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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 August 2022 (11.08.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/170195 A1

(51) International Patent Classification:
A61K 48/00 (2006.01) *C12N 15/86* (2006.01)
C12N 7/00 (2006.01)

(21) International Application Number:
PCT/US2022/015499

(22) International Filing Date:
07 February 2022 (07.02.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/147,102 08 February 2021 (08.02.2021) US

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

(72) Inventors: **DELAGRAVE, Simon**; 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US). **LEBO, Kevin James**; 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US). **DIBIASIO-WHITE, Michael James**; 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US). **NAWANDAR, Dhananjay Maniklal**; 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US).

(74) Agent: **KOYFMAN, Hannah R.** et al.; K&L Gates LLP, One Lincoln Street, State Street Financial Center, Boston, MA 02111 (US).

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

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

(54) Title: HYBRID AAV-ANELLOVECTORS

(57) Abstract: This invention relates generally to compositions for making and administering anellovectors and uses thereof.

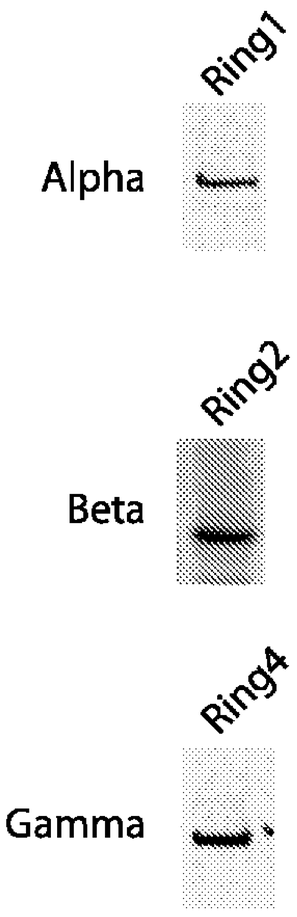


FIG. 1

[Continued on next page]

WO 2022/170195 A1

Declarations under Rule 4.17:

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

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

Published:

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

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

HYBRID AAV-ANELLOVECTORS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 63/147,102, filed February 8, 2021. The contents of the aforementioned application are hereby incorporated by reference in their entirety.

BACKGROUND

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

SUMMARY

The present disclosure provides 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 cell in a human tissue). Generally, the anellovector comprises a proteinaceous exterior comprising an *Anellovirus* ORF1 molecule (e.g., a capsid protein having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* ORF1 protein, e.g., as described herein) and a genetic element enclosed within the proteinaceous exterior, wherein the genetic element comprises at least one nucleic acid sequence (e.g., a contiguous nucleic acid sequence with a length of at least 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, or 4000 nucleotides) from a virus other than an *Anellovirus*, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the nucleic acid sequence from a virus other an *Anellovirus* is from an adeno-associated virus (AAV) (e.g., as described herein). In some embodiments, the effector (e.g., the payload), or a sequence encoding the effector, is separate from the non-*Anellovirus* sequence. In some embodiments, the proteinaceous exterior is capable of introducing the genetic element into a target cell (e.g., a mammalian cell, e.g., a human cell). The disclosure further provides compositions and methods for adminstering an anellovector (e.g., a synthetic anellovector), e.g., as described herein, 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).

An anellovector and components thereof that can be used in the methods for delivering an effector described herein (e.g., produced using a composition or method as described herein) generally

comprise 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 molecule, e.g., an Anellovirus ORF1 protein or a polypeptide encoded by an *Anellovirus* ORF1 nucleic acid, e.g., as described herein, or a polypeptide having at last 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto), which is capable of introducing the genetic element into a cell (e.g., a mammalian cell, e.g., a human cell). The genetic element generally comprises at least one nucleic acid sequence (e.g., a contiguous nucleic acid sequence with a length of at least 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, or 4000 nucleotides) from a virus other than an *Anellovirus* (e.g., from an AAV, e.g., AAV1, AAV2, or AAV5), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the non-Anellovirus sequence comprises a non-Anellovirus origin of replication, e.g., derived from a Monodnavirus, e.g., a Shotokuvirus (e.g., a Cressdnaviricota [e.g., a redondovirus, circovirus {e.g., a porcine circovirus, e.g., PCV-1 or PCV-2; or beak-and-feather disease virus}, geminivirus {e.g., tomato golden mosaic virus}, or nanovirus {e.g., BBTV, MDV1, SCSVF, or FBNYV}]), or a Parvovirus (e.g., a dependoparavirus, e.g., a bocavirus or an adeno-associated virus (AAV)). In some embodiments, the non-Anellovirus origin of replication is derived from an AAV (e.g., AAV1, AAV2, or AAV5). In some embodiments, the non-Anellovirus origin of replication comprises an AAV Rep-binding motif (RBM), e.g., as described herein, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the non-Anellovirus origin of replication comprises an AAV terminal resolution site (TRS), e.g., as described herein, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the non-Anellovirus origin of replication is comprised in an inverted terminal repeat (ITR), e.g., an AAV ITR, e.g., as described herein.

In some embodiments, the anellovector is an infectious vehicle or particle comprising a proteinaceous exterior (e.g., a capsid) 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, or a polypeptide having at last 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto). In embodiments, an anellovector described herein comprises a polypeptide encoded by an *Anellovirus* ORF1 nucleic acid, e.g., having a sequence as described in any of Tables A1, B1, B3, C1, E1, F1, F3, or F5, or a sequence having at last

80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, an anellovector described herein comprises a polypeptide having the sequence of an ORF1 protein, e.g., having a sequence as described in any of Tables A2, B2, B4, C2, E2, F2, F4, or F6, or a polypeptide having at last 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, an anellovector described herein is an infectious vehicle or particle, e.g., comprising an Anellovirus capsid encapsulating a non-Anellovirus genome. Production of an Anellovirus capsid may include in vitro production or host cell expression of an Anellovirus ORF1 molecule, e.g., as described herein.

In some embodiments, the genetic element of an anellovector of the present disclosure is a circular and/or single-stranded DNA molecule (e.g., circular and single stranded). In some embodiments, the genetic element of an anellovector of the present disclosure is a linear and/or single-stranded DNA molecule (e.g., linear and single stranded). In some embodiments, the genetic element includes 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 composition or method, as described herein.

In some instances, the anellovectors that can be used in the methods of delivering an effector described herein comprise a genetic element which comprises or encodes an effector (e.g., a nucleic acid effector, such as a non-coding RNA, or a polypeptide effector, e.g., a protein), e.g., which can be expressed in the 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., 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. In another example, the anellovector can deliver and express an effector, e.g., an exogenous protein, *in vivo*. 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; to modulate a biological response, e.g., cell or molecular response; or for treatment of

conditions such as diseases and disorders, e.g., by delivering an effector that can operate as a modulating and/or therapeutic agent to a desired cell, tissue, or subject.

In some embodiments, the compositions and methods described herein can be used to produce the genetic element of a synthetic anellovector to be used in the methods of aministering anellovectors described herein, 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., a described herein), e.g., a deletion, insertion, substitution, modification (e.g., enzymatic modification), relative to the wild-type virus. In some embodiments, the structural difference comprises the non-Anellovirus sequence of the genetic element, e.g., as described herein. 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 inflammarory 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.

