRRFRRKRV RPTYTTIPLQ QWQPPYKRTC YIKGQDCLIIY YSNLRLGMNS TMYEKSIIVPV 120  
HWPGGGSRV SMLTLDALYD IHKLCRNWWST SNQDPLVLR YKGCKITFYQ STFTDYIVRI 180  
HTELPANSNK LTYPNTHPLM MMSKYKHII PSRQTRRKKK PYTKIFVKPP PQFENKWWFA 240  
TDLYKIPPLQ IHCTACNLQN PFVKPDKLSN NVTLWSLNTI SIQNRNMSVD QGQSWPFKIL 300  
GTQSFYFYFY TGANLPGDTT QIPVADLLPL TNPRINRPGQ SLNEAKITDH ITFTEYKNKF 360  
TNWGNPNK HIQEHLDMIL YSLKSPEAIK NEWTENMKW NQLNNAGTMA LTPFNEPIFT 420  
QIQYNPDRDT GEDTQLYLLS NATGTGWDPP GIPELILEGF PLWLIYWGFA DFQKNLKVT 480

-continued

---

SEQ ID NO: 2	moltype = AA length = 688
FEATURE	Location/Qualifiers
REGION	1..688
	note = Description of Artificial Sequence: Synthetic polypeptide
source	1..688
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 2	
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPPYWRQ KYRRRYRPF SWRTRRIQR 60	
RKRWRYRKPR KTYWRRKLRV RKRFYKRKLK KIVLKQFQPK IIRRCTIFGT ICLFOGSPER 120	
ANNNYIQTYY SYVPDKEPGG GGWTLITESL SSLWEDWEHL KNVWTQSNAG LPLVRYGGVT 180	
LYFYQSAAYTD YIAQVNCYP MTDTKYTHAD SAPNRMLLKK HVirVPsRET RKKRKPYKRV 240	
RVGPPSQMQN KWYFQRDICE IPLIMIATA VDFRYPFCAS DCASNNLTLT CLNPLLFQNN 300	
DFDHPSPDTQG YFPKPGVLY STQRSNKPSS SDCIYLGNTK DNQECKSASS LMILKTQKIT 360	
DWGNPFWHYY IDGSKKIFSY FKPPS QLDSS DFEHMTELAE PMFIQVRYNP ERDTGQGNLI 420	
YVTFENFRQH WDPPSSDNLK LDGFLPLYDMC WGFIDWIEKF HETENLLTNY CFCICRSSAFN 480	
EKKTVFIPVD HSFLTGFSFY ETPVKSSDQA HWHPQIRFQI KSINDICLTG PGCARSPYGN 540	
YMQAKMSYKF HVKWGGCPKT YEKPYDPCSQ PNWTIPHNLN ETIQIQNPNPT CPQTELQEWQD 600	
WRRDIVTKKA IERIRQHTEP HETLQISTGS KHNPVHRQD SPWTDSETDS EEEKDQTQEI 660	
QIQLNKLRKH QQHLKQQLKQ YLKPQNIE 688	
SEQ ID NO: 3	moltype = AA length = 714
FEATURE	Location/Qualifiers
REGION	1..714
	note = Description of Artificial Sequence: Synthetic polypeptide
source	1..714
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 3	
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPPYWRQ KYRRRYRPF SWRTRRIQR 60	
RKRWRYRKPR KTYWRRKLRV RKRRVRPPTY TIPPLKQWP YKRTCYIKGQ DCLIIYSNLR 120	
LGMNSTMYEK SIVPVHWPGG GSFSVSMMLT DALYDIHKLC RNWWTSTNQD LPLVRYKGCK 180	
ITFYQSTFTD YIVRIHTEP ANSNKLTPYN THPLMMMSMK YKHIIPSQRT RRKKPYTKI 240	
FVKPPPQFEN KWYFATDLYK IPLLQIHCTA CNLQNPFPVKP DKLSNNVTWL SLNTISIQNR 300	
NMSVDQGQSW PFKILGTQSF YFYFVTGANL PGDTTQIPVA DLLPLTNPRI NRPGQSLNEA 360	
KITDHITFTE YKNKFTNYWG NPFNKHQIPEH LDMLYSLKS PEAIKNEWTT ENMKWNQLNN 420	
AGTMALTPFN EPIFTQIQYN PDRDTGEDTQ LYLLSNATGT GWDPPGIPEL ILEGFPLWLI 480	
YWGFADPQKQ LKKVTNIDTN YMLVAKTKFT QKPGTFYLVLI LNDTFVEGNS PYEKQPLPED 540	
NIKWYPQVQY QLEAQNKLLQ TGPFPTPNIQQ QLSDNISMFY KFYFKWGGSP PKAINVENPA 600	
HQIQYPIPQRN EHETTSQSP GEAPESILYS FDYRHGNYTT TALSRSQDW ALKDTVSKIT 660	
EPDRQQLLKQ ALECLQISEE TQEKKKEKEVQ QLISNLRQQQ QLYRERIISL LKDQ 714	
SEQ ID NO: 4	moltype = AA length = 714
FEATURE	Location/Qualifiers
REGION	1..714
	note = Description of Artificial Sequence: Synthetic polypeptide
source	1..714
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 4	
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPPYWRQ KYRRRYRPF SWRTRRIQR 60	
RKRWRYRKPR KTYWRRKLRV RKRFYKRKLK KIVLKQFQPK IIRRCTIFGT ICLFOGSNL 120	
LGMNSTMYEK SIVPVHWPGG GSFSVSMMLT DALYDIHKLC RNWWTSTNQD LPLVRYKGCK 180	
ITFYQSTFTD YIVRIHTEP ANSNKLTPYN THPLMMMSMK YKHIIPSQRT RRKKPYTKI 240	
FVKPPPQFEN KWYFATDLYK IPLLQIHCTA CNLQNPFPVKP DKLSNNVTWL SLNTISIQNR 300	
NMSVDQGQSW PFKILGTQSF YFYFVTGANL PGDTTQIPVA DLLPLTNPRI NRPGQSLNEA 360	
KITDHITFTE YKNKFTNYWG NPFNKHQIPEH LDMLYSLKS PEAIKNEWTT ENMKWNQLNN 420	
AGTMALTPFN EPIFTQIQYN PDRDTGEDTQ LYLLSNATGT GWDPPGIPEL ILEGFPLWLI 480	
YWGFADPQKQ LKKVTNIDTN YMLVAKTKFT QKPGTFYLVLI LNDTFVEGNS PYEKQPLPED 540	
NIKWYPQVQY QLEAQNKLLQ TGPFPTPNIQQ QLSDNISMFY KFYFKWGGSP PKAINVENPA 600	
HQIQYPIPQRN EHETTSQSP GEAPESILYS FDYRHGNYTT TALSRSQDW ALKDTVSKIT 660	
EPDRQQLLKQ ALECLQISEE TQEKKKEKEVQ QLISNLRQQQ QLYRERIISL LKDQ 714	
SEQ ID NO: 5	moltype = AA length = 688
FEATURE	Location/Qualifiers
REGION	1..688
	note = Description of Artificial Sequence: Synthetic polypeptide

-continued

---

```

source          1..688
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 5
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPYWRQ KYYRRRYRPF SWRTRRIIQR 60
RHRWRYRKPR HTYWRRHLRV RHRFYHRHLH HVILKQFQPK IIRRCTIFGT ICLFQGSPER 120
ANNNYIQTIY SYVPDKEPGG GGWLITESL SSLWEDWEHL KNWWTQSNAQ LPLVRYGGVT 180
LYFQSQAYTD YIAQVFNCYP MTDTKYTHAD SAPNRMILLKK H VIRVPSRET RKKRKPYKRV 240
RVGPPSQQMN KWYFQRDICE IPLIMIAATA VDFRYPFCAS DCASNLLTLC CLNPLLFQNNQ 300
DFDHPSDTQG YFPKPGVYLY STQRSNKPSS SDCIYLGNTK DNQEGKSASS LMTLKTQKIT 360
DWGNPFWHYI IDGSKKIFSY FKPPSQQLDSS DFEHMTELAE PMFIQVRYNP ERDTGQGNLI 420
YVTENFRGQH WDPPSSDNLKL LDGFPLYDMC WGFIDWIEKV HETENLLTNY CFCIRSSAFN 480
EKKTVFIPVD HSFLTGFSPY ETPVKSSDQA HWHPQIRFQTK KSINDICLTG PGCARSPYGN 540
YMQAKMSYKF HVKWWGGCPKT YEKPYDPCSQ PNWTIPHNLN ETIQIQNPNT CPQTELQEW 600
WRRDIVTKKA IERIROHTEP HELQISTGSK HNPPVHRQTS SPWTDSETDS EEEKDQTQE 660
IQQLNKLRLKH QQHLKQQLKQ YLKPNIE 688

SEQ ID NO: 6      moltype = AA length = 687
FEATURE          Location/Qualifiers
REGION           1..687
               note = Description of Artificial Sequence: Synthetic
               polypeptide
source          1..687
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 6
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMWGTSNC ACAKFQIRRR YARPYRRRH 60
RRYRRRRRRHF RRRRFITNRR KRFYKRKLK IVLKQFQPKI IIRRCTIFGTI CLFQGSPERA 120
NNNYIQTIYI YVPDKEPGGG GWTLITESLS SLWEDWEHLK NVWTQSNAGL PLVRYGGVT 180
YFQSQAYTDY IAQVFNCYP MTDTKYTHADS APNRMILLKKH VIRVPSRETR KKRKPYKVR 240
VGPPSQQMNQ KWYFQRDICEI PLIMIAATA DFRYPFCASD CASNNLLTLC LNPLLFQNNQD 300
FDHPSDTQGY FPKPGVYLYS TQRSNKPSSS DCIYLGNTKD NQEGKSASSL MTLKTQKITD 360
WGPNPFWHYI DGSKKIFSYF KPPSQQLDSSD FEHMTELAEPM FIQVRYNPE RDTGQGNLIY 420
VTEENFRGQHW DPPSSDNLKL DGFPPLYDMCW GFIDWIEKVH ETENLLTNYC FCIRSSAFNE 480
KKTIVFIPVDH SFLTGFSPEY ETPVKSSDQA WHWPQIRFQTK SIN DICLTGP GCARSPYGN 540
YMQAKMSYKFHV KWWGGCPKT YEKPYDPCSQ PNWTIPHNLN ETIQIQNPNTC PQTELQEW 600
RDRDIVTKKA IERIROHTEPH ETLQISTGSK HNPPVHRQTS PWTDSETDSE EEEKDQTQE 660
IQQLNKLRLKH QQHLKQQLKQY LKPQNIE 687

SEQ ID NO: 7      moltype = AA length = 688
FEATURE          Location/Qualifiers
REGION           1..688
               note = Description of Artificial Sequence: Synthetic
               polypeptide
source          1..688
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 7
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPYWRQ KYYRRRYRPF SWRTRRIIQR 60
RKRWRYRKPR KTYWRRKLRV RKRFYKRKLK KIVLKQFQPK IIRRCTIFGT ICLFQGSPER 120
ANNNYIQTIYI SYVPDKEPGG GGWLITESL SSLWEDWEHL KNWWTQSNAQ LPLVRYGGVT 180
LYFQSQAYTDY IAQVFNCYP MTDTKYTHAD SAPNRMILLKK HAKKPWSRET RKKRKPGFKR 240
LIGPPSQQMN KWYFQRDICE IPLIMIAATA VDFRYPFCAS DCASNLLTLC CLNPLLFQNNQ 300
DFDHPSDTQG YFPKPGVYLY STQRSNKPSS SDCIYLGNTK DNQEGKSASS LMTLKTQKIT 360
DWGNPFWHYI DGSKKIFSY FKPPSQQLDSS DFEHMTELAE PMFIQVRYNP ERDTGQGNLI 420
YVTENFRGQH WDPPSSDNLKL LDGFPLYDMC WGFIDWIEKV HETENLLTNY CFCIRSSAFN 480
EKKTVFIPVD HSFLTGFSPY ETPVKSSDQA HWHPQIRFQTK SIN DICLTGP PG CARSPYGN 540
YMQAKMSYKF HVKWWGGCPKT YEKPYDPCSQ PNWTIPHNLN ETIQIQNPNT CPQTELQEW 600
WRRDIVTKKA IERIROHTEP HELQISTGSK HNPPVHRQTS SPWTDSETDS EEEKDQTQE 660
IQQLNKLRLKH QQHLKQQLKQ YLKPNIE 688

SEQ ID NO: 8      moltype = AA length = 699
FEATURE          Location/Qualifiers
REGION           1..699
               note = Description of Artificial Sequence: Synthetic
               polypeptide
source          1..699
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 8
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPYYRRY RRYNYRRPRW YGRGWIRRP 60
RRSSRRRKRVR PRTYTTIPLK QWQPPYKRRC TIFGTICLFO GS NLRLGMNS TMYEKSIVPV 120
HWPGGGSFSV SMLTLDALYD IHKLCRNWST STNQDLPLVY YKGCKITFYQ STFTDYIVRI 180
HTELPANSNK LTYPNTHPLM MMMSKYKHI PSRQTRRKKK PYTKIFVKPP PQFENKWF 240
TDLYKIPLLQ IHCTACNLQN PFVKPDKLSN NVTLWSLNTI SIQNRNMSVD QGQSWPFKIL 300
GTQSFYFYFY TGANLPGDTT QIPVADLLPL TNPRINRPGQ SLNEAKITDH ITFTEYKNF 360

```

-continued

---

TNYWGNPFNK	HIQEHLDMIL	YSLKSPEAIK	NEWTTENMKW	NQLNNNAGTMA	LTPFNEPIFT	420
QIQQNPDRDT	GEDTQLYLSS	NATGTGWDPP	GIPELILEGF	PLWLIYWGFA	DFQKNLKVT	480
NIDTNYMLVA	KTKFTQKPGT	FYLVILNDTF	VEGNSPYEKQ	PLPEDNIKWY	PQVQYQLEAQ	540
NKLLQTGPFT	PNIQGQLSDN	ISMFYKFYFK	WGGSPPKAIN	VENPAHQIQQ	PIPRNEHETT	600
SIQSPGEAPE	SILYSFDYRH	GNYTTTALSR	ISQDWALKDT	VSKITEPDRQ	QLLKQALECL	660
QISEETQEKK	EKEVQQLISN	LRQQQQLYRE	RIISLLKDQ			699

SEQ ID NO: 9	moltype = AA	length = 699				
FEATURE	Location/Qualifiers					
REGION	1..699					
	note = Description of Artificial Sequence: Synthetic					
	polypeptide					
source	1..699					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 9						
MGSSHHHHHH	GSDYKDDDK	SGSLEVLFQG	PSGMMPYYYR	RRYNYRPRPW	YGRGWIRRPF	60
RRRFRRKRV	RPTYTTIPLK	QWPQPYKRTC	YIKQDCLII	YSPERANNY	IQTIIYSYVPD	120
KEPGGGWTL	ITESLSSLWE	DWEHLKRNWT	QSNAGLPLVR	YKGCKITFYQ	STFTDYIVRI	180
HTELPANSNK	LTYPNTHPLM	MMMSKYKHI	PSRQTRRKKK	PYTKIFVKPP	PQFENKWKYFA	240
TDLYKIPLLQ	IHCACNLQN	PFVKPDKLSN	NVTLWSLNNTI	SIQNRNMSVD	QGQSWPFKIL	300
GTQSFYFY	TGANLPGDTT	QIPVADLLPL	TNPRINRPQG	SLNEAKITDH	ITFTBEYKNAF	360
TNYWGNPFNK	HIQEHLDMIL	YSLKSPEAIK	NEWTTENMKW	NQLNNNAGTMA	LTPFNEPIFT	420
QIQQNPDRDT	GEDTQLYLSS	NATGTGWDPP	GIPELILEGF	PLWLIYWGFA	DFQKNLKVT	480
NIDTNYMLVA	KTKFTQKPGT	FYLVILNDTF	VEGNSPYEKQ	PLPEDNIKWY	PQVQYQLEAQ	540
NKLLQTGPFT	PNIQGQLSDN	ISMFYKFYFK	WGGSPPKAIN	VENPAHQIQQ	PIPRNEHETT	600
SIQSPGEAPE	SILYSFDYRH	GNYTTTALSR	ISQDWALKDT	VSKITEPDRQ	QLLKQALECL	660
QISEETQEKK	EKEVQQLISN	LRQQQQLYRE	RIISLLKDQ			699
SEQ ID NO: 10	moltype = AA	length = 699				
FEATURE	Location/Qualifiers					
REGION	1..699					
	note = Description of Artificial Sequence: Synthetic					
	polypeptide					
source	1..699					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 10						
MGSSHHHHHH	GSDYKDDDK	SGSLEVLFQG	PSGMMPYYYR	RRYNYRPRPW	YGRGWIRRPF	60
RRRFRRKRV	RPTYTTIPLK	QWPQPYKRTC	YIKQDCLII	YSPERANNY	IQTIIYSYVPD	120
HWPGGSFSV	SMLTLDALYD	IHKLCRWNWT	STNQDLPPLVR	YGGVTLFYQ	STFTDYIVRI	180
HTELPANSNK	LTYPNTHPLM	MMMSKYKHI	PSRQTRRKKK	PYTKIFVKPP	PQFENKWKYFA	240
TDLYKIPLLQ	IHCACNLQN	PFVKPDKLSN	NVTLWSLNNTI	SIQNRNMSVD	QGQSWPFKIL	300
GTQSFYFY	TGANLPGDTT	QIPVADLLPL	TNPRINRPQG	SLNEAKITDH	ITFTBEYKNAF	360
TNYWGNPFNK	HIQEHLDMIL	YSLKSPEAIK	NEWTTENMKW	NQLNNNAGTMA	LTPFNEPIFT	420
QIQQNPDRDT	GEDTQLYLSS	NATGTGWDPP	GIPELILEGF	PLWLIYWGFA	DFQKNLKVT	480
NIDTNYMLVA	KTKFTQKPGT	FYLVILNDTF	VEGNSPYEKQ	PLPEDNIKWY	PQVQYQLEAQ	540
NKLLQTGPFT	PNIQGQLSDN	ISMFYKFYFK	WGGSPPKAIN	VENPAHQIQQ	PIPRNEHETT	600
SIQSPGEAPE	SILYSFDYRH	GNYTTTALSR	ISQDWALKDT	VSKITEPDRQ	QLLKQALECL	660
QISEETQEKK	EKEVQQLISN	LRQQQQLYRE	RIISLLKDQ			699
SEQ ID NO: 11	moltype = AA	length = 699				
FEATURE	Location/Qualifiers					
REGION	1..699					
	note = Description of Artificial Sequence: Synthetic					
	polypeptide					
source	1..699					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 11						
MGSSHHHHHH	GSDYKDDDK	SGSLEVLFQG	PSGMMPYYYR	RRYNYRPRPW	YGRGWIRRPF	60
RRRFRRKRV	RPTYTTIPLK	QWPQPYKRTC	YIKQDCLII	YSPERANNY	IQTIIYSYVPD	120
HWPGGSFSV	SMLTLDALYD	IHKLCRWNWT	STNQDLPPLVR	YKGCKITFYQ	STFTDYIVRI	180
HNCYPMTDTK	YTHADSAPNK	MLLKKHHII	PSRQTRRKKK	PYTKIFVKPP	PQFENKWKYFA	240
TDLYKIPLLQ	IHCACNLQN	PFVKPDKLSN	NVTLWSLNNTI	SIQNRNMSVD	QGQSWPFKIL	300
GTQSFYFY	TGANLPGDTT	QIPVADLLPL	TNPRINRPQG	SLNEAKITDH	ITFTBEYKNAF	360
TNYWGNPFNK	HIQEHLDMIL	YSLKSPEAIK	NEWTTENMKW	NQLNNNAGTMA	LTPFNEPIFT	420
QIQQNPDRDT	GEDTQLYLSS	NATGTGWDPP	GIPELILEGF	PLWLIYWGFA	DFQKNLKVT	480
NIDTNYMLVA	KTKFTQKPGT	FYLVILNDTF	VEGNSPYEKQ	PLPEDNIKWY	PQVQYQLEAQ	540
NKLLQTGPFT	PNIQGQLSDN	ISMFYKFYFK	WGGSPPKAIN	VENPAHQIQQ	PIPRNEHETT	600
SIQSPGEAPE	SILYSFDYRH	GNYTTTALSR	ISQDWALKDT	VSKITEPDRQ	QLLKQALECL	660
QISEETQEKK	EKEVQQLISN	LRQQQQLYRE	RIISLLKDQ			699
SEQ ID NO: 12	moltype = AA	length = 699				
FEATURE	Location/Qualifiers					
REGION	1..699					

-continued

---

```

note = Description of Artificial Sequence: Synthetic
      polypeptide
source 1..699
       mol_type = protein
       organism = synthetic construct

SEQUENCE: 12
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPYYYRR RRYNYRPRW YGRGWIRRPF 60
RRRFRRKRV RPTYTTIPLK QWQPPYKRTC YIKGQDCLYY YSNLRLGMNS TMYEKSIVPV 120
HWPGGGSFSV SMLTLDALYD IHKLCRNNWWT STNQDPLPLVR YKGCKITFYQ STFTDYIVRI 180
HTELPANSNK LTYPNTHPLM MMMSKYKHI PSRQTRRKKK PYKCKRVRVKKP PQFENKWYFA 240
TDLYKIPLLQ IHCTACNLQN PVFKPDKLSN NVTLWSLNTI SIQNRNMSVD QGQSWPFKIL 300
GTQSFYFYFY TGANLPGDTT QIPVADLLPL TNPRINRPQQ SLNEAKITDH ITFTBEYKNAF 360
TNYWGNPFSNKH IQEHLDMIL YSLKSPEAIK NEWTENMKW NQLNNAGTMA LTPPNEPIFT 420
QIQYNPDRDT GEDTQLYLLS NATGTGWDPP GIPELILEGF PLWLIYWGFA DFQKNLKVKV 480
NIDTNYMLVA KTKFTQKPGT FYLVILNDTF VEGNSPYEKQ PLPEDNIKYW PQVQYQLEAQ 540
NKLLQTGPFT PNIQGQLSDN ISMFYKFYFK WGGSPPKAIN VENPAHQIQY PIPRNEHETT 600
SLQSPGEAPE SILYSFDYRH GNYTTTALSR ISQDWALKDT VSKITEPDRQ QLLKQALECL 660
QISEETQEKK EKEVQQLISN LRQQQQLYRE RIISLLKDQ 699

SEQ ID NO: 13      moltype = AA length = 699
FEATURE           Location/Qualifiers
REGION            1..699
note = Description of Artificial Sequence: Synthetic
      polypeptide
source 1..699
       mol_type = protein
       organism = synthetic construct

SEQUENCE: 13
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPYYYRR RRYNYRPRW YGRGWIRRPF 60
RRRFRRKRV RPTYTTIPLK QWQPPYKRTC YIKGQDCLYY YSNLRLGMNS TMYEKSIVPV 120
HWPGGGSFSV SMLTLDALYD IHKLCRNNWWT STNQDPLPLVR YKGCKITFYQ STFTDYIVRI 180
HTELPANSNK LTYPNTHPLM MMMSKYKHI PSRQTRRKKK PYTKIFVKKP PQFENKWYFA 240
TDLYKIPLLQ IHCTACNLQN PVFKPDKLSN NVTLWSLNTI SIQNRNMSVD QGQSWPFKIL 300
GTQSFYFYFY TGANLPGDTT QIPVADLLPL TNPRINRPQQ SLNEAKITDH ITFTBEYKNAF 360
TNYWGNPFSNKH IQEHLDMIL YSLKSPEAIK NEWTENMKW NQLNNAGTMA LTPPNEPIFT 420
QIQYNPDRDT GEDTQLYLLS NATGTGWDPP GIPELILEGF PLWLIYWGFA DFQKNLKVKV 480
NIDTNYMLVA KTKFTQKPGT FYLVILNDTF VEGNSPYEKQ PLPEDNIKYW PQVQYQLEAQ 540
NKLLQTGPFT PNIQGQLSDN ISMFYKFYFK WGGSPPKAIN VENPAHQIQY PIPRNEHETT 600
SLQSPGEAPE SILYSFDYRH GNYTTTALSR ISQDWALKDT VSKITEPDRQ QLLKQALECL 660
QISEETQEKK EKEVQQLISN LRQQQQLYRE RIISLLKDQ 699

SEQ ID NO: 14      moltype = AA length = 673
FEATURE           Location/Qualifiers
REGION            1..673
note = Description of Artificial Sequence: Synthetic
      polypeptide
source 1..673
       mol_type = protein
       organism = synthetic construct

SEQUENCE: 14
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPYYYRR RRYNYRPRW YGRGWIRRPF 60
RRRFRRKRV RPTYTTIPLK QWQPPYKRTC YIKGQDCLYY YSNLRLGMNS TMYEKSIVPV 120
HWPGGGSFSV SMLTLDALYD IHKLCRNNWWT STNQDPLPLVR YKGCKITFYQ STFTDYIVRI 180
HTELPANSNK LTYPNTHPLM MMMSKYKHI PSRQTRRKKK PYTKIFVKKP PQFENKWYFA 240
TDLYKIPLLQ IAATAVDFRY PFCASDCASN NLTLTCLNPL LFQNQDFDHP SDTQGYFPKP 300
GVYLYSTQRS NKPSSSDCIY LGNTKDNQEG KSASSLMLTLK TQKITDWGNP FWHYIIDGSK 360
KIFSYFKPQS QLDSSDFEHM TELAEPMFIC VRVNPERDTG QGNLIVVTTEN FRGQHDDPPS 420
SDNLKLDGFP LYDMCWFID WIEKVHETEN LLTNYCFCIR SSAFNEKKT VIPVDSFLT 480
GFSPYETPVK SSDQAHWHPQ IFRFQTKSIND ICLTGPFCAR SPYGNYMQAK MSYKPHVKWG 540
GCPKTYEKPY DPCSQPNWTI PHNLNETIQI QNPNTCPQTE LQEWDWRDI VTKKAIERIR 600
QHTEPHETLQ ISTGSKHNPP VHRQTPSWTD SETDSEEKKD QTQEIQIQLN KLRKHQQLK 660
QQLKQYLNPKQ NIE 673

SEQ ID NO: 15      moltype = AA length = 673
FEATURE           Location/Qualifiers
REGION            1..673
note = Description of Artificial Sequence: Synthetic
      polypeptide
source 1..673
       mol_type = protein
       organism = synthetic construct

SEQUENCE: 15
MGSSHHHHHH GSDYKDDDK SGSLEVLFQG PSGMPYYYRR RRYNYRPRW YGRGWIRRPF 60
RRRFRRKRV RPTYTTIPLK QWQPPYKRTC YIKGQDCLYY YSNLRLGMNS TMYEKSIVPV 120
HWPGGGSFSV SMLTLDALYD IHKLCRNNWWT STNQDPLPLVR YKGCKITFYQ STFTDYIVRI 180
HTELPANSNK LTYPNTHPLM MMMSKYKHI PSRQTRRKKK PYTKIFVKKP PQFENKWYFA 240