In some embodiments, the compositions and methods described herein can be used to produce the genetic element of an anellovector, e.g. an anellovector that can be used in the methods of delivering an effector described herein, comprising: (i) a genetic element comprising a promoter element and a sequence encoding an effector (e.g., an endogenous or exogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence, e.g., a packaging signal); and (ii) a proteinaceous exterior; wherein the genetic element is enclosed within the proteinaceous exterior (e.g., a capsid); and wherein the anellovector is capable of delivering the genetic element into a eukaryotic (e.g., mammalian, e.g., human) cell. In some embodiments, the genetic element is a single-stranded and/or circular DNA. Alternatively or in combination, the genetic element has one, two, three, or all of the following properties: is circular, is single-stranded, it integrates into the genome of a cell at a frequency of less than about 0.0001%, 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/or it integrates into the genome of a target cell at less than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 copies per genome. 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). In some embodiments, the genetic element is enclosed within the proteinaceous exterior. In some embodiments, the anellovector is capable of delivering the genetic element into a eukaryotic cell. In some embodiments, the genetic element comprises a nucleic acid sequence (e.g., a nucleic acid sequence of between 300-4000

nucleotides, e.g., between 300-3500 nucleotides, between 300-3000 nucleotides, between 300-2500 nucleotides, between 300- 2000 nucleotides, between 300-1500 nucleotides) having 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 sequence of a wild-type Anellovirus (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). In some embodiments, the genetic element comprises a nucleic acid sequence (e.g., a nucleic acid sequence of at least 300 nucleotides, 500 nucleotides, 1000 nucleotides, 1500 nucleotides, 2000 nucleotides, 2500 nucleotides, 3000 nucleotides or more) having 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 sequence of a wild-type Anellovirus (e.g., a wild-type Anellovirus sequence as described herein). 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 compositions and methods described herein can be used to produce the genetic element of an infectious (e.g., to a human cell) Annellovector, 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 that can be used in the methods of adminstering an anellovector described herein. In embodiments, the Anellovector is capable of delivering the genetic element into a mammalian, e.g., human, cell. In some embodiments, the genetic element has less than about 6% (e.g., less than 10%, 9.5%, 9%, 8%, 7%, 6%, 5.5%, 5%, 4.5%, 4%, 3.5%, 3%, 2.5%, 2%, 1.5%, or less) identity to a wild type Anellovirus genome sequence. In some embodiments, the genetic element has no more than 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, 5.5% or 6% identity to a wild type Anellovirus genome sequence. In some embodiments, the genetic element has at least about 2% to at least about 5.5% (e.g., 2 to 5%, 3% to 5%, 4% to 5%) identity to a wild type Anellovirus. In some embodiments, the genetic element has greater than about 2000, 3000, 4000, 4500, or 5000 nucleotides of non-viral sequence (e.g., non Anellovirus genome sequence). In some embodiments, the genetic element has greater than about 2000 to 5000, 2500 to 4500, 3000 to 4500, 2500 to 4500, 3500, or 4000, 4500 (e.g., between about 3000 to 4500) nucleotides of non-viral sequence (e.g., non Anellovirus genome sequence). In some embodiments, the genetic element is a single-stranded, circular DNA. Alternatively or in combination, the genetic element has one, two or 3 of the following properties: is circular, is single stranded, it integrates into the genome of a 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, it integrates into the genome of a target cell at less than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 copies per genome or

integrates at a frequency of less than about 0.0001%, 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 (e.g., by comparing integration frequency into genomic DNA relative to genetic element sequences from cell lysates). 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).

In some embodiments, Anelloviruses or anellovectors, administered according to the methodsdescribed 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.

In some embodiments, the compositions and methods described herein can be used to produce the genetic element of an anellovector that can be used in the methods of administration described herein, comprising a proteinaceous exterior comprising a polypeptide (e.g., a synthetic polypeptide, e.g., an ORF1 molecule) comprising (e.g., in series):

(i) a first region comprising an arginine-rich region, e.g., 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),

(ii) a second region comprising a jelly-roll domain, e.g., a sequence comprising at least 6 beta strands,

(iii) a third region comprising an N22 domain sequence described herein,

(iv) a fourth region comprising an Anellovirus ORF1 C-terminal domain (CTD) sequence described herein, and

(v) optionally wherein the polypeptide has an amino acid sequence having less than 100%, 99%, 98%, 95%, 90%, 85%, 80% sequence identity to a wild type Anellovirus ORF1 protein, e.g., as described herein.

In an aspect, the invention features an isolated nucleic acid molecule (e.g., a nucleic acid construct) comprising the sequence of a genetic element comprising a promoter element operably linked to a sequence encoding an effector, e.g., a payload, and an exterior protein binding sequence. In some embodiments, the exterior protein binding sequence includes a sequence at least 75% (at least 80%, 85%, 90%, 95%, 97%, 100%) identical to a 5'UTR sequence of an Anellovirus, e.g., as disclosed herein. In embodiments, the genetic element is a single-stranded DNA, is circular, 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 the cell, and/or integrates into the genome of a target cell at less than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 copies per genome or 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 the cell. 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). In embodiments, the effector does not originate from TTV and is not an SV40-miR-S1. In embodiments, the nucleic acid molecule does not comprise the polynucleotide sequence of TTMV-LY2. In embodiments, the promoter element is capable of directing expression of the effector in a eukaryotic (e.g., mammalian, e.g., human) cell.

In some embodiments, the nucleic acid molecule is circular. In some embodiments, the nucleic acid molecule is linear. In some embodiments, a nucleic acid molecule described herein comprises one or more modified nucleotides (e.g., a base modification, sugar modification, or backbone modification).

In some embodiments, the nucleic acid molecule comprises a sequence encoding an ORF1 molecule (e.g., an *Anellovirus* ORF1 protein, e.g., as described herein). In some embodiments, the nucleic acid molecule comprises a sequence encoding an ORF2 molecule (e.g., an *Anellovirus* ORF2 protein, e.g., as described herein). In some embodiments, the nucleic acid molecule comprises a sequence encoding an ORF3 molecule (e.g., an *Anellovirus* ORF3 protein, e.g., as described herein). In an aspect, the invention features a genetic element comprising one, two, or three of: (i) a promoter element and a sequence encoding an effector, e.g., an exogenous or endogenous effector; (ii) at least 72 contiguous nucleotides (e.g., at least 72, 73, 74, 75, 76, 77, 78, 79, 80, 90, 100, or 150 nucleotides) having 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; or at least 100 (e.g., at least 300, 500, 1000, 1500) contiguous nucleotides having at least 72% (e.g., at least 72, 73, 74, 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; and (iii) a protein binding sequence, e.g., an exterior protein binding sequence, and wherein the nucleic acid construct is a single-stranded DNA; and wherein the nucleic acid construct is circular, 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 the cell, and/or integrates into the genome of a target cell at less than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 copies per genome In some embodiments, a genetic element encoding an effector (e.g., an exogenous or endogenous effector, e.g., as described herein) is codon optimized. In some embodiments, the genetic element is circular. In some embodiments, the genetic element is linear. In some embodiments, a genetic element described herein comprises one or more modified nucleotides (e.g., a base modification, sugar modification, or backbone modification). In some embodiments, the genetic element comprises a sequence encoding an ORF1 molecule (e.g., an *Anellovirus* ORF1 protein, e.g., as described herein). In some embodiments, the genetic element comprises a sequence encoding an ORF2 molecule (e.g., an *Anellovirus* ORF2 protein, e.g., as described herein). In some embodiments, the

genetic element comprises a sequence encoding an ORF3 molecule (e.g., an *Anellovirus* ORF3 protein, e.g., as described herein).