```

-continued

---

TDLYKIPLLQ	IHCCTACNLQN	PFVKPDKLSN	NLTLCNLPL	LFQNQDFDHP	SDTQGYFPKP	300
GVLYLSTQRS	NKPSSDCIV	LGNTKDNQEG	KSASSLMTLK	TQKITDWGNP	FWHYYIDGSK	360
KIFSYFKPPS	QLDSSDFEHM	TELAEPMFIQ	VRYNPERDTG	QGNLIYVTEN	FRGQHWDDPPS	420
SDNLKLDFP	LYDMCWGFID	WIEKVHETEN	LLTNYCFCIR	SSAFNEKKTV	FIPVDHSFLT	480
GFSPYETPVK	SSDQAHWHPO	IRFQTKSIND	ICLTGPGCAR	SPYGNYMQAK	MSYKFHVKG	540
GCPKTYEKPY	DPCSOPNWFI	PHNLNETIQI	QNPNTCPOTE	LQEWDWRDI	VTKKAIERIR	600
QHTEPHETLQ	ISTGSKHNP	VHRQTSPWT	SETDSEEKD	QTQEIQIQLN	KLRKHQQHLK	660
QQLKQYLPQ	NIE					673

---

SEQ ID NO: 16	moltype = DNA	length = 3753			
FEATURE	Location/Qualifiers				
source	1..3753				
	mol_type = genomic DNA				
	organism = Alphatorquevirus sp.				
SEQUENCE: 16					
tgcgtacgtca ctaaccacag	tgtcctctac	aggccaaatcg	cagtctatgt	cgtgcacttc	60
ctggggcatgg	tctacataat	tatataaaatg	cttgcacttc	cgaaatggctg	120
gcccgcgcgc	ggagaggaggc	cacggcagggg	gatccgaacg	tcctgaggcc	180
ggtgaggtta	cacacccgaa	tcaaggggca	attcggggctc	aggactggcc	240
caaggcttt	aaaaatgcac	ttttcgtcaa	taaagcagaaa	aaaaggaaaa	300
tttgcgtgc	agcagactaa	aaaaaaccaa	ctgtatgag	cttctggaaa	360
acaatgtcac	ggggatccaa	cgcgtgtgtt	atgagtcctt	tcaccgtgac	420
tttgtgggtt	tgggaaatctt	atacttcacaa	ttactgcact	tgtgaaaca	480
caacaggccc	gagacccctt	gggcacccgg	gagtagaccc	caaccccccac	540
ccaggcctgc	ccggccgcgt	ccggagccct	cacaggtta	ttcgagacca	600
ggcatgggaa	ttgtggaaagc	gacggaggcg	ctgggtgttc	cggaagcgtt	660
caagactcgc	agacgtggcc	ctcgatcgac	tcgtcgccgc	cctagacgac	720
gaggcgcaga	cggtggaggaa	ggggggagacg	aaaacaagg	acttagacac	780
ctttagacgc	aggggacgaa	aagcaaaact	tataataaaa	ctgtggcaac	840
taaaagatgc	agaataaaagg	gatacatacc	actgtattaa	agtggaaacg	900
cacaacattt	accagtccaa	ttaatgacac	aataatgaaa	ggcccttcg	960
cagcaactatg	agggtcagcc	tctacatttt	gttgaggagg	cacctcagac	1020
cttgaccaga	agcaacagata	acccatggat	ttgggggctt	cagtaaaaat	1080
atacaggcac	ccagaccaag	actttatagt	aatatacaac	agaagaaccc	1140
caacatctac	acagccaccc	ctctacaccc	aggcaatgc	attttagcaa	1200
atttagatca	agtttacaga	caagacaaaa	gggttagaaaa	gcaatttagac	1260
acccccccaca	ctctttacag	acaagtggta	ctttcaaaag	gacatagccg	1320
tttcaacatc	atggcgatgt	aggctgactt	gcccgttccg	ttctgtcac	1380
caacatcttc	atccatgtcc	aggctcttac	ttccgttta	aacaactacc	1440
tacctttat	aatgacaact	cagactcaa	gtttaaaaagaa	tttttaataa	1500
aacaacacggc	acaaaaggaa	caagttaaa	tgcaactaat	acatttagaa	1560
cataagtca	ccacaactaa	aaaaacccaaa	cccacaataa	aacaaaccat	1620
atactttgc	ccttttagat	ccctctgggg	agaccccccata	tacttaatgt	1680
aaacaaaatgt	ttgaacata	tcatggagaa	aataactata	aaaaacatga	1740
tgcaccaact	agagaatttc	caaattcata	ccaagggaaac	aggcccttt	1800
aggcatatac	agccccaccat	accttaaacc	aggcagaata	tctccagaaa	1860
gtacacagaaa	ataattttaca	acccttacac	agacaaaggaa	actggaaaca	1920
ggacccacta	actaaaggaga	aaacacataa	taaagaaggaa	cagacaaat	1980
tgacatgccc	ctatggactt	tactttttgg	atatacagac	ttgtgtaaaa	2040
taactgggac	tttacactaa	actacagact	agtactaata	tgcccttata	2100
atttgtacaat	aaaaaaatgt	aaagatctgg	gtacatcccg	tactcttaca	2160
gggttcagatg	ccagacggca	gcaactacat	accctttca	ttttagagcaa	2220
cacagactta	caccaccaac	aggtatgtt	ggacataa	aggagccggc	2280
taaggttagaa	aaaccaagca	ctcagctgtt	aatgaagtt	ttttttact	2340
cggtaaccct	atcatgttac	agatgtttaa	agacccccc	ttccagccca	2400
accccggtacc	ggttaacatcc	ctagaatggaa	acaagtttgc	gaccgggg	2460
gcactactcg	ttccggtcat	ggacatcg	cacacacaca	ttttagag	2520
gagagtgtca	gaacaacaag	aaacttctga	ccttgttattt	tcagggccaa	2580
ggtcgcacatc	ccaaaaacaag	aaacccaaga	agaaggctca	cattcaactcc	2640
gagaccgtgg	gagaccgggg	aaaaaaaggc	gacagaagcc	ctctcgcaag	2700
ggtcccccttc	caacacggat	tgcacgca	gtaccaagaa	cagtcacgc	2760
aatcaaagtcc	cttttcgatc	agtcataa	gacccaacaa	ggggtccatg	2820
cctacggtag	gtcccaggca	gttggctgttt	ccagagagaa	agccagcccc	2880
agtggagact	ggccatggaa	gttttcgtca	gcaaaaaatgt	ttttagggcc	2940
aacctttaaag	atccccctta	atccatata	gtttaaaaac	aatacaatgt	3000
ctttaatttg	ataaaacacgc	agtttcaaa	tttcaaggcc	tttggaggtt	3060
tgtctacctc	taaaggatc	taagoactcc	gagcgttgc	gaggagg	3120
cttggaaacaac	ttttccggag	tccggcgcta	cgccttcggc	tgcggccgac	3180
ccccctccac	ccgaaacgc	tgcgcgttcc	ggaccccttcgg	cgtcggggg	3240
ttattaaacgc	gactcgaag	tgcttttgg	cactgagggg	gtgaacacga	3300
gtggggccac	attcggccat	aaggccctta	tcttcttgc	atttgcgt	3360
gccccatggct	tcgggtcg	tttttaggc	tccggactac	aaaaatgcc	3420
cgtcacggcc	ccatcttaa	gtatgttgc	cgacgggtgg	cgtgagg	3480
tcagcacac	ctactcaaaa	ttgtggacaa	tttcttcgg	acagccgcca	3540
tgttaaaaaca	cgtgacgtat	gacgtcacgg	ccgcatttt	tgacacaaag	3600
tccttcctct	ttttcaaaaa	aaacggcaag	tgccggccg	gcggccgggg	3660

---

-continued

---

gcgcgcgcgg cccagtaggg ggagccatgc gccccccccc gcgcatgcgc ggggcccccc	3720
cccgcggggg gtcggccccc ccggccccc ccg	3753
SEQ ID NO: 17	moltype = AA length = 127
FEATURE	Location/Qualifiers
source	1..127
	mol_type = protein
	organism = Alphatorquevirus sp.
SEQUENCE: 17	
MSFWKPPVHN VTGIQRMWYE SFHRGHASF C GCGNPILHIT ALAETYGHPT GPRPSGPGV	60
DPNPHIRRAR PAPAAPEPSQ VDSRPALTWH GDGGSDGGAG GSGSGGPVAD FADDGLDQLV	120
AALDDEEE	127
SEQ ID NO: 18	moltype = AA length = 268
FEATURE	Location/Qualifiers
source	1..268
	mol_type = protein
	organism = Alphatorquevirus sp.
SEQUENCE: 18	
MSFWKPPVHN VTGIQRMWYE SFHRGHASF C GCGNPILHIT ALAETYGHPT GPRPSGPGV	60
DPNPHIRRAR PAPAAPEPSQ VDSRPALTWH GDGGSDGGAG GSGSGGPVAD FADDGLDQLV	120
AALDDEELLK TPASSPPMKY PVPVITSLEYY KSSTRGSGWDR TTRSGHGTC A DTHLAEQVLR	180
ECQNNKKLLT LYSQAQKSLG STSQNKKPK KAHHSKENR DRGRPRKKAR QKPSRKRAKR	240
SPSNSSCCSSS TKSSSSSDRE SKSSSSSS	268
SEQ ID NO: 19	moltype = AA length = 276
FEATURE	Location/Qualifiers
source	1..276
	mol_type = protein
	organism = Alphatorquevirus sp.
SEQUENCE: 19	
MSFWKPPVHN VTGIQRMWYE SFHRGHASF C GCGNPILHIT ALAETYGHPT GPRPSGPGV	60
DPNPHIRRAR PAPAAPEPSQ VDSRPALTWH GDGGSDGGAG GSGSGGPVAD FADDGLDQLV	120
AALDDEEKKK ASGRHPKTRN PRRKLTFTPK RIETVGRGR KRDRSPLARE PRGPLPTAVA	180
AAVPRAAQAO TGNQSPRLAA HKDPTRGCK PMPTVGPQW LFPERKPAPA PSSGDWAMEF	240
LAALKIFDRPV RSNLKDTPY YPYVKNQYNVY FDLKFE	276
SEQ ID NO: 20	moltype = AA length = 167
FEATURE	Location/Qualifiers
source	1..167
	mol_type = protein
	organism = Alphatorquevirus sp.
SEQUENCE: 20	
MSFWKPPVHN VTGIQRMWPK KASGRHPKTR N PRRKLTFTPK KRIETVGRDR RKRDRSPLAR	60
EPRGPLPTAV AAAVPRAAQAO QTGNQSPRLA AHKDPTRGPC KPMPTVGPQW WLFPERKPAP	120
APSSGDWAME FLAAKIFDRP VRSNLKDTPY YPYVKNQYNV YFDLKFE	167
SEQ ID NO: 21	moltype = AA length = 743
FEATURE	Location/Qualifiers
source	1..743
	mol_type = protein
	organism = Alphatorquevirus sp.
SEQUENCE: 21	
MAWGWWKRRR RWWFRKRWTR GRLRRRWPRS ARRPRRRRRV RRRRRWRGR RKTRTYRRRR	60
RFRRRGRKAK LIIKLNQPAV IKRCRKIGYI PLIISGNGTF ATNFNTSHIND RIMKGPFGGG	120
HSTMRFSLYI LFEEHLRHMN FWTSRNDNLE LTRYLGASVK IYRHPDQDFI VIYNRRTPLG	180
GNIYTAPSLSH PGNAAILAKH ILVPSLQTRP KGRKAIRLRI APPTLFTDKW YFQKDIADLT	240
LFNMIVAEAD LRFPFCSPQT DNTCISFQVL SSVNNYLSI NTFNNNDNSD KLKEFLNKAF	300
PTTGTGKTSN NALNTRTEG CISHQLKKP NPQINKPLES QYFAPLDAW GDPPIYYNDLN	360
ENKSLNNDIIE KLIKINMITY HAKLREFPNS YQGNKAFCHL TGIYSPPYLN QGRISPEIFG	420
LYTEIIYNPY TDKGTRGNKVW MDPLTKENNI YKEGQSKCLL TDMPLTWLIF GYTDWCKKDT	480
NNWDLPLNYR LVLICPYTFP KLYNEKVKDY GYIPYSYKFQ AGQMPDGSNY IPFQFRAKWY	540
PTVLHQOQVM EDISRSGPFA PKVEKPSTQL VMKYCFNFWN GGNPIIIEQIV KDPSQOPTYE	600
IPTGTGNIPRQ IQVIDPRVVLG PHYSFRSWDM RRHTFSRASI KRVSEQQETS DLVFSGPKP	660
RVDIPKQETQ EESSHSLQRE SRPWETEEES ETEALSQESQ EVPFQQQLQQ QYQEQLKLR	720
GKVLFEQLI RTQQGVHVNP CLR	743
SEQ ID NO: 22	moltype = AA length = 194
FEATURE	Location/Qualifiers
source	1..194
	mol_type = protein
	organism = Alphatorquevirus sp.
SEQUENCE: 22	
MAWGWWKRRR RWWFRKRWTR GRLRRRWPRS ARRPRRRRI VKDPSFQPTY EIPGTGNIPR	60
RQVIDPRVRL GPHYSFRSWD MRRHTFSRAS IKRVSEQQET SDLVFSGPKK PRVDIPKQET	120
QEESHHSLQR ESRPWETEEE SETEALSQES QEVPFQQQLQQ QYQEQLKLR QGIKVLFEQL	180

-continued

---

IRTQQGVHVN PCLR 194

SEQ ID NO: 23 moltype = AA length = 113  
FEATURE Location/Qualifiers  
source 1..113  
mol\_type = protein  
organism = Alphatorquevirus sp.

SEQUENCE: 23  
MAWGWWKRRR RWWFRKRWTR GRLRRRWPRS ARRRPRRRRA QKSLGSTSQN KKPKKKAHIH 60  
SKENRDRGRP RKKARQKPSR KRAKRSPSNS SCSSSTKSSS SSDRESKSSS SSS 113

SEQ ID NO: 24 moltype = length =  
SEQUENCE: 24  
000

SEQ ID NO: 25 moltype = length =  
SEQUENCE: 25  
000

SEQ ID NO: 26 moltype = length =  
SEQUENCE: 26  
000

SEQ ID NO: 27 moltype = length =  
SEQUENCE: 27  
000

SEQ ID NO: 28 moltype = length =  
SEQUENCE: 28  
000

SEQ ID NO: 29 moltype = length =  
SEQUENCE: 29  
000

SEQ ID NO: 30 moltype = length =  
SEQUENCE: 30  
000

SEQ ID NO: 31 moltype = length =  
SEQUENCE: 31  
000

SEQ ID NO: 32 moltype = length =  
SEQUENCE: 32  
000

SEQ ID NO: 33 moltype = length =  
SEQUENCE: 33  
000

SEQ ID NO: 34 moltype = length =  
SEQUENCE: 34  
000

SEQ ID NO: 35 moltype = length =  
SEQUENCE: 35  
000

SEQ ID NO: 36 moltype = length =  
SEQUENCE: 36  
000

SEQ ID NO: 37 moltype = length =  
SEQUENCE: 37  
000

SEQ ID NO: 38 moltype = length =  
SEQUENCE: 38  
000

SEQ ID NO: 39 moltype = length =  
SEQUENCE: 39  
000

SEQ ID NO: 40 moltype = length =

---

-continued

---

```

SEQUENCE: 40
000

SEQ_ID NO: 41      moltype = length =
SEQUENCE: 41
000

SEQ_ID NO: 42      moltype = length =
SEQUENCE: 42
000

SEQ_ID NO: 43      moltype = length =
SEQUENCE: 43
000

SEQ_ID NO: 44      moltype = length =
SEQUENCE: 44
000

SEQ_ID NO: 45      moltype = length =
SEQUENCE: 45
000

SEQ_ID NO: 46      moltype = length =
SEQUENCE: 46
000

SEQ_ID NO: 47      moltype = length =
SEQUENCE: 47
000

SEQ_ID NO: 48      moltype = length =
SEQUENCE: 48
000

SEQ_ID NO: 49      moltype = length =
SEQUENCE: 49
000

SEQ_ID NO: 50      moltype = length =
SEQUENCE: 50
000

SEQ_ID NO: 51      moltype = length =
SEQUENCE: 51
000

SEQ_ID NO: 52      moltype = length =
SEQUENCE: 52
000

SEQ_ID NO: 53      moltype = length =
SEQUENCE: 53
000

SEQ_ID NO: 54      moltype = DNA length = 2979
FEATURE
source          Location/Qualifiers
1..2979
mol_type = genomic DNA
organism = Betatorquevirus sp.

SEQUENCE: 54
taataaaat tcaacaggaa aaccacctaa tttaaattgc cgaccacaaa ccgtcactta 60
gttccccctt ttgcaacaac ttctgctttt ttccaactgc cggaaaacca cataatttgc 120
atggctaacc acaaactgt atgctaatta acttccacaaa aacaacttcc cctttttaaa 180
ccacacatc aaattaatta ttaaacacag tcacatcctg ggaggctacta ccacactata 240
ataccaagtg cactccgaa tggctgagtt tatgccgcta gacggagaac gcatcaggta 300
ctgactgcgg actgaacttg ggcgggtgcc gaaggtgagt gaaaccaccc aagtcaaggg 360
gcaattcggg ctatgttcagt cttagcggAAC gggcaagaaa cttaaaaatTTTttttttttt 420
caagatgacgc actgtttaa accaaacatgc tacaacaacaa aaacaaaacca aactcactgg 480
atataataacc tgcatTTAAAC ccacgacactg atctgcttct gcccaacacc aactagacac 540
ttattactag cttagcaga acaacaagaa acaatttgaag tttctaaaca agaaaaagaa 600
aaaataacaa gatgccttat tactacagaa gaagacggta caactacaga cgtcttagat 660
ggatggacg aggtttggatt agacgcctt ttcgcagaag atttcaagaaaaaggg 720
taagacctac ttatactact attcctctaa agcaatggca accggccatataaaaacat 780
gttatataaa aggacaagac tggtaataat actatagcaa cttaaagactg ggaatgaata 840
gtacaatgtt tgaaaaaaagt attgttacactg tacattggcc gggggggggg tctttttctg 900

```

-continued

---

taagcatgtt	aacttttagat	gccttgtatg	atatacataa	actttgtaga	aactggtgga	960
catccacaaa	ccaagactta	ccactagtaa	gatataaagg	atgcaaaata	acattttatc	1020
aaagcacatt	tacagactac	atagtaagaa	tacatacaga	actaccagct	aacagtaaca	1080
aactaacata	cccaaacaca	catccactaa	tgatgtatg	gtctaagtac	aaacacatta	1140
taccttagtag	acaaaacaaga	agaaaaaaga	aaccatacac	aaaaatattt	gtaaaaccac	1200
ctccgcaatt	tgaaaaaacaat	tggtactttg	ctacagactc	ctacaaaattt	ccattactac	1260
aaatacactg	cacagcatgc	aacttacaaa	accatgttg	aaaaccagac	aaattatcaa	1320
acaatgttac	attatggtc	ctaaacacca	taagcataca	aatagaacac	atgtcagtgg	1380
atcaaggaca	atcatggca	ttttaaaatac	taggaacaca	aagcttttat	ttttacttt	1440
acaccggagc	aaaccttacca	ggtgacacaa	cacaatatac	agtagcagac	ctattaccac	1500
taacaaaccc	agaataaaac	agacaggac	aatcaactaa	tgaggcaaaa	attacagacc	1560
atattacttt	cacagaatac	aaaaaaacaaat	tttacaattha	ttggggtaac	ccatttaata	1620
aacacattca	agaacaccta	gatatgatac	tatactact	aaaaagtcca	gaagcaataa	1680
aaaacgatgc	gacaacagaa	aacatgaaaat	ggaaccaat	aaacaatgca	ggaacaatgg	1740
cattaaacacc	attaacagcg	ccaaatatac	cacaatatac	ataatacc	gatagagaca	1800
caaggagaaga	cactcaatta	tacctactct	ctaacgtac	aggaacagga	ttgggaccac	1860
caggaattcc	agaattaata	ctagaaggat	ttccactatg	gttaatataat	tggggatttg	1920
cgactttca	aaaaaaaccta	aaaaaaatggaa	caaacataga	cacaatatac	atgttagtag	1980
caaaaaacaaa	atttacacaa	aaacatggca	cattctactt	agtaatacta	aatgacacact	2040
ttttagaaagg	caatagcccc	tatgaaaaaa	aacccttacc	tgaagacaac	attaaatgg	2100
acccacaagt	acaataccaa	ttagaagcac	aaaacaaact	actacaaact	gggccattha	2160
cacccaaacat	acaaggacaa	ctatcagaca	atataatcaat	gttttataaa	ttttacttta	2220
aatggggagg	aaggccacca	aaagcaatcc	atgttggaaa	tcctcccccac	cagattcaat	2280
atccccatacc	ccgttaacgag	catgaaacaa	cttcgttaca	gagtcagg	gaagccccag	2340
aatccatctt	atactcttc	gactatagac	acgggaacta	cacaacaaca	gctttgtcac	2400
gaatttagcca	agactgggc	ctttaaagaca	ctgtttctaa	aattacagag	ccagatcgac	2460
agoaactgt	caaacaaggc	ctcgaatgtc	tgc当地atcc	ggaagaaacg	caggagaaaa	2520
aagaaaaaaga	agtacagcg	ctcatcagca	acccctcagaca	gcagcagcag	ctgtacagag	2580
agogaataat	atcattatta	aaggaccaat	aactttaac	tgtgtaaaaa	aggtgaaatt	2640
gtttgatgt	aaacaaaaaa	accgttagatt	tacacctgag	gaatttggaa	ctgagttaca	2700
aatagcgat	tgggtttaaaa	gaccggccaa	atccttggta	aatgtatcctc	ccttttaccc	2760
atgggttacca	cttgc当地ctg	ttgttaaactt	taagcttaat	tttactgaat	aaaggccagc	2820
attaatttcac	ttaaggagtc	tgtttattta	agttaaaccc	taataaaacgg	tcacccgctc	2880
cctaatacgc	aggcgccagaa	agggggctcc	gccccctta	accccccagg	ggctccgccc	2940
ctgaaaccc	ccaagggggc	tacccccct	tacacccccc			2979

SEQ ID NO: 55            moltype = AA length = 99  
 FEATURE                Location/Qualifiers  
 source                1..99  
 mol\_type = protein  
 organism = Betatorquevirus sp.