In an aspect, the invention features a host cell comprising: (a) one or more nucleic acid molecules comprising a sequence encoding one or more of an ORF1 molecule, an ORF2 molecule, or an ORF3 molecule (e.g, a sequence encoding an Anellovirus ORF1 polypeptide described herein), e.g., wherein the nucleic acid molecule is a plasmid, is a viral nucleic acid, or is integrated into a chromosome; and (b) a genetic element, wherein the genetic element comprises (i) a promoter element operably linked to a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector) and (ii) a protein binding sequence that binds the ORF1 molecule of (a), wherein the genetic element of (b) does not encode one or more of an ORF1 polypeptide (e.g., an ORF1 protein), an ORF2 polypeptide (e.g., an ORF2 protein), and/or an ORF3 polypeptide (e.g., an ORF3 protein). For example, the host cell comprises (a) and (b) either in *cis* (both part of the same nucleic acid molecule) or in *trans* (each part of a different nucleic acid molecule). In embodiments, the one or more nucleic acid of (a) may be circular, single-stranded DNA; in other embodiments, the one or more nucleic acid of (a) may be linear DNA. In embodiments, the genetic element of (b) is a circular, single-stranded DNA. In some embodiments, the host cell is a manufacturing cell line, e.g., as described herein. In some embodiments, the host cell is adherent or in suspension, or both. In some embodiments, the host cell or helper cell is grown in a microcarrier. In some embodiments, the host cell or helper cell is compatible with cGMP manufacturing practices. In some embodiments, the host cell or helper cell is grown in a medium suitable for promoting cell growth. In certain embodiments, once the host cell or helper cell has grown sufficiently (e.g., to an appropriate cell density), the medium may be exchanged with a medium suitable for production of anellovectors by the host cell or helper cell.

In an aspect, the invention features a pharmaceutical composition comprising an anellovector (e.g., a synthetic anellovector), e.g., an anellovector that can be administered by the methods described herein. In 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^{13}, 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).

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.

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 embodiments, the payload is a nucleic acid. In embodiments, the payload is a polypeptide.

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.

In an aspect, the invention features a method of making an anellovector, e.g., a synthetic anellovector that can be used in a method of administering an anellovector described herein. The method includes:

(a) providing a host cell comprising:

(i) a first nucleic acid molecule comprising the nucleic acid sequence of a genetic element of an anellovector, e.g., as described herein; and

(ii) a second nucleic acid molecule encoding an Anellovirus ORF1 polypeptide, or one or more of an amino acid sequence chosen from ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, e.g., as described herein, or an amino acid sequence having at least 70% (e.g., at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) sequence identity thereto; and

(b) incubating the host cell under conditions suitable for replication (e.g., rolling circle replication) of the nucleic acid sequence of the genetic element, thereby producing a genetic element; and

optionally (c) incubating the host cell under conditions suitable for enclosure of the genetic element in a proteinaceous exterior (e.g., comprising a polypeptide encoded by the second nucleic acid molecule).

In another aspect, the invention features a method of manufacturing an anellovector composition, e.g., an anellovector composition that can be used in the methods of administration described herein, the composition comprising one or more of (e.g., all of) (a), (b), and (c):

a) providing a host cell comprising, e.g., expressing one or more components (e.g., all of the components) of an anellovector, e.g., a synthetic anellovector, e.g., as described herein;

b) culturing the host cell under conditions suitable for producing a preparation of anellovectors from the host cell, wherein the anellovectors of the preparation comprise a proteinaceous exterior (e.g., comprising an Anellovector ORF1 polypeptide) encapsulating the genetic element (e.g., as described herein), thereby making a preparation of anellovectors; and

optionally, c) formulating the preparation of anellovectors, e.g., as a pharmaceutical composition suitable for administration to a subject.

For example, the host cell provided in this method of manufacturing comprises (a) a nucleic acid comprising a sequence encoding an Anellovirus ORF1 polypeptide described herein, wherein the nucleic acid is a plasmid, is a viral nucleic acid or genome, or is integrated into a helper cell chromosome; and (b) a nucleic acid construct capable of producing a genetic element (e.g., comprising a genetic element sequence and/or genetic element region, e.g., as described herein), e.g., wherein the genetic element comprises (i) a promoter element operably linked to a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector) and (i) a protein binding sequence (e.g, packaging sequence) that binds the polypeptide of (a), wherein the host cell comprises (a) and (b) either in cis or in trans. In embodiments, the genetic element of (b) is circular, single-stranded DNA. In some embodiments, the host cell is a manufacturing cell line.

In some embodiments, the components of the anellovector are introduced into the host cell at the time of production (e.g., by transient transfection). In some embodiments, the host cell stably expresses the components of the anellovector (e.g., wherein one or more nucleic acids encoding the components of the anellovector are introduced into the host cell, or a progenitor thereof, e.g., by stable transfection).

In an aspect, the invention features a method of manufacturing an anellovector composition, comprising: a) providing a plurality of anellovectors described herein, or a preparation of anellovectors described herein; and b) formulating the anellovectors or preparation thereof, e.g., as a pharmaceutical composition suitable for administration to a subject.

In an aspect, the invention features a method of making a host cell, e.g., a first host cell or a producer cell (e.g., as shown in Figure 12 of PCT/US19/65995), e.g., a population of first host cells, comprising an anellovector, the method comprising introducing a nucleic acid construct capable of producing a genetic element, e.g., as described herein, to a host cell and culturing the host cell under conditions suitable for production of the anellovector. In embodiments, the method further comprises introducing a helper, e.g., a helper virus, to the host cell. In embodiments, the introducing comprises transfection (e.g., chemical transfection) or electroporation of the host cell with the anellovector.

In an aspect, the invention features a method of making an anellovector, comprising providing a host cell, e.g., a first host cell or producer cell (e.g., as shown in Figure 12 of PCT/US19/65995), comprising an anellovector, e.g., as described herein, and purifying the anellovector from the host cell. In

some embodiments, the method further comprises, prior to the providing step, contacting the host cell with a nucleic acid construct or an anellovector, e.g., as described herein, and incubating the host cell under conditions suitable for production of the anellovector. In embodiments, the host cell is the first host cell or producer cell described in the above method of making a host cell. In embodiments, purifying the anellovector from the host cell comprises lysing the host cell.

In some embodiments, the method further comprises a second step of contacting the anellovector produced by the first host cell or producer cell with a second host cell, e.g., a permissive cell (e.g., as shown in Figure 12 of PCT/US19/65995), e.g., a population of second host cells. In some embodiments, the method further comprises incubating the second host cell inder conditions suitable for production of the anellovector. In some embodiments, the method further comprises purifying an anellovector from the second host cell, e.g., thereby producing an anellovector seed population. In embodiments, at least about 2-100-fold more of the anellovector is produced from the population of second host cells than from the population of first host cells. In embodiments, purifying the anellovector from the second host cell comprises lysing the second host cell. In some embodiments, the method further comprises a second step of contacting the anellovector produced by the second host cell with a third host cell, e.g., permissive cells (e.g., as shown in Figure 12 of PCT/US19/65995), e.g., a population of third host cells. In some embodiments, the method further comprises incubating the third host cell inder conditions suitable for production of the anellovector. In some embodiments, the method further comprises purifying a anellovector from the third host cell, e.g., thereby producing an anellovector stock population. In embodiments, purifying the anellovector from the third host cell comprises lysing the third host cell. In embodiments, at least about 2-100-fold more of the anellovector is produced from the population of third host cells than from the population of second host cells.

In some embodiments, the host cell is grown in a medium suitable for promoting cell growth. In certain embodiments, once the host cell has grown sufficiently (e.g., to an appropriate cell density), the medium may be exchanged with a medium suitable for production of anellovectors by the host cell. In some embodiments, anellovectors produced by a host cell separated from the host cell (e.g., by lysing the host cell) prior to contact with a second host cell. In some embodiments, anellovectors produced by a host cell are contacted with a second host cell without an intervening purification step.

In an aspect, the invention features a method of making a pharmaceutical anellovector preparation, e.g., a preparation to be used in the methods of administration described herein. The method comprises (a) making an anellovector preparation as described herein, (b) evaluating the preparation (e.g., a pharmaceutical anellovector preparation, anellovector seed population or the anellovector stock population) for one or more pharmaceutical quality control parameters, e.g., identity, purity, titer, potency (e.g., in genomic equivalents per anellovector particle), and/or the nucleic acid sequence, e.g., from the

genetic element comprised by the anellovector, and (c) formulating the preparation for pharmaceutical use of the evaluation meets a predetermined criterion, e.g, meets a pharmaceutical specification. In some embodiments, evaluating identity comprises evaluating (e.g., confirming) the sequence of the genetic element of the anellovector, e.g., the sequence encoding the effector. In some embodiments, evaluating purity comprises evaluating the amount of an impurity, e.g., 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 aggregates. In some embodiments, evalating titer comprises evaluating the ratio of functional versus non-functional (e.g., infectious vs non-infectious) anellovectors in the preparation (e.g., as evaluated by HPLC). In some embodiments, evaluating potency comprises evaluating the level of anellovector function (e.g., expression and/or function of an effector encoded therein or genomic equivalents) detectable in the preparation.

In embodiments, the formulated preparation is substantially free of pathogens, host cell contaminants or impurities; 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). In some embodiments, multiple anellovectors can be produced in a single batch. In embodiments, the levels of the anellovectors produced in the batch can be evaluated (e.g., individually or together).

In an aspect, the invention features a host cell comprising:

(i) a first nucleic acid molecule comprising a nucleic acid construct as described herein, and

(ii) optionally, a second nucleic acid molecule encoding one or more of an amino acid sequence chosen from ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, e.g., as described herein, or an amino acid sequence having at least about 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity thereto.

In an aspect, the invention features a reaction mixture comprising an anellovector described herein and a helper virus that can be used in the methods of admintration described herein, wherein the helper virus comprises a polynucleotide encoding an exterior protein, (e.g., an exterior protein capable of binding to the exterior protein binding sequence and, optionally, a lipid envelope), a polynucleotide encoding a replication protein (e.g., a polymerase), or any combination thereof.

In some embodiments, an anellovector (e.g., a synthetic anellovector) is isolated, e.g., isolated from a host cell and/or isolated from other constituents in a solution (e.g., a supernatant). In some embodiments, an anellovector (e.g., a synthetic anellovector) is purified, e.g., from a solution (e.g., a supernatant). In some embodiments, an anellovector is enriched in a solution relative to other constituents in the solution.

In some embodiments of any of the aforesaid anellovectors, compositions or methods, providing an anellovector comprises separating (e.g., harvesting) an anellovector from a composition comprising an anellovector-producing cell, e.g., as described herein. In other embodiments, providing an anellovector comprises obtaining an anellovector or a preparation thereof, e.g., from a third party.

In embodiments, the genetic element is not capable of self-replication and/or self-amplification. In embodiments, the genetic element is capable of replicating and/or being amplified in *trans,* e.g., in the presence of a helper, e.g., a helper virus.

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

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

1. A viral particle comprising a circular DNA comprising (i) an AAV origin of replication, (ii) a promoter operably linked to a sequence encoding a therapeutic RNA or polypeptide, and (iii) a sequence that binds an Anellovirus ORF1 molecule, the circular DNA being encapsidated by a capsid comprising an Anellovirus ORF1 molecule.

2. A viral particle comprising a circular DNA comprising (i) an AAV origin of replication, and (ii) a promoter operably linked to a sequence encoding a therapeutic RNA or polypeptide, wherein the circular DNA is encapsidated by a capsid comprising an Anellovirus ORF1 molecule.

3. A vector comprising:

a) a proteinaceous exterior comprising an Anellovirus ORF1 molecule; and

b) a genetic element comprising a non-Anellovirus origin of replication;

optionally wherein the genetic element further comprises: (i) a nucleic acid sequence encoding an exogenous effector, and/or (ii) a promoter element operatively linked to the nucleic acid sequence encoding the exogenous effector.

4. The vector of embodiment 3, wherein the non-Anellovirus origin of replication is derived from a DNA virus, e.g., a single-stranded DNA (ssDNA) virus, e.g., a linear ssDNA virus.

5. The vector of embodiment 3 or 4, wherein the non-Anellovirus origin of replication is derived from a Monodnavirus, e.g., a Shotokuvirus (e.g., a Cressdnaviricota [e.g., a redondovirus, circovirus {e.g., a porcine circovirus, e.g., PCV-1 or PCV-2; or beak-and-feather disease virus}, geminivirus {e.g., tomato golden mosaic virus}, or nanovirus {e.g., BBTV, MDV1, SCSVF, or FBNYV}]), or a Parvovirus (e.g., a dependoparavirus, e.g., a bocavirus or an AAV).

6. The vector of embodiment 5, wherein the non-Anellovirus origin of replication is derived from a Monodnavirus, e.g., Shotokuvirus, e.g., Cossaviricota, e.g., Quintoviricetes, e.g., Piccovirales, e.g., Parvoviridae, e.g., Parvovirinae, e.g., Dependoparvovirus, e.g., an Adeno-associated virus (AAV).

7. The vector of embodiment 5, wherein the non-Anellovirus origin of replication is an AAV (e.g., AAV1, AAV2, or AAV5) origin of replication.

8. The vector of embodiment 5, wherein the non-Anellovirus origin of replication is derived from a virus that replicates by rolling circle replication.

9. The vector of embodiment 5, wherein the non-Anellovirus origin of replication is derived from a virus that replicates by rolling hairpin replication.

10. The vector of embodiment 5, wherein the non-Anellovirus origin of replication is derived from a virus that infects an animal (e.g., a mammal, e.g., a human), plant, fungi, or bacteria.

11. The vector of any of the preceding embodiments, wherein the non-Anellovirus origin of replication comprises an AAV Rep-binding motif (RBM), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

12. The vector of any of the preceding embodiments, wherein the non-Anellovirus origin of replication comprises an AAV terminal resolution site (TRS), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

13. The vector of any of the preceding embodiments, wherein the non-Anellovirus origin of replication comprises an inverted terminal repeat (ITR).

14. The vector of any of the preceding embodiments, wherein the non-anellovirus origin of replication does not comprise an Anellovirus origin of replication, or a nucleic acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

15. The vector of any of the preceding embodiments, wherein the non-Anellovirus origin of replication does not substantially replicate (e.g., is incapable of replicating) by rolling circle replication.

16. The vector of any of the preceding embodiments, wherein the non-Anellovirus orign of replication does not comprise a contiguous sequence of at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides from an *Anellovirus* genome (e.g., as described herein).

17. A genetic element comprising:

a protein binding sequence that specifically binds an Anellovirus ORF1 molecule (e.g., a 5' UTR); and

an AAV origin of replication, e.g., comprised in a first AAV inverted terminal repeat (ITR); optionally, a nucleic acid sequence encoding an exogenous effector (e.g., a therapeutic exogenous effector); and

optionally, a promoter element operatively linked to the nucleic acid sequence encoding the exogenous effector.

18. A genetic element construct comprising:

a protein binding sequence that specifically binds an Anellovirus ORF1 molecule (e.g., a 5' UTR); and

an AAV origin of replication, e.g., comprised in a first AAV inverted terminal repeat (ITR); optionally, a nucleic acid sequence encoding an exogenous effector (e.g., a therapeutic exogenous effector); and

optionally, a promoter element operatively linked to the nucleic acid sequence encoding the exogenous effector.