SEQUENCE: 55  
 MSDCFKPTCY NNKTKQTHWI NNLHLTHDLI CFCPTPTRHL LLALAEQQET IEVSKQEKEK 60  
 ITRCLITTEE DGTTTDVLDG MDEVGLDALF AEDFEEKEFG 99

SEQ ID NO: 56            moltype = AA length = 203  
 FEATURE                Location/Qualifiers  
 source                1..203  
 mol\_type = protein  
 organism = Betatorquevirus sp.

SEQUENCE: 56  
 MSDCFKPTCY NNKTKQTHWI NNLHLTHDLI CFCPTPTRHL LLALAEQQET IEVSKQEKEK 60  
 ITRCLITTEE DGTTTDVLDG MDEVGLDALF AEDFEEKEFG NIPYPVTSMK QLRYRVQGKP 120  
 QNPSYTPSTI DTGTTQQQLC HELAKTGHHLK TLFLKLQSQI DSNCSNKPSN ACKSRKKRR 180  
 KKKKKYSSS ATSDSSSSCT ESE 203

SEQ ID NO: 57            moltype = AA length = 219  
 FEATURE                Location/Qualifiers  
 source                1..219  
 mol\_type = protein  
 organism = Betatorquevirus sp.

SEQUENCE: 57  
 MSDCFKPTCY NNKTKQTHWI NNLHLTHDLI CFCPTPTRHL LLALAEQQET IEVSKQEKEK 60  
 ITRCLITTEE DGTTTDVLDG MDEVGLDALF AEDFEEKEGA RSTATAQTSP RMPANLGRNA 120  
 GEKRKRSTAA HQQPQTAAAAA VQRANNIIIK GPITFNCVKK VKLFDDKPKN RRFTPEEFET 180  
 ELQIAKWLKR PPRSFVNNDPP FYPWLPPEPV VNFKLNFT 219

SEQ ID NO: 58            moltype = AA length = 666  
 FEATURE                Location/Qualifiers  
 source                1..666  
 mol\_type = protein  
 organism = Betatorquevirus sp.

SEQUENCE: 58  
 MPYYYYRRRR NYRRRPRWYGR GWIRRPFRRR FRRKRRVRPT YTTIPLKQWQ PPYKRTCYIK 60  
 GQDCLIYYSN LRLGMNSTMY EKSIVPVHWP GGGSFVSML TLDALYDIHK LCRNWWTSTN 120  
 QDLPLVRYKG CKITFYQSTF TDYIVRIHTE LPANSNKLY PNTHPLMMMM SKYKHIIPSR 180

---

-continued

---

QTRRKKKPYT	KIFVKPPPQF	ENKWKYFATDL	YKIPLLQIHC	TACNLQNPFV	KPDKLSNNVT	240
LWSLNTISIQ	NNRMSVDQGQ	SWPFKILGTQ	SFYFYFYTGA	NLPGDTTQIP	VADLLPLTNP	300
RINRPQQLN	EAKITDHITF	TEYKNKFTNY	WGNPFNKHIQ	EHLDIMLYSL	KSPEAIKNEW	360
TTENMKWNL	MNAGTMALTP	FNEPIFTQIQ	YNPDRDTGED	TQLYLLSNAT	GTGWDPPGIP	420
ELILEGFPLW	LIYWGFAFDQ	KNLKKVTVNID	TNYMLVAKTK	FTQKPGTFYL	VILNDTFVEG	480
NSPYEKQPLP	EDNIKWYPQV	QYOLEAQNL	LQTGPFTPNI	QGQLSDNISM	FYKFYFKWGG	540
SPPKAINVEN	PAHQIQYPIP	RNEHETTSLO	SPGEAPESEL	YSFDYRHGNY	TTTALSRSQ	600
DWALKDTVSK	ITEPDRQQLL	KQALECLQIS	EETQEKKKE	VQQLISNLRQ	QQQLYRERII	660
SLLKDQ						666
 SEQ ID NO: 59		moltype = AA	length = 148			
FEATURE		Location/Qualifiers				
source		1..148				
		mol_type = protein				
		organism = Betatorquevirus sp.				
SEQUENCE: 59						
MPYYYYRRRRY	NYRRPRWYGR	GWIRRPFRRR	FRRKRRIQYP	IPRNEHETTS	LQSPGEAPES	60
ILYSFDRYRGH	NYTTTALSRI	SQDWALKDTV	SKITEPDRQQ	LLKQALECLQ	ISETQEKKKE	120
KEVQQLISNL	RQQQQLYRER	IISLLKDQ				148
 SEQ ID NO: 60		moltype = AA	length = 82			
FEATURE		Location/Qualifiers				
source		1..82				
		mol_type = protein				
		organism = Betatorquevirus sp.				
SEQUENCE: 60						
MPYYYYRRRRY	NYRRPRWYGR	GWIRRPFRRR	FRRKRRIQYP	IPRNEHETTS	LQSPGEAPES	60
KKKKYSSSSA	TSDSSSSCTE	SE				82
 SEQ ID NO: 61		moltype =	length =			
SEQUENCE: 61						
000						
 SEQ ID NO: 62		moltype =	length =			
SEQUENCE: 62						
000						
 SEQ ID NO: 63		moltype =	length =			
SEQUENCE: 63						
000						
 SEQ ID NO: 64		moltype =	length =			
SEQUENCE: 64						
000						
 SEQ ID NO: 65		moltype =	length =			
SEQUENCE: 65						
000						
 SEQ ID NO: 66		moltype =	length =			
SEQUENCE: 66						
000						
 SEQ ID NO: 67		moltype =	length =			
SEQUENCE: 67						
000						
 SEQ ID NO: 68		moltype =	length =			
SEQUENCE: 68						
000						
 SEQ ID NO: 69		moltype =	length =			
SEQUENCE: 69						
000						
 SEQ ID NO: 70		moltype =	length =			
SEQUENCE: 70						
000						
 SEQ ID NO: 71		moltype =	length =			
SEQUENCE: 71						
000						
 SEQ ID NO: 72		moltype =	length =			
SEQUENCE: 72						
000						

---

-continued

---

```
SEQ ID NO: 73      moltype = length =
SEQUENCE: 73
000

SEQ ID NO: 74      moltype = length =
SEQUENCE: 74
000

SEQ ID NO: 75      moltype = length =
SEQUENCE: 75
000

SEQ ID NO: 76      moltype = length =
SEQUENCE: 76
000

SEQ ID NO: 77      moltype = length =
SEQUENCE: 77
000

SEQ ID NO: 78      moltype = length =
SEQUENCE: 78
000

SEQ ID NO: 79      moltype = length =
SEQUENCE: 79
000

SEQ ID NO: 80      moltype = length =
SEQUENCE: 80
000

SEQ ID NO: 81      moltype = length =
SEQUENCE: 81
000

SEQ ID NO: 82      moltype = length =
SEQUENCE: 82
000

SEQ ID NO: 83      moltype = length =
SEQUENCE: 83
000

SEQ ID NO: 84      moltype = length =
SEQUENCE: 84
000

SEQ ID NO: 85      moltype = length =
SEQUENCE: 85
000

SEQ ID NO: 86      moltype = length =
SEQUENCE: 86
000

SEQ ID NO: 87      moltype = length =
SEQUENCE: 87
000

SEQ ID NO: 88      moltype = length =
SEQUENCE: 88
000

SEQ ID NO: 89      moltype = length =
SEQUENCE: 89
000

SEQ ID NO: 90      moltype = length =
SEQUENCE: 90
000

SEQ ID NO: 91      moltype = length =
SEQUENCE: 91
000
```

---

-continued

---

```
SEQ ID NO: 92      moltype = length =
SEQUENCE: 92
000

SEQ ID NO: 93      moltype = length =
SEQUENCE: 93
000

SEQ ID NO: 94      moltype = length =
SEQUENCE: 94
000

SEQ ID NO: 95      moltype = length =
SEQUENCE: 95
000

SEQ ID NO: 96      moltype = length =
SEQUENCE: 96
000

SEQ ID NO: 97      moltype = length =
SEQUENCE: 97
000

SEQ ID NO: 98      moltype = length =
SEQUENCE: 98
000

SEQ ID NO: 99      moltype = length =
SEQUENCE: 99
000

SEQ ID NO: 100     moltype = length =
SEQUENCE: 100
000

SEQ ID NO: 101     moltype = length =
SEQUENCE: 101
000

SEQ ID NO: 102     moltype = length =
SEQUENCE: 102
000

SEQ ID NO: 103     moltype = length =
SEQUENCE: 103
000

SEQ ID NO: 104     moltype = length =
SEQUENCE: 104
000

SEQ ID NO: 105      moltype = DNA length = 71
FEATURE
source
1..71
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 105
cgggtgcggk aggtgagttt acacaccgma gtcaaggggc aattcgggct crggactggc 60
cgggcyhtgg g                                71

SEQ ID NO: 106      moltype = DNA length = 71
FEATURE
source
1..71
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 106
cgggtgcggg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60
cgggctwtgg g                                71

SEQ ID NO: 107      moltype = DNA length = 71
FEATURE
source
1..71
mol_type = other DNA
organism = synthetic construct
```

-continued

---

SEQUENCE: 107  
cgggtgcggc aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
cgggctatgg g 71

SEQ ID NO: 108 moltype = DNA length = 71  
FEATURE Location/Qualifiers  
source 1..71  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 108  
cgggtgcggc aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
cgggcccctgg g 71

SEQ ID NO: 109 moltype = DNA length = 71  
FEATURE Location/Qualifiers  
source 1..71  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 109  
cgggtgcggc aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
cgggctttgg g 71

SEQ ID NO: 110 moltype = DNA length = 71  
FEATURE Location/Qualifiers  
source 1..71  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 110  
cgggtgcggc aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
cgggctatgg g 71

SEQ ID NO: 111 moltype = DNA length = 71  
FEATURE Location/Qualifiers  
source 1..71  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 111  
cgggtgcggc aggtgagttt acacaccgaa gtcaaggggc aattcgggct caggactggc 60  
cgggctttgg g 71

SEQ ID NO: 112 moltype = DNA length = 71  
FEATURE Location/Qualifiers  
source 1..71  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 112  
cgggtgcggc aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
cgggcyhtgg g 71

SEQ ID NO: 113 moltype = DNA length = 71  
FEATURE Location/Qualifiers  
source 1..71  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 113  
cgggtgcggc aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
cgggctatgg g 71

SEQ ID NO: 114 moltype = DNA length = 70  
FEATURE Location/Qualifiers  
source 1..70  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 114  
cgggtgcggc aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
cgggccccggg 70

SEQ ID NO: 115 moltype = DNA length = 71  
FEATURE Location/Qualifiers  
source 1..71  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 115  
cgggtgcggc aggtgagttt acacaccgaa gtcaaggggc aattcgggct caggactggc 60  
cgggctttgg g 71

---

-continued

---

```

SEQ ID NO: 116      moltype = DNA length = 69
FEATURE
source
1..69
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 116
cgggtgcggg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggaggccc 60
ggccatggg                                     69

SEQ ID NO: 117      moltype = DNA length = 71
FEATURE
source
1..71
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 117
cgggtgcggg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60
cgggccccgg g                                     71

SEQ ID NO: 118      moltype = DNA length = 71
FEATURE
source
1..71
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 118
cgggtgcggg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60
cgggctatgg g                                     71

SEQ ID NO: 119      moltype = DNA length = 71
FEATURE
source
1..71
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 119
cgggtgcga aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60
cgggctatgg g                                     71

SEQ ID NO: 120      moltype = DNA length = 117
FEATURE
misc_difference
10
note = May or may not be present
12
note = May or may not be present
30..32
note = May or may not be present
34
note = May or may not be present
43..46
note = May or may not be present
52..54
note = May or may not be present
70..71
note = May or may not be present
89..90
note = May or may not be present
103
note = May or may not be present
1..117
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 120
cggcggsgsgs gssscgcgct dcgcgcgcsg cccrsyrggg grdssmmwgc skscCCCCC 60
cscgcgcatg cgcrccggkc ccccccccyv sggggggctc cgcccccccg gcccccc 117

SEQ ID NO: 121      moltype = DNA length = 169
FEATURE
misc_difference
20
note = a, c, t, g, unknown or other
22
note = a, c, t, g, unknown or other
40..42
note = a, c, t, g, unknown or other
53..56
note = a, c, t, g, unknown or other
62
note = a, c, t, g, unknown or other
64

```

---

---

-continued

---

```

misc_difference      note = a, c, t, g, unknown or other
                     97..98
source              note = a, c, t, g, unknown or other
                     1..169
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 121
gccccccggg cggcggsnn gnsgcgcgt dgcgcgcsn nnccrcrccc ggnnnncwgc 60
snccccccc cccgcgcatg cgcggkccc ccccccnncc gggggctcg ccccccggcc 120
ccccccctgt cttaaaccac cgcgcatgcg cgaccacgcc cccgcccc 169

SEQ ID NO: 122      moltype = DNA length = 79
FEATURE
misc_difference      Location/Qualifiers
                     20
                     note = a, c, t, g, unknown or other
                     22
                     note = a, c, t, g, unknown or other
                     40..42
                     note = a, c, t, g, unknown or other
                     53..56
                     note = a, c, t, g, unknown or other
                     62
                     note = a, c, t, g, unknown or other
                     64
                     note = a, c, t, g, unknown or other
                     1..79
                     source
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 122
gccccccggg cggcggsnn gnsgcgcgt dgcgcgcsn nnccrcrccc ggnnnncwgc 60
snccccccc cccgcgcatg 79

SEQ ID NO: 123      moltype = DNA length = 31
FEATURE
misc_difference      Location/Qualifiers
                     18..19
                     note = a, c, t, g, unknown or other
                     1..31
                     source
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 123
gccccggkcc ccccccnncc gggggctcc g 31

SEQ ID NO: 124      moltype = DNA length = 59
FEATURE
source              Location/Qualifiers
                     1..59
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 124
ccccccggc ccccccctgt cttaaaccac cgcgcatgcg cgaccacgcc cccgcccc 59

SEQ ID NO: 125      moltype = DNA length = 156
FEATURE
source              Location/Qualifiers
                     1..156
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 125
gccccgggg ggccggcccg ttcgcgcgc gcccaccagg ggggtgtcg cgccccccc 60
cgcgcatgcg cggggcccccc ccccgcccccc gctccgcggg cccggcccccc ccccggtcta 120
aaccacccgc gcatgcgcga ccacgcccc gccgcc 156

SEQ ID NO: 126      moltype = length =
SEQUENCE: 126
000

SEQ ID NO: 127      moltype = length =
SEQUENCE: 127
000

SEQ ID NO: 128      moltype = length =
SEQUENCE: 128
000

SEQ ID NO: 129      moltype = DNA length = 25
FEATURE
source              Location/Qualifiers
                     1..25
                     mol_type = other DNA

```

---

-continued

---

SEQUENCE: 129	organism = synthetic construct
ttcgcgcgcc gcccaccagg gggtg	25
SEQ ID NO: 130	moltype = length =
SEQUENCE: 130	
000	
SEQ ID NO: 131	moltype = DNA length = 17
FEATURE	Location/Qualifiers
source	1..17
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 131	
cgcgcgcgcg cgcgcat	17
SEQ ID NO: 132	moltype = DNA length = 17
FEATURE	Location/Qualifiers
source	1..17
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 132	
gwgccggggcc ccccccc	17
SEQ ID NO: 133	moltype = DNA length = 72
FEATURE	Location/Qualifiers
source	1..72
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 133	
ggggggggctc cgcccccccg gccccccccc gtgtctaaacc caccgcgcac ggcgcaccac	60
gccccccgcg cc	72
SEQ ID NO: 134	moltype = DNA length = 115
FEATURE	Location/Qualifiers
source	1..115
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 134	
cgccgggggc ggcgcgcgcg ctgcgcgcgc gcgcgggggg ggccgcagcg cccccccccc	60
cgccgcgcgcg cgggtccccc ccccaacggg gggctccgc ccccgccccc ccccc	115
SEQ ID NO: 135	moltype = DNA length = 14
FEATURE	Location/Qualifiers
source	1..14
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 135	
cggccggggc ggcgcg	14
SEQ ID NO: 136	moltype = DNA length = 17
FEATURE	Location/Qualifiers
source	1..17
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 136	
cgcgcgtgc ggcgcgcg	17
SEQ ID NO: 137	moltype = DNA length = 19
FEATURE	Location/Qualifiers
source	1..19
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 137	
cgcggggggg ggcgcagcg	19
SEQ ID NO: 138	moltype = DNA length = 17
FEATURE	Location/Qualifiers
source	1..17
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 138	
ccccccccc cgcgcat	17
SEQ ID NO: 139	moltype = DNA length = 31
FEATURE	Location/Qualifiers

---

-continued

---

```

source          1..31
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 139
gcacgggtcc ccccccccac ggggggctcc g                                31

SEQ ID NO: 140      moltype = DNA  length = 17
FEATURE
source          1..17
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 140
ccccccggcc cccccc                                17

SEQ ID NO: 141      moltype = DNA  length = 121
FEATURE
source          1..121
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 141
ccgtcggcg gggggccgca cgctgcgcgc gcggccccc ggggaggcac agcctcccc 60
cccgcgcgca atgcgcgcgg gtcccccctt ctcggggggg ctccgcccccc cggccccc 120
c                                         121

SEQ ID NO: 142      moltype = DNA  length = 37
FEATURE
source          1..37
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 142
ccgtcggcg gggggccgca cgctgcgcgc gcggccc                                37

SEQ ID NO: 143      moltype = DNA  length = 84
FEATURE
source          1..84
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 143
ccgggggagg cacagectcc ccccccgcg cgcattgcgcg cgggtccccccc ccctccgg 60
gggtccggcc ccccgcccccc cccc                                84

SEQ ID NO: 144      moltype = DNA  length = 104
FEATURE
source          1..104
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 144
ccgcggcgcc ggcgcgcgta cgcgcgcgcg cgggggggtt ggcgcccccc cccgcgcatt 60
ggcgggggcc ccccccgcg ggggctccg ccccgcccccc cccc                                104

SEQ ID NO: 145      moltype = DNA  length = 11
FEATURE
source          1..11
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 145
ccgcggcgcc g                                11

SEQ ID NO: 146      moltype = DNA  length = 17
FEATURE
source          1..17
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 146
cgcgcgctac ggcgcgcg                                17

SEQ ID NO: 147      moltype = DNA  length = 10
FEATURE
source          1..10
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 147
cgccgggggg                                10

SEQ ID NO: 148      moltype = length =
SEQUENCE: 148

```

---

-continued

---

```

000

SEQ ID NO: 149      moltype = DNA  length = 15
FEATURE
source
1..15
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 149
cccccccccgcgcat                                              15

SEQ ID NO: 150      moltype = DNA  length = 17
FEATURE
source
1..17
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 150
gcgcgggggcc cccccc                                              17

SEQ ID NO: 151      moltype = DNA  length = 13
FEATURE
source
1..13
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 151
gcggggggcgt ccg                                              13

SEQ ID NO: 152      moltype = DNA  length = 14
FEATURE
source
1..14
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 152
ccccccggcc cccc                                              14

SEQ ID NO: 153      moltype = DNA  length = 122
FEATURE
source
1..122
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 153
gcgcgcgcgg cggcgggggg cgccgcgtg cgccgcgcgc ccagtagggg gagccatgc 60
cccccccccgcgcatgcgccc gggccccccc ccgcgggggg ctccgcggcc cggccccccc 120
cg                                                               122

SEQ ID NO: 154      moltype = DNA  length = 19
FEATURE
source
1..19
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 154
gcgcgcgcgg cggcgggggg                                              19

SEQ ID NO: 155      moltype = DNA  length = 41
FEATURE
source
1..41
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 155
gcgcgcgcgt gcgcgcgcgc ccagtaggg ggagccatgc g                                              41

SEQ ID NO: 156      moltype = DNA  length = 15
FEATURE
source
1..15
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 156
cccccccccgcgcat                                              15

SEQ ID NO: 157      moltype = DNA  length = 17
FEATURE
source
1..17
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 157
gcgcggggcc cccccc                                              17

```

---

-continued

---

SEQ ID NO: 158	moltype = DNA length = 13
FEATURE	Location/Qualifiers
source	1..13
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 158	
gcggggggct ccg	13
SEQ ID NO: 159	moltype = DNA length = 17
FEATURE	Location/Qualifiers
source	1..17
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 159	
ccccccggcc ccccccgg	17
SEQ ID NO: 160	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 160	
cgcgctgcgc ggcggccca gtagggggag ccatgc	36
SEQ ID NO: 161	moltype = DNA length = 78
FEATURE	Location/Qualifiers
source	1..78
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 161	
ccggccatctt aagttagttga ggccggacggt ggccgtgagtt caaaggcac catcagccac	60
acctactcaa aatggtgg	78
SEQ ID NO: 162	moltype = DNA length = 172
FEATURE	Location/Qualifiers
source	1..172
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 162	
cttaaaggtagt tgaggcgac ggtggcgtga gttcaaaggat caccatcagc cacacctact	60
caaaaatggt gacaatttct tccgggtcaa aggttacagc cgccatgtta aaacacgtga	120
cgttatgacgt cacggccgc attttgtac acaagatggc cgacttcctt cc	172
SEQ ID NO: 163	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 163	
cgcgctgcgc ggcggccca gtagggggag ccatgc	36
SEQ ID NO: 164	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 164	
gcgcgtcgcg cgccgcgcgg ggggtgcgc cccccc	36
SEQ ID NO: 165	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 165	
gcgcgtcgcg cgccgcgcac tagggggcgt tgcgcg	36
SEQ ID NO: 166	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 166	
gcgcgtcgcg cgccgcgcag tagggggcgc aatgcg	36
SEQ ID NO: 167	moltype = DNA length = 36