19. A system comprising:

a) a first nucleic acid, wherein the first nucleic acid is a genetic element or a genetic element construct, the first nucleic acid comprising:

an AAV origin of replication, e.g., comprised in a first AAV inverted terminal repeat (ITR);

optionally, a nucleic acid sequence encoding an exogenous effector (e.g., a therapeutic exogenous effector); and

optionally, a promoter element operatively linked to the nucleic acid sequence encoding the exogenous effector;

b) a second nucleic acid encoding an Anellovirus ORF1 molecule.

20. The system of embodiment 19, wherein the first nucleic acid further comprises a protein binding sequence that specifically binds an Anellovirus ORF1 molecule (e.g., a 5' UTR or GC-rich region of an *Anellovirus*).

21. The system of embodiment 19 or 20, which further comprises a nucleic acid sequence encoding an Anellovirus ORF2 molecule.

22. The system of embodiment 21, wherein the nucleic acid sequence encoding the Anellovirus ORF2 molecule is situated on a third nucleic acid.

23. The system of any of embodiments 19-22, which further comprises a nucleic acid sequence encoding an AAV Rep2 molecule (e.g., an AAV Rep2 polypeptide, e.g., AAV Rep2 protein).

24. The system of embodiment 23, wherein the nucleic acid sequence encoding the AAV REP2 molecule is situated on a fourth nucleic acid.

25. The system of any of embodiments 19-24, which further comprises one or more nucleic acid sequence encoding one or more of (e.g., all of) an Adenovirus E2A molecule, an Adenovirus E4 molecule, and an Adenovirus VARNA molecule.

26. The system of embodiment 25, wherein the nucleic acid sequence encoding the Adenovirus E2A molecule, the Adenovirus E4 molecule, and the Adenovirus VARNA molecule is situated on a fifth nucleic acid.

27. The system of any of embodiments 19-26, wherein one or more of (e.g., all of) the first, second, third, fourth, and fifth nucleic acids are plasmids.

28. The system of any of embodiments 19-27, wherein the nucleic acids are admixed or in separate volumes.

29. The system of any of embodiments 19-28, wherein the nucleic acids are in a cell, e.g., a human cell, e.g., a 293 cell or a MOLT4 cell.

30. A DNase-protected proteinaceous complex comprising:

a) a proteinaceous exterior comprising an Anellovirus ORF1 molecule; and

b) a genetic element comprising an AAV origin of replication, e.g., comprised in a first AAV inverted terminal repeat (ITR);

optionally wherein the genetic element further comprises: (i) a nucleic acid sequence encoding an exogenous effector, and/or (ii) a promoter element operatively linked to the nucleic acid sequence encoding the exogenous effector.

31. The DNase-protected proteinaceous complex of embodiment 30, wherein:

the genetic element is substantially free of Anellovirus sequence,

the genetic element does not comprise more than 100 nucleotides of more than 50% identity to any 100 nucleotide sequence of a wild-type Anellovirus genome, or

the genetic element does not comprise an Anellovirus 5' UTR.

32. A DNase-protected proteinaceous complex comprising:

a) a proteinaceous exterior comprising an Anellovirus ORF1 molecule; and

b) a genetic element;

wherein:

the genetic element is substantially free of Anellovirus sequence,

the genetic element does not comprise more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 consecutive nucleotides of more than 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to any sequence of the same length of a wild-type Anellovirus genome, and/or

the genetic element does not comprise an Anellovirus 5' UTR;

optionally wherein the genetic element further comprises: (i) a nucleic acid sequence encoding an exogenous effector, and/or (ii) a promoter element operatively linked to the nucleic acid sequence encoding the exogenous effector.

33. The DNase-protected proteinaceous complex of embodiment 32, wherein the genetic element further comprises (iii) a first ITR, e.g., a first AAV ITR.

34. A mixture comprising:
an Anellovirus ORF1 molecule, and
a nucleic acid comprising an AAV origin of replication, e.g., comprised in a first AAV inverted terminal repeat (ITR).

35. A mixture comprising:
an Anellovirus ORF1 molecule, and
a nucleic acid (e.g., a genetic element);
wherein:
the nucleic acid is substantially free of Anellovirus sequence,
the nucleic acid does not comprise more than 100 nucleotides of more than 50% identity to any 100 nucleotide sequence of a wild-type Anellovirus genome, or
the nucleic acid does not comprise an Anellovirus 5' UTR;

36. The mixture of embodiment 34 or 35, wherein the Anellovirus ORF1 molecule is bound to the nucleic acid comprising the first AAV ITR.

37. The mixture of any of embodiments 34-36, wherein the nucleic acid comprising the first AAV origin of replication is a genetic element, e.g., a genetic element according to any of the preceding embodiments.

38. A complex comprising:
genetic element according to any of the preceding embodiments, and
a capsid protein (e.g., an ORF1 molecule) bound to the genetic element.

39. The mixture or complex of any of embodiments 34-38, which is in a cell-free system or a substantially cell-free composition.

40. The complex of embodiment 38 or 39, wherein the complex is in a cell, e.g., a host cell, e.g., a helper cell.

41. A cell comprising the genetic element or genetic element construct of any of the preceding embodiments.

42. The cell of embodiment 41, which is a human cell, e.g., a 293 cell, an Expi293 cell, an Expi293F cell, or a MOLT-4 cell.

43. A method of delivering an exogenous effector to a target cell (e.g., a vertebrate cell, e.g., a mammalian cell, e.g., a human cell), the method comprising introducing into the cell a vector of any of the preceding embodiments.

44. A method of modulating a biological activity in a subject in need thereof, the method comprising introducing into the subject a vector of any of the preceding embodiments.

45. A method of treating or preventing a disease or disorder in a subject in need thereof, the method comprising introducing into the subject a vector of any of the preceding embodiments.

46. A method of vaccinating a subject in need thereof, the method comprising introducing into the subject a vector of any of the preceding embodiments, wherein the exogenous effector comprises an antigen from an infectious agent (e.g., a virus or bacteria).

47. The method of any of embodiments 43-46, wherein the target cell is a human cell, e.g., a 293 cell, an Expi293 cell, an Expi293F cell, or a MOLT-4 cell.

48. The method of any of embodiments 43-46, wherein the target cell is a cell from an animal (e.g., an agricultural animal, e.g., a cow, sheep, pig, goat, horse, bison, or camel).

49. The method of embodiment 48, wherein the animal is an avian animal (e.g., a turkey, chicken, quail, emu, or ostrich).

50. The method of any of embodiments 43-49, wherein the target cell is *in vivo* or *in vitro*.

51. The method of any of embodiments 43-50, wherein the vector is contacted to a cell *in vitro, ex vivo,* or *in vivo*.

52. The vector of any of the preceding embodiments, wherein the genetic element is substantially protected from digestion with DNAse I.

53. The vector of any of the preceding embodiments, wherein if the exogenous effector is replaced with mKate, the vector can deliver mKate to a plurality of target cells (e.g., MOLT4 cells) in vitro, resulting in at least about 10%, 20%, 30%, 40%, 50%, or 60% of cells contacted with the vector having a fluorescence above a background levels, wherein the background level is the level excluding all but the most fluorescent 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, or 10% of cells contacted with an otherwise similar vector lacking ORF1, e.g., in a flow cytometry assay of Example 5.

54. The vector of any of the preceding embodiments, wherein if the exogenous effector is replaced with nanoLuciferase, the vector can deliver nanoLuciferase to a plurality of target cells (e.g., Vero cells or MCF7 cells) in vitro, resulting in a population of cells contacted with the vector that shows luminescence of at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, or 15 times a background level, wherein the background level is the luminescence of otherwise similar cells not contacted with the vector, e.g., in a luminescence assay of Example 4 or 8.

55. The vector of any of the preceding embodiments, which sediments at a density of about 1.2-1.4 g/ml on a CsCl gradient, e.g., according to Example 5.

56. A method of making a vector, comprising:

(a) providing a host cell comprising a genetic element of any of the preceding embodiments, and

(b) incubating the host cell under conditions suitable for enclosure of the genetic element in a proteinaceous exterior (e.g., a proteinaceous exterior comprising an Anellovirus ORF1 molecule), thereby making the vector.

57. A method of making a vector, comprising:

(a) providing a host cell comprising a system of any of the preceding embodiments, and

(b) incubating the host cell under conditions suitable for enclosure of the genetic element in a proteinaceous exterior (e.g., a proteinaceous exterior comprising an Anellovirus ORF1 molecule), thereby making the vector.

58. The method of embodiment 56 or 57, which comprises lysis of the host cell.

59. The method of any of embodiments 56-58, which comprises obtaining the vector from supernatant of the host cell.

60. The method of any of embodiments 56-59, wherein the host cell further comprises one or more additional nucleic acids encoding one or more of (e.g., all of) an Anellovirus ORF2 molecule, an AAV REP2 molecule, an Adenovirus E2A molecule, an Adenovirus E4 molecule, and an Adenovirus VARNA molecule.

61. A method of making a therapeutic composition, comprising:

(a) providing one or a plurality of host cells comprising exogenous DNA comprising:

(i) an AAV origin of replication,

(ii) a promoter operably linked to a sequence encoding a therapeutic effector (e.g., a therapeutic RNA or polypeptide),

(iii) optionally a sequence encoding an Anellovirus ORF1 molecule,

(iv) optionally a sequence encoding an Anellovirus ORF2 molecule,

(v) optionally a sequence encoding a Rep protein (e.g., an AAV Rep protein, e.g., an AAV Rep2 protein), and

(vi) optionally a sequence encoding one or a plurality of helper proteins, e.g., an Adenovirus helper protein, e.g., an E2A molecule, an Adenovirus E4 molecule, and/or an Adenovirus VARNA molecule;

(b) culturing the one or plurality of host cells under conditions suitable for formation of vectors (e.g., anellovectors, e.g., viral particles) comprising a proteinaceous exterior (e.g., capsid) comprising a sufficient number of the ORF1 molecules to enclose (e.g., encapsidate) a genetic element comprising the promoter operably linked to the sequence encoding the therapeutic effector; optionally wherein the genetic element is circular or linear;

(c) enriching, e.g., purifying the vectors produced in step (b) from the cell culture,

thereby making a therapeutic composition.

62. The method of embodiment 61, further comprising:

(d) evaluating the purified viral particles for one or more impurity selected from: endotoxin, mycoplasma, 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 particles, free viral capsid protein, adventitious agents, and aggregates;

(e) optionally reducing or removing the one or more impurity from the viral particles if detected in step (d); and

(f) optionally formulating the purified viral particles for administration to a human, thereby making a therapeutic composition.

63. The method of embodiment 61 or 62, wherein the exogenous DNA of (a) (i)-(vi) is provided in one host cell.

64. The method of any one of embodiments 61-63, wherein the exogenous DNA of (a) (i)-(vi) is provided in a plurality of host cells.

65. The method of any one of embodiments 61-64, wherein the exogenous DNA of (a) (i) and (ii) is provided in one host cell and the exogenous DNA of (a) (iii)-(vi) is provided in a second host cell.

66. The method of any one of embodiments 61-65, wherein the exogenous DNA of (a)(i)-(ii) is not part of a host cell chromosome.

67. The method of any one of embodiments 61-66, wherein the exogenous DNA of (a)(i)-(ii) is part of the same nucleic acid, e.g., a circular DNA or a linear DNA.

68. The method of any one of embodiments 61-67, wherein the exogenous DNA of (a)(i)-(ii) is a genetic element according to any of the preceding embodiments.

69. The method of any one of embodiments 61-68, wherein one or more of the exogenous DNA of (a)(iii) is integrated into a host cell chromosome.

70. The method of any one of embodiments 61-69, wherein one or more of the exogenous DNA of any of (a)(iv)-(vi), if present, is integrated into a host cell chromosome.

71. The method of any one of embodiments 61-70, wherein one or more of the exogenous DNA of (a)(iii) is part of a plasmid.

72. The method of any one of embodiments 61-71, wherein one or more of the exogenous DNA of any of (a)(iv)-(vi), if present, is part of a plasmid.

73. The method of any one of embodiments 61-72, wherein the host cell is a mammalian cell (e.g., a human cell, e.g., a HEK293 cell).

74. The method of any one of embodiments 61-73, wherein the host cell is an immortalized cell.

75. A method of making a therapeutic composition, comprising:

(a) providing a solution comprising:

(i) a genetic element comprising an AAV origin of replication and a promoter operably linked to a sequence encoding a therapeutic effector (e.g., a therapeutoic RNA or polypeptide), and

(ii) a plurality of ORF1 molecules (e.g., a plurality of copies of the same ORF1 molecule);

(b) incubating the solution under conditions suitable for formation of vectors (e.g., anellovectors, e.g., viral particles) comprising a proteinaceous exterior (e.g., capsid) comprising a sufficient number of the ORF1 molecules to enclose (e.g., encapsidate) the genetic element; and

(c) optionally enriching, e.g., purifying the vectors produced in step (b) from the solution, thereby making a therapeutic composition.

76. The method of embodiment 75, wherein the genetic element was made using …

(iii) optionally a sequence encoding an Anellovirus ORF1 molecule,

(iv) optionally a sequence encoding an Anellovirus ORF2 molecule,

(v) optionally a sequence encoding an AAV REP2 sequence

(vi) optionally a sequence encoding one or a plurality of helper proteins, e.g., an Adenovirus helper protein, e.g., an E2A molecule, an Adenovirus E4 molecule, and/or an Adenovirus VARNA molecule.

77. The method of any one of embodiments 61-76, wherein the vectors produced in step (b) are the vectors of any of the preceding embodiments.

78. A host cell (e.g., a vertebrate cell, e.g., a mammalian cell, e.g., a human cell) comprising a genetic element or genetic element construct of any of the preceding embodiments.

79. The host cell of embodiment 78, which further comprises an Anellovirus ORF1 molecule or a nucleic acid encoding the Anellovirus ORF1 molecule.

80. The host cell of embodiments 78 or 79, which further comprises one or more of (e.g., all of) an Anellovirus ORF2 molecule, an AAV REP2 molecule, an Adenovirus E2A molecule, an Adenovirus E4 molecule, and an Adenovirus VARNA molecule.

81. The host cell of any of embodiments 78-80, which further comprises one or more nucleic acids encoding one or more of (e.g., all of) an Anellovirus ORF2 molecule, an AAV REP2 molecule, an Adenovirus E2A molecule, an Adenovirus E4 molecule, and an Adenovirus VARNA molecule.

82. A host cell comprising a vector of any of the preceding embodiments.

83. A method of making a host cell of any of embodiments 78-82, comprising introducing the genetic element into a cell, e.g., wherein introducing the genetic element comprises introducing a genetic element construct into the cell under conditions that allow for production of the genetic element.

84. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element further comprises a second AAV origin of replication, e.g., comprised in a second AAV inverted terminal repeat (ITR).