---

-continued

---

FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 167	
gcccgtgcgcg cgccggcccc gggggaggca ttgcct	36
SEQ ID NO: 168	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 168	
gcccgtgcgcg cgccgcgcgg gggggcgcca gcgcgg	36
SEQ ID NO: 169	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 169	
gcccgttcgcg cgccgcgcgg ggggtccgc cccccc	36
SEQ ID NO: 170	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 170	
gcccgttcgcg cgccgcgcgg ggggtgcgc cccccc	36
SEQ ID NO: 171	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 171	
gcccgtacgcg cgccgcgcgg ggggtgcgc cccccc	36
SEQ ID NO: 172	moltype = DNA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 172	
gcccgtacgcg cgccgcgcgg ggggtctgc cccccc	36
SEQ ID NO: 173	moltype = length =
SEQUENCE: 173	000
SEQ ID NO: 174	moltype = length =
SEQUENCE: 174	000
SEQ ID NO: 175	moltype = length =
SEQUENCE: 175	000
SEQ ID NO: 176	moltype = length =
SEQUENCE: 176	000
SEQ ID NO: 177	moltype = length =
SEQUENCE: 177	000
SEQ ID NO: 178	moltype = length =
SEQUENCE: 178	000
SEQ ID NO: 179	moltype = length =
SEQUENCE: 179	000
SEQ ID NO: 180	moltype = length =

---

-continued

---

```

SEQUENCE: 180
000

SEQ ID NO: 181      moltype = length =
SEQUENCE: 181
000

SEQ ID NO: 182      moltype = length =
SEQUENCE: 182
000

SEQ ID NO: 183      moltype = length =
SEQUENCE: 183
000

SEQ ID NO: 184      moltype = length =
SEQUENCE: 184
000

SEQ ID NO: 185      moltype = AA length = 743
FEATURE           Location/Qualifiers
source            1..743
                  mol_type = protein
                  organism = Alphatorquevirus sp.

SEQUENCE: 185
MAWGWWKRRR RWWFRKRWTR GRLRRRWPRS ARRRP RRRRWRGR RKTRTYRRR 60
RFFRRGKAK LIIKLWQPAV IKRCRIKG YIPLISGNG TFIATNFTSHI NDRIMKGPF 120
HSTMRFSLYI LFEEHLRHMD FWTRSNDLLE LTRYLGASVK IYRHPDQDFI VIYNRRTPLG 180
GNIYTAPSLH PGNAILAKHK ILVPSLQTRP KGRKAIRLRI APPTLFTDKW YFQKDIADLT 240
LFNIMAVEAD LRFPFCSPQT DNTCISFQVL SSVYNNYLSI NTFNNNDNSDS KLKEFLNKAF 300
PTTGTGKTSN NALNTFRTEG CISHPQLKKP NPQINKPLES QYFAPLDALW GDPIYYNDLN 360
ENKSLNDIIE KILIKNMITY HAKLREFPNS YQGNKAFCHL TGIYSPPYLQ QGRISPEIFG 420
LYTEIIYNPY TDKGIGNKVW MDPLTKENNI YKEGQSKCLL TDMPLWTLF GYTDWCKD 480
NNWDLPLNYR LVLICPYTFP KLYNEKVKDY GYIPYSYKFG AGQMPDGSNY IPFQFRAKY 540
PTVLHQQQVM EDISRGPFPA PKVEKPSTQL VMKYCFCNFNW GGNPIIEQIV KDPSFQPTYE 600
IPGTGNIPRR IQVIDPRVLG PHYSFRSDM RRHTFSRASI KRVSEQQETS DLVFSGPKKP 660
RVDIPKQETQ EESSHSLORE SRPWETEEES ETEALSQESQ EVPFQQQLQQ QYQEQLKLRQ 720
GIKVLFEQLI RTQQGVHVNP CLR                                743

SEQ ID NO: 186      moltype = AA length = 68
FEATURE           Location/Qualifiers
source            1..68
                  mol_type = protein
                  organism = Alphatorquevirus sp.

SEQUENCE: 186
MAWGWWKRRR RWWFRKRWTR GRLRRRWPRS ARRRP RRRRWRGR RKTRTYRRR 60
RFFRRGKAK

SEQ ID NO: 187      moltype = AA length = 212
FEATURE           Location/Qualifiers
source            1..212
                  mol_type = protein
                  organism = Alphatorquevirus sp.

SEQUENCE: 187
AKLIILWQP AVIKRCRIKG YIPLISGNG TFATNFTSHI NDRIMKGPF GGHSTMRFL 60
YILFEEHLRH MNFWTRSDN LELTRYLGAS VKIYRHPDQD FIVIYNRRTPLG 120
LHPGNAILAK HKILVPSLQT RPKGRKAIRL RIAPTLFTDKW YFQKDIAD LTLFNIMAVE 180
ADLRFPFCSP QTDNTCISFQ VLSSVYNNYL SI                                212

SEQ ID NO: 188      moltype = AA length = 133
FEATURE           Location/Qualifiers
source            1..133
                  mol_type = protein
                  organism = Alphatorquevirus sp.

SEQUENCE: 188
NTFNNNDNSDS KLKEFLNKAF PTTGTGKTSN NALNTFRTEG CISHPQLKKP NPQINKPLES 60
QYFAPLDALW GDPIYYNDLN ENKSLNDIIE KILIKNMITY HAKLREFPNS YQGNKAFCHL 120
TGIYSPPYLQ QGRISPEIFG 133

SEQ ID NO: 189      moltype = AA length = 166
FEATURE           Location/Qualifiers
source            1..166
                  mol_type = protein
                  organism = Alphatorquevirus sp.

SEQUENCE: 189
ISPEIFGLYT EIIYNPYTDK GTGNKVMMDP LTKENNIYKE GQSKCLLTD PLWTLLFGYT 60

```

-continued

---

DWCKKDTNNW DLPLNYRLVL ICPYTFPKLY NEKVKDYGYI PYSYKFGAGQ MPDGSNYIPF 120  
QFRAKWYPTV LHQQQVMEDI SRSGPFAPKV EKYSTQLVMK YCFNFN 166

SEQ ID NO: 190 moltype = AA length = 164  
FEATURE Location/Qualifiers  
source 1..164  
mol\_type = protein  
organism = Alphatorquevirus sp.

SEQUENCE: 190  
WGGNPPIIEQI VKDPSFQPTY EIPGTGNIPR RIQVIDPRVL GPHYSFRSWD MRRHTFSRAS 60  
IKRVSEQQET SDLVFSGPKK PRVDIPKQET QEESSHSLQR ESRPWETEE SETEALSQES 120  
QEVPFQQQLQ QQYQEQLKLR QGIKVLFEQL IRTQQGVHVN PCLR 164

SEQ ID NO: 191 moltype = length =  
SEQUENCE: 191  
000

SEQ ID NO: 192 moltype = length =  
SEQUENCE: 192  
000

SEQ ID NO: 193 moltype = length =  
SEQUENCE: 193  
000

SEQ ID NO: 194 moltype = length =  
SEQUENCE: 194  
000

SEQ ID NO: 195 moltype = length =  
SEQUENCE: 195  
000

SEQ ID NO: 196 moltype = length =  
SEQUENCE: 196  
000

SEQ ID NO: 197 moltype = length =  
SEQUENCE: 197  
000

SEQ ID NO: 198 moltype = length =  
SEQUENCE: 198  
000

SEQ ID NO: 199 moltype = length =  
SEQUENCE: 199  
000

SEQ ID NO: 200 moltype = length =  
SEQUENCE: 200  
000

SEQ ID NO: 201 moltype = length =  
SEQUENCE: 201  
000

SEQ ID NO: 202 moltype = length =  
SEQUENCE: 202  
000

SEQ ID NO: 203 moltype = length =  
SEQUENCE: 203  
000

SEQ ID NO: 204 moltype = length =  
SEQUENCE: 204  
000

SEQ ID NO: 205 moltype = length =  
SEQUENCE: 205  
000

SEQ ID NO: 206 moltype = length =  
SEQUENCE: 206  
000

---

-continued

---

```

SEQ ID NO: 207      moltype =   length =
SEQUENCE: 207
000

SEQ ID NO: 208      moltype =   length =
SEQUENCE: 208
000

SEQ ID NO: 209      moltype =   length =
SEQUENCE: 209
000

SEQ ID NO: 210      moltype =   length =
SEQUENCE: 210
000

SEQ ID NO: 211      moltype =   length =
SEQUENCE: 211
000

SEQ ID NO: 212      moltype =   length =
SEQUENCE: 212
000

SEQ ID NO: 213      moltype =   length =
SEQUENCE: 213
000

SEQ ID NO: 214      moltype =   length =
SEQUENCE: 214
000

SEQ ID NO: 215      moltype = AA  length = 666
FEATURE
source
1..666
mol_type = protein
organism = Betatorquevirus sp.
SEQUENCE: 215
MPYYYRRRY NYRRPRWYGR GWIRRPFRRR FRRKRRVRPRT YTTIPLKQWQ PPYKRTCYIK 60
GQDCLIIYYSN LRLGMNSTMY EKSIVPVHWP GGGSFSVSML TLDALYDIHK LCRNWWTSTN 120
QDLPLVRYKG CKITFYQSTF TDYIVRIHTE LPANSNLITY PNTHPPLMMMM SKYKHIIPSR 180
QTRRKKKPYT KIFVKPPPPQ ENWKWFATDYL YKIPLLQIHC TACNLQNPVFK KPDKLSNNVT 240
LWSLNNTISIQ NRNMSVDDQGQ SWPKFILGTQ SFYFYFYTGNA NLPGDTQIP VADILPLTNP 300
RINRPGQSLN EAKITDHITF TEYKNKFTNY WGPNPFNKHIQ EHLDMLILYSL KSPEAIKNEW 360
TTENMKWNQL NNAGTMALTP FNEPIFTQIQ YNPDRDTGED TQLYLLSNAT GTGWDPPGIP 420
ELILEGFPWL LIYWGPAFDFQ KNLKKVVTNID TNYMLVAKTK FTQKPGTFYL VILNDTFVEG 480
NSPYEKQPLW EDNIKWYPQV QYQLEAQNLQ LQTGPFTPNI QGQLSDNISM FYKFYFKWGG 540
SPPKAINVEN PAHQIQYPIP RNEHETTSLQ SPGEAPESEL YSFDYRHGNY TTTALSRSIQ 600
DWALKDTVSK ITEPDRQOLL KQALECLQIS EETQEKKKE VQQLISNLRQ QQQLYRERII 660
SLLKDQ                                     666

SEQ ID NO: 216      moltype = AA  length = 38
FEATURE
source
1..38
mol_type = protein
organism = Betatorquevirus sp.
SEQUENCE: 216
MPYYYRRRY NYRRPRWYGR GWIRRPFRRR FRRKRRVRPRT                                38
PTYTTIPLKQ WQPPYKRTCY IKGQDCLIYY SNLRLGMNST MYEKSIVPVH WPGGGSFSVS 60
MLTLDALYDI HKLCRNWWTS TNQDPLVRY KGCKITFYQS TFTDYIVRIH TELPANSNKL 120
TYPNTHPLMM MMSKYKHIIP SRQTRKKKP YTKIFVKPPP QFENKWFAT DLYKIPLLQI 180
HCTACNLQNP FVKPDKLSNN VTLWSLNT                                         208

SEQ ID NO: 218      moltype = AA  length = 128
FEATURE
source
1..128
mol_type = protein
organism = Betatorquevirus sp.

```

---

-continued

---

```

SEQUENCE: 218
ISIQNQRNMSV DQGQSWPFKI LGTQSFYFYF YTGANLPGDT TQIPVADLLP LTNPRINRPG 60
QSLNEAKITD HITFTEYKNK FTNYWGNPNF KHIQEHLDMI LYSLKSPEAI KNEWTENMK 120
WNQLNNAG 128

SEQ ID NO: 219      moltype = AA length = 163
FEATURE          Location/Qualifiers
source           1..163
                 mol_type = protein
                 organism = Betatorquevirus sp.

SEQUENCE: 219
TMALTPFNEP IFTQIQQYNPD RDTGEDTQLY LLSNATGTGW DPPGIPELIL EGFPLWLIW 60
GFADFQKNLK KVTNIDTNYM LVAKTKFTQK PGTFYLVILN DTFVEGNSPY EKQPLPEDNI 120
KWYPQVQYQL EAQNKLQGT PFTPNIQGQL SDNISMFYKF YFK 163

SEQ ID NO: 220      moltype = AA length = 129
FEATURE          Location/Qualifiers
source           1..129
                 mol_type = protein
                 organism = Betatorquevirus sp.

SEQUENCE: 220
WGGSPPKAIN VENPAHQIQY PIPRNEHETT SLOSPGEAPE SILYSFDYRH GNYTTTALSR 60
ISQDWALKDT VSKITEPDRQ QLLKQALECL QISEETQEKK EKEVQQLISN LRQQQOLYRE 120
RIISLLKDQ 129

SEQ ID NO: 221      moltype = length =
SEQUENCE: 221
000

SEQ ID NO: 222      moltype = length =
SEQUENCE: 222
000

SEQ ID NO: 223      moltype = length =
SEQUENCE: 223
000

SEQ ID NO: 224      moltype = length =
SEQUENCE: 224
000

SEQ ID NO: 225      moltype = length =
SEQUENCE: 225
000

SEQ ID NO: 226      moltype = length =
SEQUENCE: 226
000

SEQ ID NO: 227      moltype = AA length = 220
FEATURE          Location/Qualifiers
VARIANT          29..31
                 note = Any amino acid
SITE             29..31
                 note = This region may encompass 0-3 residues
VARIANT          100
                 note = Any amino acid
VARIANT          125..129
                 note = Any amino acid
SITE             125..129
                 note = This region may encompass 1-5 residues
VARIANT          181
                 note = Any amino acid
VARIANT          211
                 note = Any amino acid
source           1..220
                 mol_type = protein
                 organism = synthetic construct
REGION           1..220
                 note = Description of Artificial Sequence: Synthetic
                           polypeptide

SEQUENCE: 227
LVLTQWQPNT VRRCYIRGYL PLIICGENXX XTTSRNYATH SDDTIQKGPF GGGMSTTFS 60
LRVLYDEYQR FMNRWTYSNE DLDLARYLGC KFTFYRHPDX DFIVQYNTNP PFKDTKLTA 120
SIHPXXXXXG MLMLSKRKIL IPSLKTRPKG KHYVKVRIGP PKLFEDKWTY QSDLCDVPLV 180
XYNATAADLQ HPFGSPQTDN PCVTFQVLGS XYNKHLISISP 220

```

---

-continued

---

```

SEQ ID NO: 228      moltype = AA length = 172
FEATURE          Location/Qualifiers
VARIANT          38
                  note = Any amino acid
VARIANT          44..46
                  note = Any amino acid
SITE             44..46
                  note = This region may encompass 0-3 residues
VARIANT          77
                  note = Any amino acid
VARIANT          79
                  note = Any amino acid
VARIANT          98..101
                  note = Any amino acid
SITE             98..101
                  note = This region may encompass 0-4 residues
source            1..172
                  mol_type = protein
                  organism = synthetic construct
REGION            1..172
                  note = Description of Artificial Sequence: Synthetic
                  polypeptide

SEQUENCE: 228
SNFEFPGAYT DITYNPLTDK GVGNNMVIQY LTKPDTIXDK TQSXXXKCLI EDLPLWAALY 60
GYVDFCEKET GDSAIIXNXG RVLIRCPYTK PPLYDKTXXX XNKGFVPYST NFGNGKMPGG 120
SGYVPIYWRA RWYPTLFHQK EVLEDIVQSG PFAVKDEKPS TQLVMKYCFN FN        172

SEQ ID NO: 229      moltype = AA length = 258
FEATURE          Location/Qualifiers
VARIANT          20..22
                  note = Any amino acid
SITE             20..22
                  note = This region may encompass 0-3 residues
VARIANT          25
                  note = Any amino acid
VARIANT          78
                  note = Any amino acid
VARIANT          89
                  note = Any amino acid
VARIANT          91
                  note = Any amino acid
VARIANT          95..98
                  note = Any amino acid
SITE             95..98
                  note = This region may encompass 1-4 residues
VARIANT          107..120
                  note = Any amino acid
SITE             107..120
                  note = This region may encompass 2-14 residues
VARIANT          129
                  note = Any amino acid
VARIANT          139..168
                  note = Any amino acid
SITE             139..168
                  note = This region may encompass 0-30 residues
VARIANT          201..204
                  note = Any amino acid
SITE             201..204
                  note = This region may encompass 0-4 residues
SITE             219..258
                  note = This region may encompass 0-40 residues
VARIANT          219..258
                  note = Any amino acid
source            1..258
                  mol_type = protein
                  organism = synthetic construct
REGION            1..258
                  note = Description of Artificial Sequence: Synthetic
                  polypeptide

SEQUENCE: 229
WGGNPISQQV VRNPCKDSGX XXSGXGRQPR SVQVVDPKYM GPEYTFHSDW WRRGLFGEKA 60
IKRMSEQPTD DEIFTGGXPK RPRRDPPTXQ XPEEXXXXQK ESSSFRXXXX XXXXXXXXXX 120
PWESSSQEXE SESQEEEEXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXEQ TVQQQLRQQL 180
REQRRLRVQL QLLFQQLLKT XXXXQAGLHI NPULLSQAXX XXXXXXXXXXXX XXXXXXXXXX 240
XXXXXXXXXX XXXXXXXXX 258

```

---

-continued

---

```

SEQ ID NO: 230      moltype = AA  length = 214
FEATURE          Location/Qualifiers
VARIANT          136
                  note = Any amino acid
VARIANT          138..141
                  note = Any amino acid
SITE             138..141
                  note = This region may encompass 1-4 residues
VARIANT          179
                  note = Any amino acid
source            1..214
                  mol_type = protein
                  organism = synthetic construct
REGION            1..214
                  note = Description of Artificial Sequence: Synthetic
                  polypeptide

SEQUENCE: 230
LKQWQPSTIR KCKIKGYLPL FQCGKGRISN NYTQYKESIV PHHEPGGGGW SIQQFTLGLA 60
YEEHLKLRLNW WTKSNNDGLPL VRYLGCTIKL YRSEDTDYIV TYQRCYPMTA TKLTYLSTQP 120
SRMLLMNKHKI IVPSKXTXXX XNKKKKPYKK IFIGKPPSQMQ NKWYFQQDIA NTPLLQLTXT 180
ACSLDRMMLYS SDSISNNITF TSLNTNFFQN PNFO 214

SEQ ID NO: 231      moltype = AA  length = 187
FEATURE          Location/Qualifiers
VARIANT          1..10
                  note = Any amino acid
SITE             1..10
                  note = This region may encompass 4-10 residues
VARIANT          38..45
                  note = Any amino acid
SITE             38..45
                  note = This region may encompass 1-8 residues
VARIANT          94
                  note = Any amino acid
VARIANT          100..102
                  note = Any amino acid
SITE             100..102
                  note = This region may encompass 1-3 residues
VARIANT          112
                  note = Any amino acid
VARIANT          114..115
                  note = Any amino acid
SITE             114..115
                  note = This region may encompass 0-2 residues
VARIANT          124..139
                  note = Any amino acid
SITE             124..139
                  note = This region may encompass 3-16 residues
VARIANT          154
                  note = Any amino acid
source            1..187
                  mol_type = protein
                  organism = synthetic construct
REGION            1..187
                  note = Description of Artificial Sequence: Synthetic
                  polypeptide

SEQUENCE: 231
XXXXXXXXXX TPLYFECRYN PFKDKGTGNK VYLVSNXXXX XXXXXTGWDPT TDTPDLIIEG 60
FPLWLLLWGWD LDWQKKGKI QNIDTDYILV IQSXYYIPPX XXKLPPYYVPL DXDXXFLHGR 120
SPYXXXXXXX XXXXXXXXXXP SDKQHWHPKV RFQXETINNNI ALTGPGBTPKL PNQKSIQAHM 180
KYKFYFK 187

SEQ ID NO: 232      moltype = AA  length = 163
FEATURE          Location/Qualifiers
VARIANT          34
                  note = Any amino acid
VARIANT          65
                  note = Any amino acid
VARIANT          77..78
                  note = Any amino acid
VARIANT          86..87
                  note = Any amino acid
VARIANT          96
                  note = Any amino acid
VARIANT          102..106

```

---

-continued

---

```

SITE note = Any amino acid
      102..106
VARIANT note = This region may encompass 0-5 residues
      125
VARIANT note = Any amino acid
      135
VARIANT note = Any amino acid
      138..163
VARIANT note = Any amino acid
      138..163
SITE note = This region may encompass 0-26 residues
      1..163
source mol_type = protein
organism = synthetic construct
      1..163
REGION note = Description of Artificial Sequence: Synthetic
polypeptide

SEQUENCE: 232
WGGCPAPMET ITDPCKQPKY PIPNNNLLQTT SLQXPPTTPIE TYLYKFDERR GLLTKKAAKR 60
IKKDXTTETT LFTDTGXXTS TTLPTXXQTE TTQEEXTSEE EXXXXXETLL QQLQQLRRKQ 120
KQLRXRILQL LQLLXLLXXX XXXXXXXXXXXX XXXXXXXXXXXX XXX 163

SEQ ID NO: 233      moltype = AA length = 203
FEATURE Location/Qualifiers
VARIANT 79
note = Any amino acid
VARIANT 104
note = Any amino acid
VARIANT 116
note = Any amino acid
VARIANT 120..121
note = Any amino acid
VARIANT 125
note = Any amino acid
VARIANT 170
note = Any amino acid
source 1..203
mol_type = protein
organism = synthetic construct
      1..203
REGION note = Description of Artificial Sequence: Synthetic
polypeptide

SEQUENCE: 233
TIPLKQWQPE SIRKCKIKGY GTLVLGAEGR QFYCYTNEKD EYTPPKAPGG GGFGVELFSL 60
EVLYEIQWKAR NNIWTKSNXY KDLCRYTGCK ITFYRHPTTD FIVXYSRQPP FEIDKXTYMX 120
XHPQXLLLRLK HKKIILSKAT NPKGKLKKKI KIKPPKQMLN KWFFQKQFAX YGLVQLQAAA 180
CBLRYPRLGC CNENRLITLY YLN 203

SEQ ID NO: 234      moltype = AA length = 162
FEATURE Location/Qualifiers
VARIANT 12
note = Any amino acid
VARIANT 20
note = Any amino acid
VARIANT 23
note = Any amino acid
VARIANT 30
note = Any amino acid
VARIANT 58
note = I or L
VARIANT 84
note = Any amino acid
VARIANT 90
note = Any amino acid
VARIANT 95
note = Any amino acid
VARIANT 105
note = Any amino acid
VARIANT 111
note = I or L
VARIANT 113
note = Any amino acid
VARIANT 154
note = Any amino acid
VARIANT 156
note = Any amino acid

```

---

-continued

---

```

source          1..162
               mol_type = protein
               organism = synthetic construct
REGION          1..162
               note = Description of Artificial Sequence: Synthetic
               polypeptide
SEQUENCE: 234
LPPIVVARYNP AXDTGKGNKX WLXSTLNGSX WAPPPTDKDL IIEGLPLWLA LYGYWSYJKK  60
VKDKDKGILQS HMFVVVKSPA QPLXTATTQX TFYPXIDNSF IQGKXPYDEP JTXNQKKLWY 120
PTLEHQQETI NAIVESGPYV PKLDNQKNST WELXYXYTFY FK                           162

SEQ ID NO: 235      moltype = AA length = 177
FEATURE          Location/Qualifiers
VARIANT          16
note = Any amino acid
VARIANT          26
note = Any amino acid
VARIANT          33
note = Any amino acid
VARIANT          73
note = Any amino acid
VARIANT          81..82
note = Any amino acid
SITE             81..82
note = This region may encompass 0-2 residues
VARIANT          90
note = Any amino acid
VARIANT          94
note = Any amino acid
VARIANT          119..124
note = Any amino acid
SITE             119..124
note = This region may encompass 0-2 residues
VARIANT          168..177
note = Any amino acid
SITE             168..177
note = This region may encompass 1-6 residues
VARIANT          1..177
source          mol_type = protein
               organism = synthetic construct
REGION          1..177
               note = Description of Artificial Sequence: Synthetic
               polypeptide
SEQUENCE: 235
WGPPQIPDQP VEDPKXQGTY PVPDTXQQTI QIXNPLKQKP ETMFHDWDYR RGIITSTALK  60
RMQENLETDS SFXSDSEETP XXXKKKRRLTX ELPXPQEETE EIQSCLLSLC EESTCQEEXX 120
XXXXENLQQL IHQQQQQQQQ LKHNLKLSS DLKZKQRLLQ LQTGILEXXX XXXXXXXX      177

SEQ ID NO: 236      moltype = length =
SEQUENCE: 236
000

SEQ ID NO: 237      moltype = length =
SEQUENCE: 237
000

SEQ ID NO: 238      moltype = length =
SEQUENCE: 238
000

SEQ ID NO: 239      moltype = length =
SEQUENCE: 239
000

SEQ ID NO: 240      moltype = length =
SEQUENCE: 240
000

SEQ ID NO: 241      moltype = length =
SEQUENCE: 241
000

SEQ ID NO: 242      moltype = length =
SEQUENCE: 242
000