85. The genetic element, genetic element construct, system, cell, method, or vector of embodiment 84, wherein the second ITR is oriented inversely to the first ITR.

86. The genetic element, genetic element construct, system, cell, method, or vector of embodiment 84, wherein the second ITR has the same orientation relative to the first ITR.

87. The genetic element, genetic element construct, system, cell, method, or vector of any of embodiments 84-86, wherein the second ITR has the same sequence as the first ITR.

88. The genetic element, genetic element construct, system, cell, method, or vector of any of embodiments 84-86, wherein the second ITR has one or more sequence differences relative to the first ITR.

89. The genetic element, genetic element construct, system, cell, method, or vector of any of embodiments 84-88, wherein the nucleic acid sequence encoding the exogenous effector is situated between the first ITR and the second ITR.

90. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the first AAV ITR comprises the sequence of any of SEQ ID NOs: 1051-1059, or a sequence having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

91. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element is linear.

92. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element is circular.

93. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element construct is circular.

94. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element construct is linear.

95. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element has a length of about 500-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, 4000-4100, 4100-4200, 4200-4300, 4300-4400, 4400-4500, 4500-4600, 4600-4700, 4700-4800, 4800-4900, 4900-5000, 5000-5500, 5500-6000, or 6000-7000 nucleotides.

96. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element has a length of at least 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5100, 5200, 5300, 5400, 5500, or 6000 nucleotides.

97. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element comprises DNA.

98. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element consists of DNA.

99. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element consists at least of 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% DNA.

100. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element is single stranded DNA or double stranded DNA.

101. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element construct is single stranded DNA or double stranded DNA.

102. The genetic element of any of the preceding embodiments, which was produced using a circularized double-stranded DNA, e.g., wherein the circularized DNA was produced by in vitro circularization.

103. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the promoter element is endogenous to an Anellovirus.

104. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the promoter element is endogenous to an AAV.

105. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the promoter element is exogenous to an Anellovirus.

106. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the promoter element is exogenous to an AAV.

107. The genetic element construct of any of the preceding embodiments, which comprises a backbone region suitable for replication of the genetic element construct, e.g., for replication in a bacterial cell.

108. The genetic element construct of any of the preceding embodiments, wherein the backbone region comprises one or both of an origin of replication and a selectable marker.

109. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element further comprises an Anellovirus 5' UTR, an Anellovirus GC-rich region, and Anellovirus 3' UTR, or any combination thereof.

110. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element further comprises an Anellovirus 5' UTR of any of Tables A1, B1, B3, C1, E1, F1, F3, or F5.

111. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element further comprises an Anellovirus GC-rich region of any of Tables A1, B1, B3, C1, E1, F1, F3, or F5.

112. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the genetic element further comprises an Anellovirus 3' UTR of any of Tables A1, B1, B3, C1, E1, F1, F3, or F5.

113. The genetic element, genetic element construct, system, cell, method, or vector of any of the preceding embodiments, wherein the nucleic acid sequence encoding the exogenous effector is about 20-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1,000 nucleotides in length.

114. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the effector comprises a miRNA.

115. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the effector, e.g., miRNA, targets a host gene, e.g., modulates expression of the gene, e.g., increases or decreases expression of the gene.

116. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the effector comprises a miRNA, and decreases expression of a host gene.

117. The genetic element, nucleic acid construct, CAVector, complex, method, or host cell of any of the preceding embodiments, wherein the effector comprises a nucleic acid sequence about 20-200, 30-180, 40-160, 50-140, or 60-120 nucleotides in length.

118. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the nucleic acid sequence encoding the effector is about 20-200, 30-180, 40-160, 50-140, or 60-120 nucleotides in length.

119. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the sequence encoding the effector has a size of at least about 100 nucleotides.

120. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the sequence encoding the effector has a size of about 100 to about 5000 nucleotides.

121. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the sequence encoding the effector has a size of about 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, or 1500-2000 nucleotides.

122. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the genetic element is DNA.

123. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the vector is replication-deficient.

124. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein:

(i) the genetic element is substantially free of Anellovirus sequence,

(ii) the genetic element does not comprise more than 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 consecutive nucleotides of more than 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to any sequence of the same length of a wild-type Anellovirus genome, and/or

(iii) the genetic element does not comprise an Anellovirus 5' UTR;

125. The genetic element, genetic element construct, vector, mixture, complex, method, or host cell of any of the preceding embodiments, wherein the vector is a viral particle.

126. A pharmaceutical composition comprising the vector of any of the preceding embodiments, and a pharmaceutically acceptable carrier and/or excipient.

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

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 incorporated 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

FIG. 1 is a Western blot demonstrating expression of N-terminally 3xFlag-tagged anellovirus ORF1 proteins. Top, Alphatorquevirus Ring1 ORF1 (91 kda). Middle, Betatorquevirus Ring2 ORF1 (79 kda). Bottom, Gammatorquevirus Ring4 ORF1 (82 kda).

FIG. 2 is a series of diagrams demonstrating replication of ITR-flanked payloads by Cap-free AAV-Rep expression constructs. Depicted is a Southern blot probed for hrGFP and pHelper. Lanes 1-3 contain untransfected control DNAs, lanes 4-6 contain total DNA from cells transfected with different Rep constructs. Arrows indicate band positions for pHelper plasmid, pITR-hrGFP plasmid, and replicated ITR-hrGFP DNA.

FIGS. 3A-3B are a series of graphs showing purification of R2 anellovectors encompassing an nLuc transgene from CsCl linear gradients. Vectors were quantified through qPCR against the nLuc reporter gene. (A) Vectors were produced through trans-expression of both AnelloVirus ORF1, ORF2 proteins and particles containing the nLuc transgenes. (B) Quantification of nLuc transgenes when Anellovirus ORF1 and ORF2 were not expressed in trans.

FIG. 4 is a graph showing transduction of non-human primate cells with R2-nLuc anellovectors. Vero cells were seeded at 1e5 cells per well in a 24 well plate. Transductions were performed via the addition of vector at a MOI of 0.4 (based on qPCR titre). 2 days later luciferase assays were performed.

FIG. 5 is a graph showing transduction of human cells with R2-nLuc anellovectors. IGR-OV1 cells were seeded at 1e5 cells per well in a 24 well plate. Transductions were performed via the addition of vector at a MOI of 0.4 (based on qPCR titre). 2 days later luciferase assays were performed.

FIG. 6 is a series of diagrams showing generation of Anellovirus/AAV vectors and successful transduction in MOLT4 cells. The top panel shows an exemplary workflow for producing Anello/AAV hybrid vectors varying an mKate payload in Expi-293 cells and transduction of vectors into MOLT4 cells, followed by flow cytometry analysis for mKate fluorescence. The bottom left panel shows a diagram of an Anello/AAV hybrid vector comprising an ORF1 protein capsid enclosing a genetic element comprising an mKate-encoding gene flanked by inverse terminal repeats (ITRs). The bottom right panel shows the results of flow cytometry analysis of MOLT4 cells transduced with vectors generated using the indicated plasmids.

FIGS. 7A-7B is a series of diagram showing that engineered Ring2 Anellovirus DNA replicates through AAV Rep protein. (A) Diagram showing Ring2 dsDNA genome incorporating a minimal region required for AAV replication, including a Rep binding motif (RBM) and a terminal resolution site (TRS). (B) Southern blots showing linear plasmid and Dpn1 digestion products from DNA samples obtained from Expi-293 cells transfected with indicated combinations of AAV-Rep plasmids and WT Ring2 genome or Ring2 + RBM/TRS DNA (as shown in FIG. 7A).