```

-continued

---

SEQ ID NO: 243	moltype =	length =
SEQUENCE: 243		
000		
SEQ ID NO: 244	moltype =	length =
SEQUENCE: 244		
000		
SEQ ID NO: 245	moltype =	length =
SEQUENCE: 245		
000		
SEQ ID NO: 246	moltype =	length =
SEQUENCE: 246		
000		
SEQ ID NO: 247	moltype =	length =
SEQUENCE: 247		
000		
SEQ ID NO: 248	moltype =	length =
SEQUENCE: 248		
000		
SEQ ID NO: 249	moltype =	length =
SEQUENCE: 249		
000		
SEQ ID NO: 250	moltype =	length =
SEQUENCE: 250		
000		
SEQ ID NO: 251	moltype =	length =
SEQUENCE: 251		
000		
SEQ ID NO: 252	moltype =	length =
SEQUENCE: 252		
000		
SEQ ID NO: 253	moltype =	length =
SEQUENCE: 253		
000		
SEQ ID NO: 254	moltype =	length =
SEQUENCE: 254		
000		
SEQ ID NO: 255	moltype =	length =
SEQUENCE: 255		
000		
SEQ ID NO: 256	moltype =	length =
SEQUENCE: 256		
000		
SEQ ID NO: 257	moltype =	length =
SEQUENCE: 257		
000		
SEQ ID NO: 258	moltype =	length =
SEQUENCE: 258		
000		
SEQ ID NO: 259	moltype =	length =
SEQUENCE: 259		
000		
SEQ ID NO: 260	moltype =	length =
SEQUENCE: 260		
000		
SEQ ID NO: 261	moltype =	length =
SEQUENCE: 261		
000		

-continued

---

SEQ ID NO: 262	moltype =	length =
SEQUENCE: 262		
000		
SEQ ID NO: 263	moltype =	length =
SEQUENCE: 263		
000		
SEQ ID NO: 264	moltype =	length =
SEQUENCE: 264		
000		
SEQ ID NO: 265	moltype =	length =
SEQUENCE: 265		
000		
SEQ ID NO: 266	moltype =	length =
SEQUENCE: 266		
000		
SEQ ID NO: 267	moltype =	length =
SEQUENCE: 267		
000		
SEQ ID NO: 268	moltype =	length =
SEQUENCE: 268		
000		
SEQ ID NO: 269	moltype =	length =
SEQUENCE: 269		
000		
SEQ ID NO: 270	moltype =	length =
SEQUENCE: 270		
000		
SEQ ID NO: 271	moltype =	length =
SEQUENCE: 271		
000		
SEQ ID NO: 272	moltype =	length =
SEQUENCE: 272		
000		
SEQ ID NO: 273	moltype =	length =
SEQUENCE: 273		
000		
SEQ ID NO: 274	moltype =	length =
SEQUENCE: 274		
000		
SEQ ID NO: 275	moltype =	length =
SEQUENCE: 275		
000		
SEQ ID NO: 276	moltype =	length =
SEQUENCE: 276		
000		
SEQ ID NO: 277	moltype =	length =
SEQUENCE: 277		
000		
SEQ ID NO: 278	moltype =	length =
SEQUENCE: 278		
000		
SEQ ID NO: 279	moltype =	length =
SEQUENCE: 279		
000		
SEQ ID NO: 280	moltype =	length =
SEQUENCE: 280		
000		

-continued

---

SEQ ID NO: 281	moltype =	length =
SEQUENCE: 281		
000		
SEQ ID NO: 282	moltype =	length =
SEQUENCE: 282		
000		
SEQ ID NO: 283	moltype =	length =
SEQUENCE: 283		
000		
SEQ ID NO: 284	moltype =	length =
SEQUENCE: 284		
000		
SEQ ID NO: 285	moltype =	length =
SEQUENCE: 285		
000		
SEQ ID NO: 286	moltype =	length =
SEQUENCE: 286		
000		
SEQ ID NO: 287	moltype =	length =
SEQUENCE: 287		
000		
SEQ ID NO: 288	moltype =	length =
SEQUENCE: 288		
000		
SEQ ID NO: 289	moltype =	length =
SEQUENCE: 289		
000		
SEQ ID NO: 290	moltype =	length =
SEQUENCE: 290		
000		
SEQ ID NO: 291	moltype =	length =
SEQUENCE: 291		
000		
SEQ ID NO: 292	moltype =	length =
SEQUENCE: 292		
000		
SEQ ID NO: 293	moltype =	length =
SEQUENCE: 293		
000		
SEQ ID NO: 294	moltype =	length =
SEQUENCE: 294		
000		
SEQ ID NO: 295	moltype =	length =
SEQUENCE: 295		
000		
SEQ ID NO: 296	moltype =	length =
SEQUENCE: 296		
000		
SEQ ID NO: 297	moltype =	length =
SEQUENCE: 297		
000		
SEQ ID NO: 298	moltype =	length =
SEQUENCE: 298		
000		
SEQ ID NO: 299	moltype =	length =
SEQUENCE: 299		
000		

-continued

---

SEQ ID NO: 300	moltype =	length =
SEQUENCE: 300		
000		
SEQ ID NO: 301	moltype =	length =
SEQUENCE: 301		
000		
SEQ ID NO: 302	moltype =	length =
SEQUENCE: 302		
000		
SEQ ID NO: 303	moltype =	length =
SEQUENCE: 303		
000		
SEQ ID NO: 304	moltype =	length =
SEQUENCE: 304		
000		
SEQ ID NO: 305	moltype =	length =
SEQUENCE: 305		
000		
SEQ ID NO: 306	moltype =	length =
SEQUENCE: 306		
000		
SEQ ID NO: 307	moltype =	length =
SEQUENCE: 307		
000		
SEQ ID NO: 308	moltype =	length =
SEQUENCE: 308		
000		
SEQ ID NO: 309	moltype =	length =
SEQUENCE: 309		
000		
SEQ ID NO: 310	moltype =	length =
SEQUENCE: 310		
000		
SEQ ID NO: 311	moltype =	length =
SEQUENCE: 311		
000		
SEQ ID NO: 312	moltype =	length =
SEQUENCE: 312		
000		
SEQ ID NO: 313	moltype =	length =
SEQUENCE: 313		
000		
SEQ ID NO: 314	moltype =	length =
SEQUENCE: 314		
000		
SEQ ID NO: 315	moltype =	length =
SEQUENCE: 315		
000		
SEQ ID NO: 316	moltype =	length =
SEQUENCE: 316		
000		
SEQ ID NO: 317	moltype =	length =
SEQUENCE: 317		
000		
SEQ ID NO: 318	moltype =	length =
SEQUENCE: 318		
000		

-continued

---

SEQ ID NO: 319	moltype =	length =
SEQUENCE: 319		
000		
SEQ ID NO: 320	moltype =	length =
SEQUENCE: 320		
000		
SEQ ID NO: 321	moltype =	length =
SEQUENCE: 321		
000		
SEQ ID NO: 322	moltype =	length =
SEQUENCE: 322		
000		
SEQ ID NO: 323	moltype =	length =
SEQUENCE: 323		
000		
SEQ ID NO: 324	moltype =	length =
SEQUENCE: 324		
000		
SEQ ID NO: 325	moltype =	length =
SEQUENCE: 325		
000		
SEQ ID NO: 326	moltype =	length =
SEQUENCE: 326		
000		
SEQ ID NO: 327	moltype =	length =
SEQUENCE: 327		
000		
SEQ ID NO: 328	moltype =	length =
SEQUENCE: 328		
000		
SEQ ID NO: 329	moltype =	length =
SEQUENCE: 329		
000		
SEQ ID NO: 330	moltype =	length =
SEQUENCE: 330		
000		
SEQ ID NO: 331	moltype =	length =
SEQUENCE: 331		
000		
SEQ ID NO: 332	moltype =	length =
SEQUENCE: 332		
000		
SEQ ID NO: 333	moltype =	length =
SEQUENCE: 333		
000		
SEQ ID NO: 334	moltype =	length =
SEQUENCE: 334		
000		
SEQ ID NO: 335	moltype =	length =
SEQUENCE: 335		
000		
SEQ ID NO: 336	moltype =	length =
SEQUENCE: 336		
000		
SEQ ID NO: 337	moltype =	length =
SEQUENCE: 337		
000		

-continued

---

SEQ ID NO: 338	moltype =	length =
SEQUENCE: 338		
000		
SEQ ID NO: 339	moltype =	length =
SEQUENCE: 339		
000		
SEQ ID NO: 340	moltype =	length =
SEQUENCE: 340		
000		
SEQ ID NO: 341	moltype =	length =
SEQUENCE: 341		
000		
SEQ ID NO: 342	moltype =	length =
SEQUENCE: 342		
000		
SEQ ID NO: 343	moltype =	length =
SEQUENCE: 343		
000		
SEQ ID NO: 344	moltype =	length =
SEQUENCE: 344		
000		
SEQ ID NO: 345	moltype =	length =
SEQUENCE: 345		
000		
SEQ ID NO: 346	moltype =	length =
SEQUENCE: 346		
000		
SEQ ID NO: 347	moltype =	length =
SEQUENCE: 347		
000		
SEQ ID NO: 348	moltype =	length =
SEQUENCE: 348		
000		
SEQ ID NO: 349	moltype =	length =
SEQUENCE: 349		
000		
SEQ ID NO: 350	moltype =	length =
SEQUENCE: 350		
000		
SEQ ID NO: 351	moltype =	length =
SEQUENCE: 351		
000		
SEQ ID NO: 352	moltype =	length =
SEQUENCE: 352		
000		
SEQ ID NO: 353	moltype =	length =
SEQUENCE: 353		
000		
SEQ ID NO: 354	moltype =	length =
SEQUENCE: 354		
000		
SEQ ID NO: 355	moltype =	length =
SEQUENCE: 355		
000		
SEQ ID NO: 356	moltype =	length =
SEQUENCE: 356		
000		

-continued

---

SEQ ID NO: 357	moltype =	length =
SEQUENCE: 357		
000		
SEQ ID NO: 358	moltype =	length =
SEQUENCE: 358		
000		
SEQ ID NO: 359	moltype =	length =
SEQUENCE: 359		
000		
SEQ ID NO: 360	moltype =	length =
SEQUENCE: 360		
000		
SEQ ID NO: 361	moltype =	length =
SEQUENCE: 361		
000		
SEQ ID NO: 362	moltype =	length =
SEQUENCE: 362		
000		
SEQ ID NO: 363	moltype =	length =
SEQUENCE: 363		
000		
SEQ ID NO: 364	moltype =	length =
SEQUENCE: 364		
000		
SEQ ID NO: 365	moltype =	length =
SEQUENCE: 365		
000		
SEQ ID NO: 366	moltype =	length =
SEQUENCE: 366		
000		
SEQ ID NO: 367	moltype =	length =
SEQUENCE: 367		
000		
SEQ ID NO: 368	moltype =	length =
SEQUENCE: 368		
000		
SEQ ID NO: 369	moltype =	length =
SEQUENCE: 369		
000		
SEQ ID NO: 370	moltype =	length =
SEQUENCE: 370		
000		
SEQ ID NO: 371	moltype =	length =
SEQUENCE: 371		
000		
SEQ ID NO: 372	moltype =	length =
SEQUENCE: 372		
000		
SEQ ID NO: 373	moltype =	length =
SEQUENCE: 373		
000		
SEQ ID NO: 374	moltype =	length =
SEQUENCE: 374		
000		
SEQ ID NO: 375	moltype =	length =
SEQUENCE: 375		
000		

-continued

---

SEQ ID NO: 376	moltype =	length =
SEQUENCE: 376		
000		
SEQ ID NO: 377	moltype =	length =
SEQUENCE: 377		
000		
SEQ ID NO: 378	moltype =	length =
SEQUENCE: 378		
000		
SEQ ID NO: 379	moltype =	length =
SEQUENCE: 379		
000		
SEQ ID NO: 380	moltype =	length =
SEQUENCE: 380		
000		
SEQ ID NO: 381	moltype =	length =
SEQUENCE: 381		
000		
SEQ ID NO: 382	moltype =	length =
SEQUENCE: 382		
000		
SEQ ID NO: 383	moltype =	length =
SEQUENCE: 383		
000		
SEQ ID NO: 384	moltype =	length =
SEQUENCE: 384		
000		
SEQ ID NO: 385	moltype =	length =
SEQUENCE: 385		
000		
SEQ ID NO: 386	moltype =	length =
SEQUENCE: 386		
000		
SEQ ID NO: 387	moltype =	length =
SEQUENCE: 387		
000		
SEQ ID NO: 388	moltype =	length =
SEQUENCE: 388		
000		
SEQ ID NO: 389	moltype =	length =
SEQUENCE: 389		
000		
SEQ ID NO: 390	moltype =	length =
SEQUENCE: 390		
000		
SEQ ID NO: 391	moltype =	length =
SEQUENCE: 391		
000		
SEQ ID NO: 392	moltype =	length =
SEQUENCE: 392		
000		
SEQ ID NO: 393	moltype =	length =
SEQUENCE: 393		
000		
SEQ ID NO: 394	moltype =	length =
SEQUENCE: 394		
000		

-continued

---

SEQ ID NO: 395	moltype =	length =
SEQUENCE: 395		
000		
SEQ ID NO: 396	moltype =	length =
SEQUENCE: 396		
000		
SEQ ID NO: 397	moltype =	length =
SEQUENCE: 397		
000		
SEQ ID NO: 398	moltype =	length =
SEQUENCE: 398		
000		
SEQ ID NO: 399	moltype =	length =
SEQUENCE: 399		
000		
SEQ ID NO: 400	moltype =	length =
SEQUENCE: 400		
000		
SEQ ID NO: 401	moltype =	length =
SEQUENCE: 401		
000		
SEQ ID NO: 402	moltype =	length =
SEQUENCE: 402		
000		
SEQ ID NO: 403	moltype =	length =
SEQUENCE: 403		
000		
SEQ ID NO: 404	moltype =	length =
SEQUENCE: 404		
000		
SEQ ID NO: 405	moltype =	length =
SEQUENCE: 405		
000		
SEQ ID NO: 406	moltype =	length =
SEQUENCE: 406		
000		
SEQ ID NO: 407	moltype =	length =
SEQUENCE: 407		
000		
SEQ ID NO: 408	moltype =	length =
SEQUENCE: 408		
000		
SEQ ID NO: 409	moltype =	length =
SEQUENCE: 409		
000		
SEQ ID NO: 410	moltype =	length =
SEQUENCE: 410		
000		
SEQ ID NO: 411	moltype =	length =
SEQUENCE: 411		
000		
SEQ ID NO: 412	moltype =	length =
SEQUENCE: 412		
000		
SEQ ID NO: 413	moltype =	length =
SEQUENCE: 413		
000		

-continued

---

SEQ ID NO: 414	moltype =	length =
SEQUENCE: 414		
000		
SEQ ID NO: 415	moltype =	length =
SEQUENCE: 415		
000		
SEQ ID NO: 416	moltype =	length =
SEQUENCE: 416		
000		
SEQ ID NO: 417	moltype =	length =
SEQUENCE: 417		
000		
SEQ ID NO: 418	moltype =	length =
SEQUENCE: 418		
000		
SEQ ID NO: 419	moltype =	length =
SEQUENCE: 419		
000		
SEQ ID NO: 420	moltype =	length =
SEQUENCE: 420		
000		
SEQ ID NO: 421	moltype =	length =
SEQUENCE: 421		
000		
SEQ ID NO: 422	moltype =	length =
SEQUENCE: 422		
000		
SEQ ID NO: 423	moltype =	length =
SEQUENCE: 423		
000		
SEQ ID NO: 424	moltype =	length =
SEQUENCE: 424		
000		
SEQ ID NO: 425	moltype =	length =
SEQUENCE: 425		
000		
SEQ ID NO: 426	moltype =	length =
SEQUENCE: 426		
000		
SEQ ID NO: 427	moltype =	length =
SEQUENCE: 427		
000		
SEQ ID NO: 428	moltype =	length =
SEQUENCE: 428		
000		
SEQ ID NO: 429	moltype =	length =
SEQUENCE: 429		
000		
SEQ ID NO: 430	moltype =	length =
SEQUENCE: 430		
000		
SEQ ID NO: 431	moltype =	length =
SEQUENCE: 431		
000		
SEQ ID NO: 432	moltype =	length =
SEQUENCE: 432		
000		

-continued

---

SEQ ID NO: 433	moltype =	length =
SEQUENCE: 433		
000		
SEQ ID NO: 434	moltype =	length =
SEQUENCE: 434		
000		
SEQ ID NO: 435	moltype =	length =
SEQUENCE: 435		
000		
SEQ ID NO: 436	moltype =	length =
SEQUENCE: 436		
000		
SEQ ID NO: 437	moltype =	length =
SEQUENCE: 437		
000		
SEQ ID NO: 438	moltype =	length =
SEQUENCE: 438		
000		
SEQ ID NO: 439	moltype =	length =
SEQUENCE: 439		
000		
SEQ ID NO: 440	moltype =	length =
SEQUENCE: 440		
000		
SEQ ID NO: 441	moltype =	length =
SEQUENCE: 441		
000		
SEQ ID NO: 442	moltype =	length =
SEQUENCE: 442		
000		
SEQ ID NO: 443	moltype =	length =
SEQUENCE: 443		
000		
SEQ ID NO: 444	moltype =	length =
SEQUENCE: 444		
000		
SEQ ID NO: 445	moltype =	length =
SEQUENCE: 445		
000		
SEQ ID NO: 446	moltype =	length =
SEQUENCE: 446		
000		
SEQ ID NO: 447	moltype =	length =
SEQUENCE: 447		
000		
SEQ ID NO: 448	moltype =	length =
SEQUENCE: 448		
000		
SEQ ID NO: 449	moltype =	length =
SEQUENCE: 449		
000		
SEQ ID NO: 450	moltype =	length =
SEQUENCE: 450		
000		
SEQ ID NO: 451	moltype =	length =
SEQUENCE: 451		
000		

-continued

---

SEQ ID NO: 452	moltype =	length =
SEQUENCE: 452		
000		
SEQ ID NO: 453	moltype =	length =
SEQUENCE: 453		
000		
SEQ ID NO: 454	moltype =	length =
SEQUENCE: 454		
000		
SEQ ID NO: 455	moltype =	length =
SEQUENCE: 455		
000		
SEQ ID NO: 456	moltype =	length =
SEQUENCE: 456		
000		
SEQ ID NO: 457	moltype =	length =
SEQUENCE: 457		
000		
SEQ ID NO: 458	moltype =	length =
SEQUENCE: 458		
000		
SEQ ID NO: 459	moltype =	length =
SEQUENCE: 459		
000		
SEQ ID NO: 460	moltype =	length =
SEQUENCE: 460		
000		
SEQ ID NO: 461	moltype =	length =
SEQUENCE: 461		
000		
SEQ ID NO: 462	moltype =	length =
SEQUENCE: 462		
000		
SEQ ID NO: 463	moltype =	length =
SEQUENCE: 463		
000		
SEQ ID NO: 464	moltype =	length =
SEQUENCE: 464		
000		
SEQ ID NO: 465	moltype =	length =
SEQUENCE: 465		
000		
SEQ ID NO: 466	moltype =	length =
SEQUENCE: 466		
000		
SEQ ID NO: 467	moltype =	length =
SEQUENCE: 467		
000		
SEQ ID NO: 468	moltype =	length =
SEQUENCE: 468		
000		
SEQ ID NO: 469	moltype =	length =
SEQUENCE: 469		
000		
SEQ ID NO: 470	moltype =	length =
SEQUENCE: 470		
000		

-continued

---

SEQ ID NO: 471	moltype =	length =
SEQUENCE: 471		
000		
SEQ ID NO: 472	moltype =	length =
SEQUENCE: 472		
000		
SEQ ID NO: 473	moltype =	length =
SEQUENCE: 473		
000		
SEQ ID NO: 474	moltype =	length =
SEQUENCE: 474		
000		
SEQ ID NO: 475	moltype =	length =
SEQUENCE: 475		
000		
SEQ ID NO: 476	moltype =	length =
SEQUENCE: 476		
000		
SEQ ID NO: 477	moltype =	length =
SEQUENCE: 477		
000		
SEQ ID NO: 478	moltype =	length =
SEQUENCE: 478		
000		
SEQ ID NO: 479	moltype =	length =
SEQUENCE: 479		
000		
SEQ ID NO: 480	moltype =	length =
SEQUENCE: 480		
000		
SEQ ID NO: 481	moltype =	length =
SEQUENCE: 481		
000		
SEQ ID NO: 482	moltype =	length =
SEQUENCE: 482		
000		
SEQ ID NO: 483	moltype =	length =
SEQUENCE: 483		
000		
SEQ ID NO: 484	moltype =	length =
SEQUENCE: 484		
000		
SEQ ID NO: 485	moltype =	length =
SEQUENCE: 485		
000		
SEQ ID NO: 486	moltype =	length =
SEQUENCE: 486		
000		
SEQ ID NO: 487	moltype =	length =
SEQUENCE: 487		
000		
SEQ ID NO: 488	moltype =	length =
SEQUENCE: 488		
000		
SEQ ID NO: 489	moltype =	length =
SEQUENCE: 489		
000		

-continued

---

SEQ ID NO: 490	moltype =	length =
SEQUENCE: 490		
000		
SEQ ID NO: 491	moltype =	length =
SEQUENCE: 491		
000		
SEQ ID NO: 492	moltype =	length =
SEQUENCE: 492		
000		
SEQ ID NO: 493	moltype =	length =
SEQUENCE: 493		
000		
SEQ ID NO: 494	moltype =	length =
SEQUENCE: 494		
000		
SEQ ID NO: 495	moltype =	length =
SEQUENCE: 495		
000		
SEQ ID NO: 496	moltype =	length =
SEQUENCE: 496		
000		
SEQ ID NO: 497	moltype =	length =
SEQUENCE: 497		
000		
SEQ ID NO: 498	moltype =	length =
SEQUENCE: 498		
000		
SEQ ID NO: 499	moltype =	length =
SEQUENCE: 499		
000		
SEQ ID NO: 500	moltype =	length =
SEQUENCE: 500		
000		
SEQ ID NO: 501	moltype =	length =
SEQUENCE: 501		
000		
SEQ ID NO: 502	moltype =	length =
SEQUENCE: 502		
000		
SEQ ID NO: 503	moltype =	length =
SEQUENCE: 503		
000		
SEQ ID NO: 504	moltype =	length =
SEQUENCE: 504		
000		
SEQ ID NO: 505	moltype =	length =
SEQUENCE: 505		
000		
SEQ ID NO: 506	moltype =	length =
SEQUENCE: 506		
000		
SEQ ID NO: 507	moltype =	length =
SEQUENCE: 507		
000		
SEQ ID NO: 508	moltype =	length =
SEQUENCE: 508		
000		

-continued

---

SEQ ID NO: 509	moltype =	length =
SEQUENCE: 509		
000		
SEQ ID NO: 510	moltype =	length =
SEQUENCE: 510		
000		
SEQ ID NO: 511	moltype =	length =
SEQUENCE: 511		
000		
SEQ ID NO: 512	moltype =	length =
SEQUENCE: 512		
000		
SEQ ID NO: 513	moltype =	length =
SEQUENCE: 513		
000		
SEQ ID NO: 514	moltype =	length =
SEQUENCE: 514		
000		
SEQ ID NO: 515	moltype =	length =
SEQUENCE: 515		
000		
SEQ ID NO: 516	moltype =	length =
SEQUENCE: 516		
000		
SEQ ID NO: 517	moltype =	length =
SEQUENCE: 517		
000		
SEQ ID NO: 518	moltype =	length =
SEQUENCE: 518		
000		
SEQ ID NO: 519	moltype =	length =
SEQUENCE: 519		
000		
SEQ ID NO: 520	moltype =	length =
SEQUENCE: 520		
000		
SEQ ID NO: 521	moltype =	length =
SEQUENCE: 521		
000		
SEQ ID NO: 522	moltype =	length =
SEQUENCE: 522		
000		
SEQ ID NO: 523	moltype =	length =
SEQUENCE: 523		
000		
SEQ ID NO: 524	moltype =	length =
SEQUENCE: 524		
000		
SEQ ID NO: 525	moltype =	length =
SEQUENCE: 525		
000		
SEQ ID NO: 526	moltype =	length =
SEQUENCE: 526		
000		
SEQ ID NO: 527	moltype =	length =
SEQUENCE: 527		
000		

-continued

---

SEQ ID NO: 528	moltype =	length =
SEQUENCE: 528		
000		
SEQ ID NO: 529	moltype =	length =
SEQUENCE: 529		
000		
SEQ ID NO: 530	moltype =	length =
SEQUENCE: 530		
000		
SEQ ID NO: 531	moltype =	length =
SEQUENCE: 531		
000		
SEQ ID NO: 532	moltype =	length =
SEQUENCE: 532		
000		
SEQ ID NO: 533	moltype =	length =
SEQUENCE: 533		
000		
SEQ ID NO: 534	moltype =	length =
SEQUENCE: 534		
000		
SEQ ID NO: 535	moltype =	length =
SEQUENCE: 535		
000		
SEQ ID NO: 536	moltype =	length =
SEQUENCE: 536		
000		
SEQ ID NO: 537	moltype =	length =
SEQUENCE: 537		
000		
SEQ ID NO: 538	moltype =	length =
SEQUENCE: 538		
000		
SEQ ID NO: 539	moltype =	length =
SEQUENCE: 539		
000		
SEQ ID NO: 540	moltype =	length =
SEQUENCE: 540		
000		
SEQ ID NO: 541	moltype =	length =
SEQUENCE: 541		
000		
SEQ ID NO: 542	moltype =	length =
SEQUENCE: 542		
000		
SEQ ID NO: 543	moltype =	length =
SEQUENCE: 543		
000		
SEQ ID NO: 544	moltype =	length =
SEQUENCE: 544		
000		
SEQ ID NO: 545	moltype =	length =
SEQUENCE: 545		
000		
SEQ ID NO: 546	moltype =	length =
SEQUENCE: 546		
000		

-continued

---

SEQ ID NO: 547	moltype =	length =
SEQUENCE: 547		
000		
SEQ ID NO: 548	moltype =	length =
SEQUENCE: 548		
000		
SEQ ID NO: 549	moltype =	length =
SEQUENCE: 549		
000		
SEQ ID NO: 550	moltype =	length =
SEQUENCE: 550		
000		
SEQ ID NO: 551	moltype =	length =
SEQUENCE: 551		
000		
SEQ ID NO: 552	moltype =	length =
SEQUENCE: 552		
000		
SEQ ID NO: 553	moltype =	length =
SEQUENCE: 553		
000		
SEQ ID NO: 554	moltype =	length =
SEQUENCE: 554		
000		
SEQ ID NO: 555	moltype =	length =
SEQUENCE: 555		
000		
SEQ ID NO: 556	moltype =	length =
SEQUENCE: 556		
000		
SEQ ID NO: 557	moltype =	length =
SEQUENCE: 557		
000		
SEQ ID NO: 558	moltype =	length =
SEQUENCE: 558		
000		
SEQ ID NO: 559	moltype =	length =
SEQUENCE: 559		
000		
SEQ ID NO: 560	moltype =	length =
SEQUENCE: 560		
000		
SEQ ID NO: 561	moltype =	length =
SEQUENCE: 561		
000		
SEQ ID NO: 562	moltype =	length =
SEQUENCE: 562		
000		
SEQ ID NO: 563	moltype =	length =
SEQUENCE: 563		
000		
SEQ ID NO: 564	moltype =	length =
SEQUENCE: 564		
000		
SEQ ID NO: 565	moltype =	length =
SEQUENCE: 565		
000		

-continued

---

SEQ ID NO: 566	moltype =	length =
SEQUENCE: 566		
000		
SEQ ID NO: 567	moltype =	length =
SEQUENCE: 567		
000		
SEQ ID NO: 568	moltype =	length =
SEQUENCE: 568		
000		
SEQ ID NO: 569	moltype =	length =
SEQUENCE: 569		
000		
SEQ ID NO: 570	moltype =	length =
SEQUENCE: 570		
000		
SEQ ID NO: 571	moltype =	length =
SEQUENCE: 571		
000		
SEQ ID NO: 572	moltype =	length =
SEQUENCE: 572		
000		
SEQ ID NO: 573	moltype =	length =
SEQUENCE: 573		
000		
SEQ ID NO: 574	moltype =	length =
SEQUENCE: 574		
000		
SEQ ID NO: 575	moltype =	length =
SEQUENCE: 575		
000		
SEQ ID NO: 576	moltype =	length =
SEQUENCE: 576		
000		
SEQ ID NO: 577	moltype =	length =
SEQUENCE: 577		
000		
SEQ ID NO: 578	moltype =	length =
SEQUENCE: 578		
000		
SEQ ID NO: 579	moltype =	length =
SEQUENCE: 579		
000		
SEQ ID NO: 580	moltype =	length =
SEQUENCE: 580		
000		
SEQ ID NO: 581	moltype =	length =
SEQUENCE: 581		
000		
SEQ ID NO: 582	moltype =	length =
SEQUENCE: 582		
000		
SEQ ID NO: 583	moltype =	length =
SEQUENCE: 583		
000		
SEQ ID NO: 584	moltype =	length =
SEQUENCE: 584		
000		

-continued

---

SEQ ID NO: 585	moltype =	length =
SEQUENCE: 585		
000		
SEQ ID NO: 586	moltype =	length =
SEQUENCE: 586		
000		
SEQ ID NO: 587	moltype =	length =
SEQUENCE: 587		
000		
SEQ ID NO: 588	moltype =	length =
SEQUENCE: 588		
000		
SEQ ID NO: 589	moltype =	length =
SEQUENCE: 589		
000		
SEQ ID NO: 590	moltype =	length =
SEQUENCE: 590		
000		
SEQ ID NO: 591	moltype =	length =
SEQUENCE: 591		
000		
SEQ ID NO: 592	moltype =	length =
SEQUENCE: 592		
000		
SEQ ID NO: 593	moltype =	length =
SEQUENCE: 593		
000		
SEQ ID NO: 594	moltype =	length =
SEQUENCE: 594		
000		
SEQ ID NO: 595	moltype =	length =
SEQUENCE: 595		
000		
SEQ ID NO: 596	moltype =	length =
SEQUENCE: 596		
000		
SEQ ID NO: 597	moltype =	length =
SEQUENCE: 597		
000		
SEQ ID NO: 598	moltype =	length =
SEQUENCE: 598		
000		
SEQ ID NO: 599	moltype =	length =
SEQUENCE: 599		
000		
SEQ ID NO: 600	moltype =	length =
SEQUENCE: 600		
000		
SEQ ID NO: 601	moltype =	length =
SEQUENCE: 601		
000		
SEQ ID NO: 602	moltype =	length =
SEQUENCE: 602		
000		
SEQ ID NO: 603	moltype =	length =
SEQUENCE: 603		
000		

-continued

---

SEQ ID NO: 604	moltype =	length =
SEQUENCE: 604		
000		
SEQ ID NO: 605	moltype =	length =
SEQUENCE: 605		
000		
SEQ ID NO: 606	moltype =	length =
SEQUENCE: 606		
000		
SEQ ID NO: 607	moltype =	length =
SEQUENCE: 607		
000		
SEQ ID NO: 608	moltype =	length =
SEQUENCE: 608		
000		
SEQ ID NO: 609	moltype =	length =
SEQUENCE: 609		
000		
SEQ ID NO: 610	moltype =	length =
SEQUENCE: 610		
000		
SEQ ID NO: 611	moltype =	length =
SEQUENCE: 611		
000		
SEQ ID NO: 612	moltype =	length =
SEQUENCE: 612		
000		
SEQ ID NO: 613	moltype =	length =
SEQUENCE: 613		
000		
SEQ ID NO: 614	moltype =	length =
SEQUENCE: 614		
000		
SEQ ID NO: 615	moltype =	length =
SEQUENCE: 615		
000		
SEQ ID NO: 616	moltype =	length =
SEQUENCE: 616		
000		
SEQ ID NO: 617	moltype =	length =
SEQUENCE: 617		
000		
SEQ ID NO: 618	moltype =	length =
SEQUENCE: 618		
000		
SEQ ID NO: 619	moltype =	length =
SEQUENCE: 619		
000		
SEQ ID NO: 620	moltype =	length =
SEQUENCE: 620		
000		
SEQ ID NO: 621	moltype =	length =
SEQUENCE: 621		
000		
SEQ ID NO: 622	moltype =	length =
SEQUENCE: 622		
000		

-continued

---

SEQ ID NO: 623	moltype =	length =
SEQUENCE: 623		
000		
SEQ ID NO: 624	moltype =	length =
SEQUENCE: 624		
000		
SEQ ID NO: 625	moltype =	length =
SEQUENCE: 625		
000		
SEQ ID NO: 626	moltype =	length =
SEQUENCE: 626		
000		
SEQ ID NO: 627	moltype =	length =
SEQUENCE: 627		
000		
SEQ ID NO: 628	moltype =	length =
SEQUENCE: 628		
000		
SEQ ID NO: 629	moltype =	length =
SEQUENCE: 629		
000		
SEQ ID NO: 630	moltype =	length =
SEQUENCE: 630		
000		
SEQ ID NO: 631	moltype =	length =
SEQUENCE: 631		
000		
SEQ ID NO: 632	moltype =	length =
SEQUENCE: 632		
000		
SEQ ID NO: 633	moltype =	length =
SEQUENCE: 633		
000		
SEQ ID NO: 634	moltype =	length =
SEQUENCE: 634		
000		
SEQ ID NO: 635	moltype =	length =
SEQUENCE: 635		
000		
SEQ ID NO: 636	moltype =	length =
SEQUENCE: 636		
000		
SEQ ID NO: 637	moltype =	length =
SEQUENCE: 637		
000		
SEQ ID NO: 638	moltype =	length =
SEQUENCE: 638		
000		
SEQ ID NO: 639	moltype =	length =
SEQUENCE: 639		
000		
SEQ ID NO: 640	moltype =	length =
SEQUENCE: 640		
000		
SEQ ID NO: 641	moltype =	length =
SEQUENCE: 641		
000		

-continued

---

SEQ ID NO: 642	moltype =	length =
SEQUENCE: 642		
000		
SEQ ID NO: 643	moltype =	length =
SEQUENCE: 643		
000		
SEQ ID NO: 644	moltype =	length =
SEQUENCE: 644		
000		
SEQ ID NO: 645	moltype =	length =
SEQUENCE: 645		
000		
SEQ ID NO: 646	moltype =	length =
SEQUENCE: 646		
000		
SEQ ID NO: 647	moltype =	length =
SEQUENCE: 647		
000		
SEQ ID NO: 648	moltype =	length =
SEQUENCE: 648		
000		
SEQ ID NO: 649	moltype =	length =
SEQUENCE: 649		
000		
SEQ ID NO: 650	moltype =	length =
SEQUENCE: 650		
000		
SEQ ID NO: 651	moltype =	length =
SEQUENCE: 651		
000		
SEQ ID NO: 652	moltype =	length =
SEQUENCE: 652		
000		
SEQ ID NO: 653	moltype =	length =
SEQUENCE: 653		
000		
SEQ ID NO: 654	moltype =	length =
SEQUENCE: 654		
000		
SEQ ID NO: 655	moltype =	length =
SEQUENCE: 655		
000		
SEQ ID NO: 656	moltype =	length =
SEQUENCE: 656		
000		
SEQ ID NO: 657	moltype =	length =
SEQUENCE: 657		
000		
SEQ ID NO: 658	moltype =	length =
SEQUENCE: 658		
000		
SEQ ID NO: 659	moltype =	length =
SEQUENCE: 659		
000		
SEQ ID NO: 660	moltype =	length =
SEQUENCE: 660		
000		

-continued

---

SEQ ID NO: 661	moltype =	length =
SEQUENCE: 661		
000		
SEQ ID NO: 662	moltype =	length =
SEQUENCE: 662		
000		
SEQ ID NO: 663	moltype =	length =
SEQUENCE: 663		
000		
SEQ ID NO: 664	moltype =	length =
SEQUENCE: 664		
000		
SEQ ID NO: 665	moltype =	length =
SEQUENCE: 665		
000		
SEQ ID NO: 666	moltype =	length =
SEQUENCE: 666		
000		
SEQ ID NO: 667	moltype =	length =
SEQUENCE: 667		
000		
SEQ ID NO: 668	moltype =	length =
SEQUENCE: 668		
000		
SEQ ID NO: 669	moltype =	length =
SEQUENCE: 669		
000		
SEQ ID NO: 670	moltype =	length =
SEQUENCE: 670		
000		
SEQ ID NO: 671	moltype =	length =
SEQUENCE: 671		
000		
SEQ ID NO: 672	moltype =	length =
SEQUENCE: 672		
000		
SEQ ID NO: 673	moltype =	length =
SEQUENCE: 673		
000		
SEQ ID NO: 674	moltype =	length =
SEQUENCE: 674		
000		
SEQ ID NO: 675	moltype =	length =
SEQUENCE: 675		
000		
SEQ ID NO: 676	moltype =	length =
SEQUENCE: 676		
000		
SEQ ID NO: 677	moltype =	length =
SEQUENCE: 677		
000		
SEQ ID NO: 678	moltype =	length =
SEQUENCE: 678		
000		
SEQ ID NO: 679	moltype =	length =
SEQUENCE: 679		
000		

-continued

---

SEQ ID NO: 680	moltype =	length =
SEQUENCE: 680		
000		
SEQ ID NO: 681	moltype =	length =
SEQUENCE: 681		
000		
SEQ ID NO: 682	moltype =	length =
SEQUENCE: 682		
000		
SEQ ID NO: 683	moltype =	length =
SEQUENCE: 683		
000		
SEQ ID NO: 684	moltype =	length =
SEQUENCE: 684		
000		
SEQ ID NO: 685	moltype =	length =
SEQUENCE: 685		
000		
SEQ ID NO: 686	moltype =	length =
SEQUENCE: 686		
000		
SEQ ID NO: 687	moltype =	length =
SEQUENCE: 687		
000		
SEQ ID NO: 688	moltype =	length =
SEQUENCE: 688		
000		
SEQ ID NO: 689	moltype =	length =
SEQUENCE: 689		
000		
SEQ ID NO: 690	moltype =	length =
SEQUENCE: 690		
000		
SEQ ID NO: 691	moltype =	length =
SEQUENCE: 691		
000		
SEQ ID NO: 692	moltype =	length =
SEQUENCE: 692		
000		
SEQ ID NO: 693	moltype =	length =
SEQUENCE: 693		
000		
SEQ ID NO: 694	moltype =	length =
SEQUENCE: 694		
000		
SEQ ID NO: 695	moltype =	length =
SEQUENCE: 695		
000		
SEQ ID NO: 696	moltype =	length =
SEQUENCE: 696		
000		
SEQ ID NO: 697	moltype =	length =
SEQUENCE: 697		
000		
SEQ ID NO: 698	moltype =	length =
SEQUENCE: 698		
000		

-continued

---

SEQ ID NO: 699	moltype =	length =
SEQUENCE: 699		
000		
SEQ ID NO: 700	moltype =	length =
SEQUENCE: 700		
000		
SEQ ID NO: 701	moltype =	length =
SEQUENCE: 701		
000		
SEQ ID NO: 702	moltype =	length =
SEQUENCE: 702		
000		
SEQ ID NO: 703	moltype =	length =
SEQUENCE: 703		
000		
SEQ ID NO: 704	moltype =	length =
SEQUENCE: 704		
000		
SEQ ID NO: 705	moltype =	length =
SEQUENCE: 705		
000		
SEQ ID NO: 706	moltype =	length =
SEQUENCE: 706		
000		
SEQ ID NO: 707	moltype =	length =
SEQUENCE: 707		
000		
SEQ ID NO: 708	moltype =	length =
SEQUENCE: 708		
000		
SEQ ID NO: 709	moltype =	length =
SEQUENCE: 709		
000		
SEQ ID NO: 710	moltype =	length =
SEQUENCE: 710		
000		
SEQ ID NO: 711	moltype =	length =
SEQUENCE: 711		
000		
SEQ ID NO: 712	moltype =	length =
SEQUENCE: 712		
000		
SEQ ID NO: 713	moltype =	length =
SEQUENCE: 713		
000		
SEQ ID NO: 714	moltype =	length =
SEQUENCE: 714		
000		
SEQ ID NO: 715	moltype =	length =
SEQUENCE: 715		
000		
SEQ ID NO: 716	moltype =	length =
SEQUENCE: 716		
000		
SEQ ID NO: 717	moltype =	length =
SEQUENCE: 717		
000		

-continued

---

SEQ ID NO: 718	moltype =	length =
SEQUENCE: 718		
000		
SEQ ID NO: 719	moltype =	length =
SEQUENCE: 719		
000		
SEQ ID NO: 720	moltype =	length =
SEQUENCE: 720		
000		
SEQ ID NO: 721	moltype =	length =
SEQUENCE: 721		
000		
SEQ ID NO: 722	moltype =	length =
SEQUENCE: 722		
000		
SEQ ID NO: 723	moltype =	length =
SEQUENCE: 723		
000		
SEQ ID NO: 724	moltype =	length =
SEQUENCE: 724		
000		
SEQ ID NO: 725	moltype =	length =
SEQUENCE: 725		
000		
SEQ ID NO: 726	moltype =	length =
SEQUENCE: 726		
000		
SEQ ID NO: 727	moltype =	length =
SEQUENCE: 727		
000		
SEQ ID NO: 728	moltype =	length =
SEQUENCE: 728		
000		
SEQ ID NO: 729	moltype =	length =
SEQUENCE: 729		
000		
SEQ ID NO: 730	moltype =	length =
SEQUENCE: 730		
000		
SEQ ID NO: 731	moltype =	length =
SEQUENCE: 731		
000		
SEQ ID NO: 732	moltype =	length =
SEQUENCE: 732		
000		
SEQ ID NO: 733	moltype =	length =
SEQUENCE: 733		
000		
SEQ ID NO: 734	moltype =	length =
SEQUENCE: 734		
000		
SEQ ID NO: 735	moltype =	length =
SEQUENCE: 735		
000		
SEQ ID NO: 736	moltype =	length =
SEQUENCE: 736		
000		

-continued

---

SEQ ID NO: 737	moltype =	length =
SEQUENCE: 737		
000		
SEQ ID NO: 738	moltype =	length =
SEQUENCE: 738		
000		
SEQ ID NO: 739	moltype =	length =
SEQUENCE: 739		
000		
SEQ ID NO: 740	moltype =	length =
SEQUENCE: 740		
000		
SEQ ID NO: 741	moltype =	length =
SEQUENCE: 741		
000		
SEQ ID NO: 742	moltype =	length =
SEQUENCE: 742		
000		
SEQ ID NO: 743	moltype =	length =
SEQUENCE: 743		
000		
SEQ ID NO: 744	moltype =	length =
SEQUENCE: 744		
000		
SEQ ID NO: 745	moltype =	length =
SEQUENCE: 745		
000		
SEQ ID NO: 746	moltype =	length =
SEQUENCE: 746		
000		
SEQ ID NO: 747	moltype =	length =
SEQUENCE: 747		
000		
SEQ ID NO: 748	moltype =	length =
SEQUENCE: 748		
000		
SEQ ID NO: 749	moltype =	length =
SEQUENCE: 749		
000		
SEQ ID NO: 750	moltype =	length =
SEQUENCE: 750		
000		
SEQ ID NO: 751	moltype =	length =
SEQUENCE: 751		
000		
SEQ ID NO: 752	moltype =	length =
SEQUENCE: 752		
000		
SEQ ID NO: 753	moltype =	length =
SEQUENCE: 753		
000		
SEQ ID NO: 754	moltype =	length =
SEQUENCE: 754		
000		
SEQ ID NO: 755	moltype =	length =
SEQUENCE: 755		
000		

-continued

---

SEQ ID NO: 756	moltype =	length =
SEQUENCE: 756		
000		
SEQ ID NO: 757	moltype =	length =
SEQUENCE: 757		
000		
SEQ ID NO: 758	moltype =	length =
SEQUENCE: 758		
000		
SEQ ID NO: 759	moltype =	length =
SEQUENCE: 759		
000		
SEQ ID NO: 760	moltype =	length =
SEQUENCE: 760		
000		
SEQ ID NO: 761	moltype =	length =
SEQUENCE: 761		
000		
SEQ ID NO: 762	moltype =	length =
SEQUENCE: 762		
000		
SEQ ID NO: 763	moltype =	length =
SEQUENCE: 763		
000		
SEQ ID NO: 764	moltype =	length =
SEQUENCE: 764		
000		
SEQ ID NO: 765	moltype =	length =
SEQUENCE: 765		
000		
SEQ ID NO: 766	moltype =	length =
SEQUENCE: 766		
000		
SEQ ID NO: 767	moltype =	length =
SEQUENCE: 767		
000		
SEQ ID NO: 768	moltype =	length =
SEQUENCE: 768		
000		
SEQ ID NO: 769	moltype =	length =
SEQUENCE: 769		
000		
SEQ ID NO: 770	moltype =	length =
SEQUENCE: 770		
000		
SEQ ID NO: 771	moltype =	length =
SEQUENCE: 771		
000		
SEQ ID NO: 772	moltype =	length =
SEQUENCE: 772		
000		
SEQ ID NO: 773	moltype =	length =
SEQUENCE: 773		
000		
SEQ ID NO: 774	moltype =	length =
SEQUENCE: 774		
000		

-continued

---

SEQ ID NO: 775	moltype =	length =
SEQUENCE: 775		
000		
SEQ ID NO: 776	moltype =	length =
SEQUENCE: 776		
000		
SEQ ID NO: 777	moltype =	length =
SEQUENCE: 777		
000		
SEQ ID NO: 778	moltype =	length =
SEQUENCE: 778		
000		
SEQ ID NO: 779	moltype =	length =
SEQUENCE: 779		
000		
SEQ ID NO: 780	moltype =	length =
SEQUENCE: 780		
000		
SEQ ID NO: 781	moltype =	length =
SEQUENCE: 781		
000		
SEQ ID NO: 782	moltype =	length =
SEQUENCE: 782		
000		
SEQ ID NO: 783	moltype =	length =
SEQUENCE: 783		
000		
SEQ ID NO: 784	moltype =	length =
SEQUENCE: 784		
000		
SEQ ID NO: 785	moltype =	length =
SEQUENCE: 785		
000		
SEQ ID NO: 786	moltype =	length =
SEQUENCE: 786		
000		
SEQ ID NO: 787	moltype =	length =
SEQUENCE: 787		
000		
SEQ ID NO: 788	moltype =	length =
SEQUENCE: 788		
000		
SEQ ID NO: 789	moltype =	length =
SEQUENCE: 789		
000		
SEQ ID NO: 790	moltype =	length =
SEQUENCE: 790		
000		
SEQ ID NO: 791	moltype =	length =
SEQUENCE: 791		
000		
SEQ ID NO: 792	moltype =	length =
SEQUENCE: 792		
000		
SEQ ID NO: 793	moltype =	length =
SEQUENCE: 793		
000		

-continued

SEQ ID NO: 794 moltype = length =  
SEQUENCE: 794  
000

SEQ ID NO: 795 moltype = length =  
SEQUENCE: 795  
000

SEQ ID NO: 796 moltype = length =  
SEQUENCE: 796  
000

SEQ ID NO: 797 moltype = length =  
SEQUENCE: 797  
000

SEQ ID NO: 798 moltype = length =  
SEQUENCE: 798  
000

SEQ ID NO: 799 moltype = length =  
SEQUENCE: 799  
000

SEQ ID NO: 800 moltype = length =  
SEQUENCE: 800  
000

SEQ ID NO: 801 moltype = DNA length = 156  
FEATURE Location/Qualifiers  
source 1..156  
mol\_type = other DNA  
organism = synthetic construct  
SEQUENCE: 801  
ggggcggggg ggcggccgca ttcgcgcgca gcccacagg ggggtgtcgc cggccccccc 60  
cgcgcatgcg cggggccccc ccccggggg gtcggcccc cccggggccc ccccggtgtca 120  
aaccacccgc gcatgogcga ccacgcccc gccggc 156

SEQ ID NO: 802 moltype = DNA length = 150  
FEATURE Location/Qualifiers  
source 1..150  
mol\_type = other DNA  
organism = synthetic construct  
SEQUENCE: 802  
ccgagcgta gcgaggagtgc gacccctacc ccctggccca acttcttcgg agccgcgcgc 60  
taaccccttcg gtcgcgcgc gacactcaga ccccgctcgc tgctgacacg cttgcgcgtg 120  
tcagaccact tcgggctcgc ggggttcggg 150

SEQ ID NO: 803 moltype = DNA length = 122  
FEATURE Location/Qualifiers  
source 1..122  
mol\_type = other DNA  
organism = synthetic construct  
SEQUENCE: 803  
gcgcgcgcg cggggggggg cgccgcgcgtg cgccgcgcgc ccagtaggggg gagccatgcg 60  
ccccccccc cgcgcgcgcg gggccccccc ccgggggggg ctccgcgcgc cggccccccc 120  
cg 122

SEQ ID NO: 804 moltype = DNA length = 111  
FEATURE Location/Qualifiers  
source 1..111  
mol\_type = other DNA  
organism = synthetic construct  
SEQUENCE: 804  
cgccccacgc gggggcgccg cgcttcgcgc gggccgggg gggctccggcc ccccccccg 60  
catgcgcggg gccccccccc ggggggggt ccggccccc ggtccccccc g 111

SEQ ID NO: 805 moltype = DNA length = 115  
FEATURE Location/Qualifiers  
source 1..115  
mol\_type = other DNA  
organism = synthetic construct  
SEQUENCE: 805  
cgccgcgtcg gggggcgccg cgcttcgcgc gggccgggg ggctgcgcgc ccccccccg 60  
catgcgcggg gggccccccc ccgggggggg ctccgcgcgc cggccccccc ccccg 115

-continued

---

SEQ ID NO: 806 moltype = DNA length = 104  
FEATURE Location/Qualifiers  
source 1..104  
mol\_type = other DNA  
organism = synthetic construct  
SEQUENCE: 806 cggcgccggc ggcgcgctac cgcgcgcg cgggggggtc ggcgcggccc cccgcgcat 60  
gcccggggcc cccccccgcg gggggctccg ccccccggcc cccc 104  
SEQ ID NO: 807 moltype = DNA length = 108  
FEATURE Location/Qualifiers  
source 1..108  
mol\_type = other DNA  
organism = synthetic construct  
SEQUENCE: 807 ggccgcggcg cgccgcgctac gcgcgcgcgc cggggagctc tgcccccccg cgccatgcg 60  
cgccgggtccg ccccccgcgg ggggatccgc ccccccggtc ccccccgg 108  
SEQ ID NO: 808 moltype = length =  
SEQUENCE: 808  
000  
SEQ ID NO: 809 moltype = length =  
SEQUENCE: 809  
000  
SEQ ID NO: 810 moltype = length =  
SEQUENCE: 810  
000  
SEQ ID NO: 811 moltype = length =  
SEQUENCE: 811  
000  
SEQ ID NO: 812 moltype = length =  
SEQUENCE: 812  
000  
SEQ ID NO: 813 moltype = length =  
SEQUENCE: 813  
000  
SEQ ID NO: 814 moltype = length =  
SEQUENCE: 814  
000  
SEQ ID NO: 815 moltype = length =  
SEQUENCE: 815  
000  
SEQ ID NO: 816 moltype = length =  
SEQUENCE: 816  
000  
SEQ ID NO: 817 moltype = length =  
SEQUENCE: 817  
000  
SEQ ID NO: 818 moltype = length =  
SEQUENCE: 818  
000  
SEQ ID NO: 819 moltype = length =  
SEQUENCE: 819  
000  
SEQ ID NO: 820 moltype = length =  
SEQUENCE: 820  
000  
SEQ ID NO: 821 moltype = length =  
SEQUENCE: 821  
000  
SEQ ID NO: 822 moltype = length =  
SEQUENCE: 822

-continued

---

000  
SEQ ID NO: 823 moltype = length =  
SEQUENCE: 823  
000  
SEQ ID NO: 824 moltype = length =  
SEQUENCE: 824  
000  
SEQ ID NO: 825 moltype = length =  
SEQUENCE: 825  
000  
SEQ ID NO: 826 moltype = length =  
SEQUENCE: 826  
000  
SEQ ID NO: 827 moltype = length =  
SEQUENCE: 827  
000  
SEQ ID NO: 828 moltype = length =  
SEQUENCE: 828  
000  
SEQ ID NO: 829 moltype = AA length = 11  
FEATURE Location/Qualifiers  
VARIANT 4..5  
note = Any amino acid  
VARIANT 7  
note = Any amino acid  
VARIANT 9..10  
note = Any amino acid  
source 1..11  
mol\_type = protein  
REGION organism = synthetic construct  
1..11  
note = Description of Artificial Sequence: Synthetic peptide  
SEQUENCE: 829  
YNPXXDXGXX N  
11  
SEQ ID NO: 830 moltype = length =  
SEQUENCE: 830  
000  
SEQ ID NO: 831 moltype = length =  
SEQUENCE: 831  
000  
SEQ ID NO: 832 moltype = length =  
SEQUENCE: 832  
000  
SEQ ID NO: 833 moltype = length =  
SEQUENCE: 833  
000  
SEQ ID NO: 834 moltype = length =  
SEQUENCE: 834  
000  
SEQ ID NO: 835 moltype = length =  
SEQUENCE: 835  
000  
SEQ ID NO: 836 moltype = length =  
SEQUENCE: 836  
000  
SEQ ID NO: 837 moltype = length =  
SEQUENCE: 837  
000  
SEQ ID NO: 838 moltype = length =  
SEQUENCE: 838

-continued

---

000  
SEQ ID NO: 839 moltype = length =  
SEQUENCE: 839  
000  
SEQ ID NO: 840 moltype = length =  
SEQUENCE: 840  
000  
SEQ ID NO: 841 moltype = length =  
SEQUENCE: 841  
000  
SEQ ID NO: 842 moltype = length =  
SEQUENCE: 842  
000  
SEQ ID NO: 843 moltype = length =  
SEQUENCE: 843  
000  
SEQ ID NO: 844 moltype = length =  
SEQUENCE: 844  
000  
SEQ ID NO: 845 moltype = length =  
SEQUENCE: 845  
000  
SEQ ID NO: 846 moltype = length =  
SEQUENCE: 846  
000  
SEQ ID NO: 847 moltype = length =  
SEQUENCE: 847  
000  
SEQ ID NO: 848 moltype = length =  
SEQUENCE: 848  
000  
SEQ ID NO: 849 moltype = length =  
SEQUENCE: 849  
000  
SEQ ID NO: 850 moltype = length =  
SEQUENCE: 850  
000  
SEQ ID NO: 851 moltype = length =  
SEQUENCE: 851  
000  
SEQ ID NO: 852 moltype = length =  
SEQUENCE: 852  
000  
SEQ ID NO: 853 moltype = length =  
SEQUENCE: 853  
000  
SEQ ID NO: 854 moltype = length =  
SEQUENCE: 854  
000  
SEQ ID NO: 855 moltype = length =  
SEQUENCE: 855  
000  
SEQ ID NO: 856 moltype = length =  
SEQUENCE: 856  
000  
SEQ ID NO: 857 moltype = length =  
SEQUENCE: 857

-continued

---

000  
SEQ ID NO: 858 moltype = length =  
SEQUENCE: 858  
000  
SEQ ID NO: 859 moltype = length =  
SEQUENCE: 859  
000  
SEQ ID NO: 860 moltype = length =  
SEQUENCE: 860  
000  
SEQ ID NO: 861 moltype = length =  
SEQUENCE: 861  
000  
SEQ ID NO: 862 moltype = length =  
SEQUENCE: 862  
000  
SEQ ID NO: 863 moltype = length =  
SEQUENCE: 863  
000  
SEQ ID NO: 864 moltype = length =  
SEQUENCE: 864  
000  
SEQ ID NO: 865 moltype = length =  
SEQUENCE: 865  
000  
SEQ ID NO: 866 moltype = length =  
SEQUENCE: 866  
000  
SEQ ID NO: 867 moltype = length =  
SEQUENCE: 867  
000  
SEQ ID NO: 868 moltype = length =  
SEQUENCE: 868  
000  
SEQ ID NO: 869 moltype = length =  
SEQUENCE: 869  
000  
SEQ ID NO: 870 moltype = length =  
SEQUENCE: 870  
000  
SEQ ID NO: 871 moltype = length =  
SEQUENCE: 871  
000  
SEQ ID NO: 872 moltype = length =  
SEQUENCE: 872  
000  
SEQ ID NO: 873 moltype = length =  
SEQUENCE: 873  
000  
SEQ ID NO: 874 moltype = length =  
SEQUENCE: 874  
000  
SEQ ID NO: 875 moltype = length =  
SEQUENCE: 875  
000  
SEQ ID NO: 876 moltype = length =  
SEQUENCE: 876

---

-continued

---

```

000

SEQ ID NO: 877      moltype =    length =
SEQUENCE: 877
000

SEQ ID NO: 878      moltype =    length =
SEQUENCE: 878
000

SEQ ID NO: 879      moltype =    length =
SEQUENCE: 879
000

SEQ ID NO: 880      moltype =    length =
SEQUENCE: 880
000

SEQ ID NO: 881      moltype =    length =
SEQUENCE: 881
000

SEQ ID NO: 882      moltype =    length =
SEQUENCE: 882
000

SEQ ID NO: 883      moltype =    length =
SEQUENCE: 883
000

SEQ ID NO: 884      moltype =    length =
SEQUENCE: 884
000

SEQ ID NO: 885      moltype =    length =
SEQUENCE: 885
000

SEQ ID NO: 886      moltype = DNA  length = 3176
FEATURE
source
1..3176
mol_type = genomic DNA
organism = Gammatorquevirus sp.
SEQUENCE: 886
taaaatggc ggagccatc atttatact ttcaacttcc aattaaaaat ggccacgtca 60
caaacaaggg gtggagccat ttaaactata taactaagt gggtggcgaa tggctgagtt 120
taccccgcta gacggtcgac ggacccggat gagcgcacgg aggagggtccc cggctgccca 180
tggcggggat ccggaggtag tgaaaccacc gagggtctagg ggcaattccg gctaggccag 240
tctagcggaa cgggaagaa actaaaaca atattgtt tacagatggt tagtatatcc 300
tcaagtgtt ttttaagaa aacgaaattt aatgaggaga cgcagaacca agtatggat 360
tctcaattt ctgactctca tgataatato tcgagtttgc ggcattccatt tgctcacct 420
cttgcttcca tatttccctc tggccaaaaa gatcgtgatc ttacttaa ccaaattctt 480
cttaagagatt ataaaagaaa atgcatttctt ggtggagaag aaggagaaaa ttctggacca 540
acaacaggtt taatttacc accatcaca aaaagaagaa gatataaaaa aagatggccc agaaggcc 600
gcagaagaag accatcaca cgcctgttc gccgcgcggc tagaaaaactt cgaaaaggtaa 660
agagaaaaaaa aaaatttta attgttagac aatggcaacc acagactata agaacttgta 720
aaattatagg acagtcaatc atgttgtt gggctgatc aagcaaatg tacttgtata 780
ctgtcaataa gtttaattat gtgccttccaa aaaccaccata tgggggaggg tttggatgt 840
accaacacac actgaaataac ttatgtatgaa aatacagatc tgcacaaaaac atttggacac 900
aatctaatgt actgaaagac ttatgttagat acataaaatgt taactcta ttctacagag 960
acaacaaaaac acactttgc ctttctatg acagaaaaacc acctttccaa ctaacaaaat 1020
ttacataccc aggacacac ccacaacaaa tcatgttca aaaacaccac aaattcatac 1080
tatcacaat gacaaaggctt aatggaaagac taacaaaaaa actcaaaattt aaacctctta 1140
acaaaatgtt ttctaatgg ttctttcaaa aacaatttctg taaataccct ttactatctc 1200
ttaaagcttc tgcactagac cttaggcact cttaccttagg ctgctgtat gaaaatccac 1260
aggtatttt ttattttta aaccatggat actacacaaat aacaaactgg ggagccaaat 1320
cctcaacagc atacagacct aactocaagg tgacagacac aacataactac agataaaaa 1380
atgacagaaaa aataatataac attaaaagcc atgaataatcga aaaaagtata tcatatgaaa 1440
acggttatcc tcaatcttagt ttcttacaaa cacagtcat atataccgt gagcgtggtg 1500
aaggctgtat agcagaaaaaa ccacttagaa tagctattta caatccgatgaaagacaatg 1560
gagatggtaa tatgtatata cttgtaaagca cttctagaaat cacttggac cagctccaa 1620
aagacagtgc tattttataa caaggagtac ccattatggct aggcttattt ggatatttag 1680
actactgttag acaaattaaa gctgacaaaa catggctaga cagtcatgtt ctgtatcc 1740
aaagtctgc tattttact taccctaaatc caggagcagg caaatggat tgcactat 1800
cacaagttt tataatggc aatggccgt ttaatcaacc acctacactg ctacaaaaag 1860
cacaagttt tccacaaaata caataccaaac aagaaattat taatagctt stagaatcag 1920

```

---

-continued

---

gaccatttgt	tcccaaataat	gcaaatcaaa	ctgaaagcaa	ctgggaacta	aaataataat	1980
atgttttac	atthaatgg	ggtgaccac	aattccatga	accagaatt	gctgacccta	2040
gcaaaaacaaga	gcagtatgt	gtccccgata	ctttctacca	aacaatacaa	attgaagatc	2100
cagaaggaca	agaccccaga	tctctcatcc	atgattggga	ctacagacga	ggcttattta	2160
aagaaaagatc	tcttaaaaga	atgtcaactt	acttctcaac	tcatacagat	cagcaagcaa	2220
cttcagagga	agacattccc	aaaaagaaaa	agagaattgg	accccaactc	acagtcccac	2280
aacaaaaaaga	agaggagaca	ctgtcatgtc	tctctctctc	ctgcaaaaaaa	gataccttc	2340
aagaaaacaga	gacacaagaa	gacctccago	agtcatcaa	gcagcagcag	gagcagcagc	2400
tccteectca	gagaaaacatc	cttcagctca	tccacaact	aaaagagaat	caacaatgc	2460
tttagcttca	cacaggcatg	ttaccttaad	caagatttaaa	ctggatttgc	aagagcaac	2520
agagagagaa	tttagcaatta	tatttcatag	gccccctaga	acctacaatag	aggacttcc	2580
attctatccc	ttggctaccac	ctgcacccct	tgtacaattt	aaccttaact	tcaaaggcta	2640
ggccaacaat	gtacacttag	taaagcatgt	ttattaaagc	acaacccca	aaataaatgt	2700
aaaaataaaa	aaaaaaaaaa	aaaaataaaa	aattgcaaaa	attcggcgct	cgcgccatg	2760
tgcgcctcg	tgccaaatca	cgcaacgcgt	gcccgcgcgc	gtatgtctt	ttaccacgca	2820
cctagatgg	ggtgcgcgcg	ctagcgcgcg	caccccaatg	ccccccgcgc	tcgttccgac	2880
ccgcttgcgc	gggtcgacc	acttcgggt	cggggggggc	cgccctgcggc	gctttttac	2940
taaacagact	ccgagccgcg	atttgcggcc	ctaagctccg	ccccccat	gaatattcat	3000
aaaggaaacc	acataatagg	aattgcgcac	cacaaactgc	catatgcata	ttagttcccc	3060
ttttacaaag	taaaaaggga	agtAACATA	ccccccacacc	cgccaggcgc	aggcccccgc	3120
cccttacgtc	actaaccacg	ccccccgcgc	catcttgggt	gcccggggc	gggggc	3176

SEQ ID NO: 887            moltype = AA length = 124  
 FEATURE                    Location/Qualifiers  
 source                    1..124  
 mol\_type = protein  
 organism = Gammatorquevirus sp.

SEQUENCE: 887  
 MVSISSSDFF KKTGFNEETQ NQVWMSQIAD SHDNICSCWH PFAHLLASIF PPGHKDRDLT 60  
 INQILLRDYK EKCHSGGEEG ENSGPTTGLI TPKEEDIEKD GPEGAAEEDH TDALFAAAVE 120  
 NFER 124

SEQ ID NO: 888            moltype = AA length = 271  
 FEATURE                    Location/Qualifiers  
 source                    1..271  
 mol\_type = protein  
 organism = Gammatorquevirus sp.

SEQUENCE: 888  
 MVSISSSDFF KKTGFNEETQ NQVWMSQIAD SHDNICSCWH PFAHLLASIF PPGHKDRDLT 60  
 INQILLRDYK EKCHSGGEEG ENSGPTTGLI TPKEEDIEKD GPEGAAEEDH TDALFAAAVE 120  
 NFESGVDHNS MNQKLTLAN KSSMMSPILS TKQYKLKIQK DKTPDLSMI GTTDEALLKK 180  
 DLLKECOLTS QLIQISKQLQ RKTFPKRKR LDPNSQSHNK KKRRHCHVSS LSAKKIPSKK 240  
 QRHKKTSSS SSSSRSSSS SRETSSSSST N 271

SEQ ID NO: 889            moltype = AA length = 267  
 FEATURE                    Location/Qualifiers  
 source                    1..267  
 mol\_type = protein  
 organism = Gammatorquevirus sp.

SEQUENCE: 889  
 MVSISSSDFF KKTGFNEETQ NQVWMSQIAD SHDNICSCWH PFAHLLASIF PPGHKDRDLT 60  
 INQILLRDYK EKCHSGGEEG ENSGPTTGLI TPKEEDIEKD GPEGAAEEDH TDALFAAAVE 120  
 NFERSASNFR GRHSQKEKEN WTPTHSPPTK RRGDTVMSPL SLQKRYLPRN RDTRRPPAAH 180  
 QAAAGAAAAPP QEKHPPAHPQ TKRESTNASA SHRHTLTFR KPGFEEQTER ELAIIFHRPP 240  
 RTYKEDLPFY PWLPPAPLVQ FNLFNPKG 267

SEQ ID NO: 890            moltype = AA length = 50  
 FEATURE                    Location/Qualifiers  
 source                    1..50  
 mol\_type = protein  
 organism = Gammatorquevirus sp.

SEQUENCE: 890  
 MRRRRTKYGC LKLLTLMIIS AVAGIHLTF LLPYFLLATK IVILLLTKFF 50

SEQ ID NO: 891            moltype = AA length = 662  
 FEATURE                    Location/Qualifiers  
 source                    1..662  
 mol\_type = protein  
 organism = Gammatorquevirus sp.

SEQUENCE: 891  
 MPFWWRRLRK FWTNNRFNYT KRRRYKRWP RRRRRRPRY RPVRRRRK RLKVKKKSL 60  
 IVRQWQPSDI RTCKIIQGSA IVVGAEGKQM YCYTVNKLIN VPPKTPYGGG FGVDQYTLKY 120  
 LYBEYRFQAQN IWTQSNVLKD LCRYINVKL FYRDNKTDVF LSYDRNPFPQ LTKFTYPGAH 180  
 PQQIMLQKHH KFILSQMTKP NGRLTKKLKI KPPKQMLSKW FFSKQFCCKYP LLSLKASALD 240  
 LRHSYLGCCN ENPQVFFYYL NHGYYTITNW GAQSSTAYRP NSKVTDTTTY RYKNDRKNIN 300  
 IKSHYEYKSI SYENGYFQSS FLQTQCIYTS ERGEACIAEK PLGIAIYNPV KDNGDGNMIV 360

---

-continued

---

LVSTLANTWD QPPKDSAILI QGVPPIWLGLF GYLDYCRQIK ADKTWLD SHV LVIQSPAIFT	420
YPNPGAGKWWY CPLSQSFING NGPFNQPPTL LQAKAWFPOI QYQQEINSF VESGFVPKY	480
ANQTESNWEL KYKYVFTFKW GGPQFHEPEI ADPSKQEYD VPDTFYQTIQ IEDPEGQDPR	540
SLIHWDWYRR GFIKERSLKK MSTYFSTHTD QQATSEEDIP KKKKRIGPQL TVPQQKEEET	600
LSCLLSLCKK DTFQETETQE DLQQLIKQQQ EQQQLLKRN1 LQLIHKLKEN QQMLQLHTGM	660
LP	662
 SEQ ID NO: 892	moltype = AA length = 215
FEATURE	Location/Qualifiers
source	1..215
	mol_type = protein
	organism = Gammatorquevirus sp.
SEQUENCE: 892	
MPFWWRRRK FWTNNRNFNYT KRRRYRKRWP RRRRRRPRYR RPVRRRRKRL RKWGGPQFHE	60
PEIADPSKQE QYDVPDTFYQ TIQIEDPEGQ DPRSLIHWD YRGFIFIERS LKRMSTYFST	120
HTDQQATSEE DIPKKKKRIG PQLTVQQKE EETLSCLLLS CKKDTFQETE TQEDLQQLIK	180
QQQEQQQLLK RNILQLIHKL KENQQMLQLH TGMLP	215
 SEQ ID NO: 893	moltype = AA length = 129
FEATURE	Location/Qualifiers
source	1..129
	mol_type = protein
	organism = Gammatorquevirus sp.
SEQUENCE: 893	
MPFWWRRRK FWTNNRNFNYT KRRRYRKRWP RRRRRRPRYR RPVRRRRKRL RKISKQLQRK	60
TFPKRKRELD PNSQSHNKKK RRHCHVSSL AKKIPSKKQR HKKTSSSSS SSRSSSSSR	120
ETSSSSSTN	129
 SEQ ID NO: 894	moltype = length =
SEQUENCE: 894	
000	
 SEQ ID NO: 895	moltype = length =
SEQUENCE: 895	
000	
 SEQ ID NO: 896	moltype = length =
SEQUENCE: 896	
000	
 SEQ ID NO: 897	moltype = length =
SEQUENCE: 897	
000	
 SEQ ID NO: 898	moltype = length =
SEQUENCE: 898	
000	
 SEQ ID NO: 899	moltype = length =
SEQUENCE: 899	
000	
 SEQ ID NO: 900	moltype = length =
SEQUENCE: 900	
000	
 SEQ ID NO: 901	moltype = length =
SEQUENCE: 901	
000	
 SEQ ID NO: 902	moltype = length =
SEQUENCE: 902	
000	
 SEQ ID NO: 903	moltype = length =
SEQUENCE: 903	
000	
 SEQ ID NO: 904	moltype = length =
SEQUENCE: 904	
000	
 SEQ ID NO: 905	moltype = length =
SEQUENCE: 905	
000	

-continued

---

SEQ ID NO: 906	moltype =	length =
SEQUENCE: 906		
000		
SEQ ID NO: 907	moltype =	length =
SEQUENCE: 907		
000		
SEQ ID NO: 908	moltype =	length =
SEQUENCE: 908		
000		
SEQ ID NO: 909	moltype =	length =
SEQUENCE: 909		
000		
SEQ ID NO: 910	moltype =	length =
SEQUENCE: 910		
000		
SEQ ID NO: 911	moltype =	length =
SEQUENCE: 911		
000		
SEQ ID NO: 912	moltype =	length =
SEQUENCE: 912		
000		
SEQ ID NO: 913	moltype =	length =
SEQUENCE: 913		
000		
SEQ ID NO: 914	moltype =	length =
SEQUENCE: 914		
000		
SEQ ID NO: 915	moltype =	length =
SEQUENCE: 915		
000		
SEQ ID NO: 916	moltype =	length =
SEQUENCE: 916		
000		
SEQ ID NO: 917	moltype =	length =
SEQUENCE: 917		
000		
SEQ ID NO: 918	moltype =	length =
SEQUENCE: 918		
000		
SEQ ID NO: 919	moltype =	length =
SEQUENCE: 919		
000		
SEQ ID NO: 920	moltype =	length =
SEQUENCE: 920		
000		
SEQ ID NO: 921	moltype =	length =
SEQUENCE: 921		
000		
SEQ ID NO: 922	moltype =	length =
SEQUENCE: 922		
000		
SEQ ID NO: 923	moltype =	length =
SEQUENCE: 923		
000		
SEQ ID NO: 924	moltype =	length =
SEQUENCE: 924		
000		

-continued

---

SEQ ID NO: 925	moltype = AA length = 662
FEATURE	Location/Qualifiers
source	1..662
	mol_type = protein
	organism = Gammatorquevirus sp.
SEQUENCE: 925	
MPFWWWRRRK FWTNNRFNYT KRRRYRKWP RRRRRRPPYR RPVRRRRKRL RKVKRKKKSL	60
IVRQWQPSI RTCKIIGQSA IVVGAEGKQM YCYTVNKLIN VPPKTPYGGG FGVDQYTLKY	120
LYEEYRFQAQN IWTQSNVLKD LCRYINVKLI FYRDNKTDVF LSYDRNPPFQ LTKFTYPGAH	180
PQQIMLQKH KFILSQMTKP NGRLTKKLKI KPPKQMLSKW FFSKQFCYQP LLSLKASALD	240
LRHSYLGCCN ENPQVFFYI NHGYYTITNW GAQSSTAYRP NSKVTDTYY RYKNDRKNIN	300
IKSHEYEKSI SYENGYFQSS FLQTQCIYTS ERGEACTAEK PLGIAIYNPV KDNGDGNMIY	360
LVSTLANTWD QPPKDSAILI QGVPWLGLF GYLDYCRQIK ADKTWLDSHV LVIQSPAIFT	420
YMPNGAGKWKY CPLSQSFING NGPFPNPPTL LQAKAKWFPOI QYQQEIINSF VESGFVPKY	480
ANQTESNWEI KYKVFTFKW GGPQFHPEI ADPSKQEQYD VPDTFYQTIQ IEDPEGQDPR	540
SILHDWDYRR GFIKERSLK MSTYTFSTHTD QQATSEEDIP KKKKRIGPQL TVPQQKEEET	600
LSCLLSLCKK DTFQETETQE DLQQLIKQQ EQQLLLKRNI LQLIHKLKEN QQMLQLHTGM	660
LP	662
SEQ ID NO: 926	moltype = AA length = 58
FEATURE	Location/Qualifiers
source	1..58
	mol_type = protein
	organism = Gammatorquevirus sp.
SEQUENCE: 926	
MPFWWWRRRK FWTNNRFNYT KRRRYRKWP RRRRRRPPYR RPVRRRRKRL RKVKRKKK	58
SEQ ID NO: 927	moltype = AA length = 202
FEATURE	Location/Qualifiers
source	1..202
	mol_type = protein
	organism = Gammatorquevirus sp.
SEQUENCE: 927	
SLIVRQWQPD SIRTCKIIGQ SAIVVGAEKG QMYCYTVNKL INVPPKTPYG GGFGVDQYTL	60
KYLYEEYRFQAQN QNIWTQSNVL KDLCRYINVK LIFYRDNKTD FVLSYDRNPP FQLTKFTYPG	120
AHQQQIMLQK HHKFILSQMT KPNGRLTKKL KIKPPKQMLS KWFFSKQFCYK YPLLSLKASA	180
LDLRHSGYLGC CNENPQVFFY YL	202
SEQ ID NO: 928	moltype = AA length = 79
FEATURE	Location/Qualifiers
source	1..79
	mol_type = protein
	organism = Gammatorquevirus sp.
SEQUENCE: 928	
NHGGYTITNW GAQSSTAYRP NSKVTDTYY RYKNDRKNIN IKSHEYEKSI SYENGYFQSS	60
FLQTQCIYTS ERGEACIAE	79
SEQ ID NO: 929	moltype = AA length = 160
FEATURE	Location/Qualifiers
source	1..160
	mol_type = protein
	organism = Gammatorquevirus sp.
SEQUENCE: 929	
KPLGIAIYNP VKDNGDGNMI YLVSTLANTW DQPPKDSAIL IQGVPWLGL FGYLDYCRQI	60
KADKTWLDSH VLVIQSPAIF TYPNGAGKWKY YCPLSQSFING GNPGFNQPPT LLQAKAKWFPO	120
IQYQQEIINS FVESGFVPK YANQTESNWE LYKVFTFKW	160
SEQ ID NO: 930	moltype = AA length = 163
FEATURE	Location/Qualifiers
source	1..163
	mol_type = protein
	organism = Gammatorquevirus sp.
SEQUENCE: 930	
WGGPQFHPEI IADPSKQEQYD DVDPDTFYQTQ IEDPEGQDPR RSLIHDWDYR RGFIKERSLK	60
RMSTYFSTHTD QQATSEEDI KKKKRIGPQL TVPQQKEEET TLSCLLSLCK KDTFQETETQ	120
EDLQQLIKQQ EQQQLLLKRNI LQLIHKLKE NQQMLQLHTGM MLP	163
SEQ ID NO: 931	moltype = length =
SEQUENCE: 931	000
SEQ ID NO: 932	moltype = length =
SEQUENCE: 932	000
SEQ ID NO: 933	moltype = length =

-continued

---

SEQUENCE: 933  
000

SEQ\_ID NO: 934 moltype = length =  
SEQUENCE: 934  
000

SEQ\_ID NO: 935 moltype = length =  
SEQUENCE: 935  
000

SEQ\_ID NO: 936 moltype = length =  
SEQUENCE: 936  
000

SEQ\_ID NO: 937 moltype = length =  
SEQUENCE: 937  
000

SEQ\_ID NO: 938 moltype = length =  
SEQUENCE: 938  
000

SEQ\_ID NO: 939 moltype = length =  
SEQUENCE: 939  
000

SEQ\_ID NO: 940 moltype = length =  
SEQUENCE: 940  
000

SEQ\_ID NO: 941 moltype = length =  
SEQUENCE: 941  
000

SEQ\_ID NO: 942 moltype = length =  
SEQUENCE: 942  
000

SEQ\_ID NO: 943 moltype = length =  
SEQUENCE: 943  
000

SEQ\_ID NO: 944 moltype = length =  
SEQUENCE: 944  
000

SEQ\_ID NO: 945 moltype = length =  
SEQUENCE: 945  
000

SEQ\_ID NO: 946 moltype = length =  
SEQUENCE: 946  
000

SEQ\_ID NO: 947 moltype = length =  
SEQUENCE: 947  
000

SEQ\_ID NO: 948 moltype = length =  
SEQUENCE: 948  
000

SEQ\_ID NO: 949 moltype = AA length = 21  
FEATURE Location/Qualifiers  
VARIANT 1  
note = W or F  
VARIANT 2..8  
note = Any amino acid  
VARIANT 10..12  
note = Any amino acid  
VARIANT 14  
note = Any amino acid  
VARIANT 16..20  
note = Any amino acid  
source 1..21

---

-continued

---

```

mol_type = protein
organism = synthetic construct
1..21
note = Description of Artificial Sequence: Synthetic peptide
REGION
SEQUENCE: 949
XXXXXXXXXXHXXXXCXXXXXX H                                         21

SEQ ID NO: 950      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
mol_type = protein
organism = synthetic construct
REGION
VARIANT          1..22
note = Description of Artificial Sequence: Synthetic peptide
6..7
note = Any amino acid
VARIANT          9
note = Any amino acid
VARIANT          15..16
note = Any amino acid
SEQUENCE: 950
YNCSPXXDXG ASKRXXNTSV AK                                         22

SEQ ID NO: 951      moltype = DNA  length = 51
FEATURE          Location/Qualifiers
misc_difference  45
note = May be absent
source           1..51
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 951
aggtagtga aaccaccgaa gtcaaggggc aattcgggct agggacagt c t                                         51

SEQ ID NO: 952      moltype = DNA  length = 50
FEATURE          Location/Qualifiers
source           1..50
mol_type = other DNA
organism = Alphatorquevirus sp.
SEQUENCE: 952
aggtagttt acacacccgca gtcaaggggc aattcgggct cgggactggc                                         50

SEQ ID NO: 953      moltype = DNA  length = 50
FEATURE          Location/Qualifiers
source           1..50
mol_type = other DNA
organism = Betatorquevirus sp.
SEQUENCE: 953
aggtagtga aaccaccgaa gtcaaggggc aattcgggct agatcagtct                                         50

SEQ ID NO: 954      moltype = DNA  length = 50
FEATURE          Location/Qualifiers
source           1..50
mol_type = other DNA
organism = Gammatorquevirus sp.
SEQUENCE: 954
aggtagtga aaccaccgag gtctaggggc aattcgggct agggcagtct                                         50

SEQ ID NO: 955      moltype = AA  length = 6
FEATURE          Location/Qualifiers
REGION
source           1..6
note = Description of Artificial Sequence: Synthetic 6xHis
tag
1..6
mol_type = protein
organism = synthetic construct
SEQUENCE: 955
HHHHHHH                                         6

SEQ ID NO: 956      moltype = AA  length = 237
FEATURE          Location/Qualifiers
REGION
source           1..237
note = Description of Unknown: Anelloviridae family sequence
1..237
mol_type = protein
organism = unidentified
SEQUENCE: 956

```

---

-continued

---

RRKLRVRKRF YKRKLKKIVL KQFQPKIIRR CTIFGTICLF QGSUPERANNN YIQTIISYVP 60  
DEKEPGGGGWT LITESLSSLW EDWEHLKNVW TQSNAGLPLW RYGGVTLYFY QSAYTDYIAQ 120  
VFNCYCPTDT KYTHADSAPN RMLLKKHVIR VPSRETRKKR PKYKRVVRGP PSQMQNWKWF 180  
QRDICEIPLI MIAATAVDFR YPFCASDCAS NNLTLCNLNP LLFQNQDFDH PSDTQGY 237

SEQ ID NO: 957 moltype = AA length = 232  
FEATURE Location/Qualifiers  
REGION 1..232  
note = Description of Unknown: Anelloviridae family sequence  
source 1..232  
mol\_type = protein  
organism = unidentified

SEQUENCE: 957  
RKRRKRVPRTY TTIPLKQWP PYKRTCYIKG QDCLIYYSNL RLGMNSTMYE KSIVPVHWP 60  
GGSFVSVSLMT LDALYDIHKL CRNWNTSTNQ DLPLVRYKG C KITFYQSTFT DYIVRIHTEL 120  
PANSNKLTYP NTHPLMMMS KYKHIIPSQO TRRKKKPYTK IFVKPPPQFE NKWYFATDLY 180  
KIPLLQIHCT ACNLQNPVFK PDKLSNNVTL WSLNTISION RNMSVVDQGQS WP 232

SEQ ID NO: 958 moltype = AA length = 238  
FEATURE Location/Qualifiers  
REGION 1..238  
note = Description of Unknown: Anelloviridae family sequence  
source 1..238  
mol\_type = protein  
organism = unidentified

SEQUENCE: 958  
RRRRRWKRKP TVRRKLKKLT IQQWQPKTIR KCCIQGLHCL FLVTEDTISR NYRMYEHSYT 60  
GEYYPGGGGF SITRYSLDGL YEHQHQLDRNW WTNNTNLPL VRYTGCKIKF YQSWSVDYIC 120  
NYSLTWPVMV TQLLYQSCQP SFMMMNKNSI MIPSKLTKPI KKGYKTIKEK PPHEMLNRWY 180  
FAKDLSKVGL LMLTAASASF DHYYQATDSL SNNCTFESLN PYFYMRHDFI LFPVTGYV 238

SEQ ID NO: 959 moltype = AA length = 238  
FEATURE Location/Qualifiers  
REGION 1..238  
note = Description of Unknown: Anelloviridae family sequence  
source 1..238  
mol\_type = protein  
organism = unidentified

SEQUENCE: 959  
RRRRRWVRK PFYKRKIKRL NIVEWQPKSI RKCRIKGMLC LFQTTEDRLS YNFDMYEESI 60  
IPEKLPGGGF FSIKNISLYA LYQEHIHAWN IFTHTNTDRP LARYTGCSLK FYQSKDIDYV 120  
VTVYSTSLPLR SSMGMYNNSMQ PSIHLMQQN K LIVPSKQTQK RRKPYIKKHI SPPTQMKSQW 180  
YFQHNIAINIP LLMIRTALT LDNYYIGSRQ LSTNVTIHTL NTYIQNRDW GDRNKTYY 238

SEQ ID NO: 960 moltype = AA length = 240  
FEATURE Location/Qualifiers  
REGION 1..240  
note = Description of Unknown: Anelloviridae family sequence  
source 1..240  
mol\_type = protein  
organism = unidentified

SEQUENCE: 960  
RPFVRRRRRK LRKVKRKKS LIVRQWQPD S INTCKIIGQS AIVVGAEGKQ MYCYTVNKL 60  
NVPPKTPYGG GFGVDQYTLK YLYEEYRFAQ NIWTQSNVLK DLCRYINVKL IFYRDNKTDF 120  
VLSYDRNPNF QLTKFTYPGA HPQQIMLQKH HKFILSQMTK PNGRLTKKLK IKPPKQMLSK 180  
WFFSKQFCKY PLLSLKASAL DLRHSYLGCC NENPQVFFYY LNHGYYTITN WGAQSSTAYR 240

SEQ ID NO: 961 moltype = AA length = 240  
FEATURE Location/Qualifiers  
REGION 1..240  
note = Description of Unknown: Anelloviridae family sequence  
source 1..240  
mol\_type = protein  
organism = unidentified

SEQUENCE: 961  
RFTKTRRRRK RKKVRRKLKK ITIKQWQPD S VKKCKIKGYS TLVMGAQGKQ YNCYTNQASD 60  
YVQPKAPQGG GFGCEVFNLK WLYQEYTAHR NIWTKTNEYT DLCRYTGAQI ILYRHPDVDF 120  
IVSWDNQPF LLNKYTYPEL QPQNLLARR KRIILSQKSN PKGKLRIKLR IPPPKQMITK 180  
WFFQRDFCDV NLFKLCASAA SFRYPGISHG AQSTIFSAYA LNTDFYQCSD WCQTNTEGY 240

SEQ ID NO: 962 moltype = AA length = 239  
FEATURE Location/Qualifiers  
REGION 1..239  
note = Description of Unknown: Anelloviridae family sequence  
source 1..239  
mol\_type = protein

-continued

---

```

SEQUENCE: 962          organism = unidentified
GRRTYTRRAV RRRRPRKRL VLTQWSPQTV RNCSIRGIVP MVICGHTKAG RNYAIHSDF 60
TTCIQPFGGS FSTTTWSLVN LWDEHQKFQN RWSYPNQTLQD LARYRGVTFW FYRDQKTDYI 120
VQWSRNPPFK LNKYSSAMYH PGMMMQAKRK LVVPSFQTRP KGKKRYRTI KPPNMFADKW 180
YTQEDLCPVPV LVQIVVSAAS LLHPFCPPQT NNPCITFQVL KDIYDECIGV NETMKDKYK 239

SEQ ID NO: 963          moltype = AA length = 240
FEATURE           Location/Qualifiers
REGION            1..240
note = Description of Unknown: Anelloviridae family sequence
source             1..240
mol_type = protein
organism = unidentified

SEQUENCE: 963          moltype = AA length = 240
RERRRRRGRRR RRRRRRHKPT LILRQWQPD C IRHCKITGWM PLIICGKGST QFNYITHADD 60
ITPRGASYGG NFTNMTFSLE AIYEQFLYHR NRWSASNHDL ELCRYKGTTL KLYRHPEVDY 120
IVVYSRTGPF EISHMTYLST HPMILMLLNKH HIVVPSLTKT PRGRKAIKVR IRPPKLMNNK 180
WYFTRDFCNI GLFQLWATGL ELRNPNWLMS TLSPCIGFNVN LKNSIYTNLS NLPQYKNERL 240

SEQ ID NO: 964          moltype = AA length = 233
FEATURE           Location/Qualifiers
REGION            1..233
note = Description of Unknown: Anelloviridae family sequence
source             1..233
mol_type = protein
organism = unidentified

SEQUENCE: 964          moltype = AA length = 233
RERRRRFRRRG RKAKLIIKLW QPAVIKRCRI KGVIPLIIS NGTFATNFTS HINDRIMKGP 60
FGGGHSTMRF SLYILFEHL RHMNWTWRSN DNLELTRYLG ASVKIYRHPD QDFIVIYNRR 120
TPLGGNIYTA PSLHPGNAIL AKHKILVPSL QTRPKGRKAI RLRIAPPLTF TDKWYFQKDI 180
ADLTLFNIMA VEADLRFPPFC SPQTDNTCIS FQVLSSVYNN YLSINTFNNND NSD      233

SEQ ID NO: 965          moltype = AA length = 233
FEATURE           Location/Qualifiers
REGION            1..233
note = Description of Unknown: Anelloviridae family sequence
source             1..233
mol_type = protein
organism = unidentified

SEQUENCE: 965          moltype = AA length = 233
RRWRRKGKRS RKKKIIIRQW QPNYTRRCNI VGYMPLLICG ENTVATNYAT HSDDSYYPGP 60
FGGGMTTDKF TLRLIYDEYK RFMNYWTSSN EDLDLCRYLG CTLYVFRHPE VDFIIIINTS 120
PPFLDTEITG PSIHPGMAL NKRSRWPISI KNRPGRKHYI KIKVGAPRMF TDKWYFQTDL 180
CDMTLLTIFA SAADMQYPFG SPLTDTIVVS FQVLQSMYND CLSVLPDNFA ETS      233

SEQ ID NO: 966          moltype = AA length = 232
FEATURE           Location/Qualifiers
REGION            1..232
note = Description of Unknown: Anelloviridae family sequence
source             1..232
mol_type = protein
organism = unidentified

SEQUENCE: 966          moltype = AA length = 232
RRWKRKGRR RKAKIIRQW QPNYTRRCNI VGYLPILICG GNTVSRNYAT HSDDTNYPGP 60
FGGGMTTDKF SLRILYDEYK RFMNYWTASN EDLDLCRYLG CTFYFFRHPE VDFIIKINTM 120
PPFLDTTITA PSIHPGLMAL DKRARWIPSL KNRPGKKHYI KIRVGAPKMF TDKWYFQTDL 180
CDMTLLTIFA TAADMQYPFG SPLTDVVVN SQVLQSMYDE TISILPDEKT KR      232

SEQ ID NO: 967          moltype = AA length = 170
FEATURE           Location/Qualifiers
source             1..170
mol_type = protein
organism = Beak and feather disease virus

SEQUENCE: 967          moltype = AA length = 170
RRRRRRFSTN RIYTLRLTRQ FQFKINKOLI FNADYITFAL DDFLQAVPNP HTLNPFEDYRI 60
KLAKMEMRPT GGHYTGFGHT AVIQDSRITR FKTTADPLAP FDGAKKWFVS RGFKRLLRPK 120
PQITIEGPNS AGTKVRHYGI AFSFPQPEQT VTKLTLVQF RQFAPNNPST      170

SEQ ID NO: 968          moltype = AA length = 179
FEATURE           Location/Qualifiers
source             1..179
mol_type = protein
organism = Orthohepevirus A

SEQUENCE: 968          moltype = AA length = 179
GAILRRQYNL STSPLTSSVA SGTNLVLVYAA PLNPLLPLQD GTNTHIMATE ASNYAQYRVV 60

```

-continued

---

RATIRYRPLV PNAVSISFWP QTTTPTSV	D MNSITSTDVR ILVQPGIASE LVIPSERLHY	120
RTRTGVAAEE ATSGLVMLCI HGSPVNSALG LLDFALELEF RNLTPGNTNT RVSRYTSTA		179

---

What is claimed is:

1. A particle comprising:
  - a) a proteinaceous exterior comprising an Anellovirus open reading frame 1 (ORF1) molecule; and
  - b) a genetic element comprising RNA and encoding an exogenous effector, wherein the genetic element is enclosed within the proteinaceous exterior; and
  - wherein the genetic element is protected from digestion by an RNase; and
  - wherein the genetic element binds the Anellovirus ORF1 molecule.
2. The particle of claim 1, wherein the exogenous effector comprises a therapeutic polypeptide.
3. The particle of claim 1, wherein the exogenous effector comprises a human protein.
4. The particle of claim 1, wherein the exogenous effector comprises:
  - (i) a cytosolic polypeptide or cytosolic peptide;
  - (ii) a regulatory intracellular polypeptide;
  - (iii) a secreted polypeptide or peptide;
  - (iv) a protein replacement therapeutic;
  - (v) an enzyme;
  - (vi) a component of a gene editing system;
  - (vii) a membrane receptor; or
  - (viii) a membrane transporter.
5. The particle of claim 1, wherein the genetic element comprises a chemically modified ribonucleotide.
6. The particle of claim 1, wherein the genetic element consists of or consists essentially of RNA.
7. The particle of claim 1, wherein:
  - (a) the genetic element is about 50-60, 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, or 4000-4500 nucleotides in length; and/or
  - (b) the sequence encoding the exogenous effector is at least about 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, or 3000 nucleotides in length.
8. The particle of claim 1, wherein the particle comprises a plurality of genetic elements.
9. The particle of claim 8, wherein the genetic elements of the plurality each comprise the same sequence.
10. The particle of claim 1, wherein the RNA is an mRNA.
11. The particle of claim 10, wherein the genetic element comprises an mRNA cap and/or a poly-A tail.
12. The particle of claim 11, wherein the poly-A tail is at least about 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 adenosines in length.
13. The particle of claim 1, wherein the Anellovirus ORF1 molecule comprises an amino acid sequence of any one of SEQ ID NOs: 21, 58, 891, 1005, or 10<sup>12</sup>, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.
14. The particle of claim 1, wherein the genetic element:
  - (i) lacks a sequence encoding an Anellovirus ORF1 protein;
  - (ii) lacks a sequence encoding an Anellovirus ORF2 protein; and/or
  - (iii) lacks a sequence encoding an Anellovirus ORF3 protein.
15. The particle of claim 1, wherein the genetic element comprises a 5' UTR.
16. A composition comprising a plurality of the particles of claim 1.
17. A pharmaceutical composition comprising the composition of claim 16, and a pharmaceutically acceptable carrier or excipient.
18. A method of making a particle according to claim 1, the method comprising:
  - (a) providing a mixture comprising:
    - (i) a genetic element comprising RNA, and
    - (ii) an Anellovirus open reading frame 1 (ORF1) molecule; and
  - (b) incubating the mixture under conditions suitable for enclosing the genetic element within a proteinaceous exterior comprising the Anellovirus ORF1 molecule, thereby making the particle.
19. The method of claim 18, wherein the mixture is not comprised in a cell.
20. A method of purifying a particle, the method comprising:
  - (a) providing a particle of claim 1; and
  - (b) purifying the particle.
21. A method of treating a disease or disorder in a subject in need thereof, the method comprising administering to the subject the particle of claim 1, thereby treating a disease or disorder in the subject.

\* \* \* \* \*