FIGS. 8A-8B are a series of graphs showing transduction of mammalian cell lines by anellovectors encoding human growth hormone (hGH) as a payload. (A) IGR-OV1 cells were transfected with an AAV Rep vector, a pHelper vector, and one of: (i) Ring2 capsid anellovector encoding hGH, (ii) Ring9 capsid anellovector encoding hGH, encoding hGH, (iii) an AAV2 capsid viral vector encoding hGH (positive control), or (iv) a no-capsid negative control. hGH levels were quantified by ELISA at day 0, day 2, and day 3. (A) Vero cells were transfected with an AAV Rep vector, a pHelper vector, and one of: (i) Ring2 capsid anellovector encoding hGH, (ii) Ring9 capsid anellovector encoding hGH, encoding hGH, (iii) an AAV2 capsid viral vector encoding hGH (positive control), or (iv) a no-capsid negative control. hGH levels were quantified by ELISA at day 0, day 2, and day 3.

FIG. 9 is a graph showing nano-luciferase luminescence in cell lysates from 293F cells transfected with Ring2-AAV ITR-nLuc anellovectors produced either in the presence or absence of AAV Rep (+AAV Rep or -AAV Rep, respectively).

FIGS. 10A-10L are a series of diagrams showing schematics of exemplary genetic element constructs that can be used to produced genetic elements for anellovectors as described herein. The individual schematics correspond to the plasmids indicated in Table 61 below. Black = Ring2 genome sequence (e.g., as described herein); Green = exogenous effector sequence; Blue = AAV origin of replication.

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

DETAILED DESCRIPTION OF THE INVENTION

Definitions

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.

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 "consisting 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.

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.

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.

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.

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.

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.

As used herein, the term "anellovector" refers to a vehicle comprising a genetic element, e.g., a circular DNA, enclosed in a proteinaceous exterior, e.g, the genetic element is substantially protected from digestion with DNAse I 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 DNAse I, 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 Anellovirus viral particle.

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 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.

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).

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.

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.

As used herein, the term "genetic element" refers to a nucleic acid molecule that is or can be enclosed within (e.g., protected from DNAse I 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 DNA and optionally further assembled into a proteinaceous 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.

As used herein, "genetic element construct" refers to a nucleic acid construct (e.g., a plasmid, bacmid, cosmid, or minicircle) comprising at least one (e.g., two) genetic element sequence(s), or fragment thereof. In some embodiments, a genetic element construct comprises at least one full length genetic element sequence. In some embodiments, a genetic element comprises a full length genetic element sequence and a partial genetic element sequence. In some embodiments, a genetic element comprises two or more partial genetic element sequences (e.g., in 5' to 3' order, a 5'-truncated genetic element sequence arranged in tandem with a 3'-truncated genetic element sequence, e.g., as shown in FIG. 27C).

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 genetic element region can undergo rolling circle replication. In some embodiments, the genetic element comprises a Rep protein binding site. In some embodiments, the genetic element comprises a Rep protein displacement site. 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.

As used herein, the term "inverted terminal repeat" ("ITR") refers to a nucleic acid sequence comprising an origin of replication suitable for replication of the surrounding nucleic acid sequence (or a portion thereof) by a viral Rep molecule (e.g., a non-Anellovirus Rep molecule, e.g., an AAV Rep protein), or a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. Generally, an ITR (or the viral sequence from which an ITR is derived)

comprises a contiguous sequence of nucleotides followed (e.g., directly adjacent to, or separated by about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides) by its reverse complement. A copy of an ITR may, in some instances, be comprised at one or both terminal ends of the genome of a single-stranded viral genome (e.g., the genome of a non-Anellovirus, e.g., as described herein, e.g., an AAV). An ITR sequence may be capable of forming a hairpin. An ITR may comprise a Rep-binding motif (RBM) and/or a terminal resolution site (TRS), e.g., as described herein. In some instances, an ITR sequence is present in a genetic element of an anellovector, e.g., as described herein. In some instances, an ITR present in a genetic element of an anellovector may be positioned at a terminal end (e.g., a 5' terminal end or a 3' terminal end) of the genetic element. In some instances, an ITR present in a genetic element of an anellovector may not be positioned at a terminal end (e.g., a 5' terminal end or a 3' terminal end) of the genetic element, e.g., may be flanked by nucleic acid sequences at its 5' and 3' ends (e.g., in a circular genetic element or in a linear genetic element).

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.

As used herein, the term "non-Anellovirus" sequence refers to a sequence from a virus that is not classified in the family *Anelloviridae*. A non-Anellovirus sequence generally: (i) does not comprise a nucleic acid sequence identical to a genome, gene, or non-coding functional element (e.g., an origin of replication) of a virus classified in the family *Anelloviridae* (e.g., an Alphatorquevirus, a Betatorquevirus, or a Gammatorquevirus, e.g., as describd herein); and/or does not encode one or more proteins from a virus not classified in the family *Anelloviridae* (e.g., a capsid protein or a Rep protein). In some instances, a non-Anellovirus sequence has no more than 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, or 90% sequence identity to a genome, gene, or non-coding functional element (e.g., an origin of replication) of any virus classified in the family *Anelloviridae* (e.g., an Alphatorquevirus, a Betatorquevirus, or a

Gammatorquevirus, e.g., as describd herein). In some embodiments, the non-Anellovirus sequence is a wild-type sequence from a virus not classified in the family *Anelloviridae*. In other embodiments, the non-Anellovirus sequence from the virus not classified in the family *Anelloviridae* comprises one or more non-naturally occurring mutations from the genome of the virus. In some instances, a non-Anellovirus sequence is from a virus that infects a non-human organism (e.g., a non-human primate, a non-human mammal, or a bird). In some instances, a non-Anellovirus sequence is from a virus that infects humans. In some instances, a non-Anellovirus sequence is from a virus selected from the group consisting of: a Monodnavirus, e.g., a Shotokuvirus (e.g., a Cressdnaviricota [e.g., a redondovirus, circovirus {e.g., a porcine circovirus, e.g., PCV-1 or PCV-2; or beak-and-feather disease virus}, geminivirus {e.g., tomato golden mosaic virus}, or nanovirus {e.g., BBTV, MDV1, SCSVF, or FBNYV}]), and a Parvovirus (e.g., a dependoparavirus, e.g., a bocavirus or an AAV).

"ORF molecule" refers to a polypeptide having an activity and/or a structural feature of an Anellovirus ORF protein (e.g., 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., 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.

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 compising 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).

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).

"Origin of replication," as used herein, refers to a nucleic acid sequence comprising a sequence which, in the presence of a Rep molecule (e.g., a viral Rep protein, e.g., a non-Anellovirus Rep protein, e.g., an AAV Rep protein, e.g., as described herein), promotes DNA replication. In some instances, an origin of replication situated within a nucleic acid molecule (e.g., a genetic element as described herein) promotes replication of the genetic element, or a portion thereof, in the presence of a Rep molecule to a greater degree than an otherwise similar nucleic acid molecule lacking the origin of replication. In some instances, an origin of replication is comprised in an inverted terminal repeat (ITR) sequence, e.g., of a non-Anellovirus genome, e.g., an AAV genome, e.g., as described herein. In some instances, an origin of replication comprises one or both of a Rep-binding motif (RBM) and/or a terminal resolution site (TRS), e.g., from a non-Anellovirus (e.g., an AAV), e.g., as described herein. In other instances, an origin of replication comprises an Anellovirus origin of replication. As used herein, an "AAV origin of replication" refers to a nucleic acid sequence comprising a sequence, which, in the presence of an AAV Rep molecule (e.g., an AAV Rep protein), promotes DNA replication. In some instances, an AAV origin of replication is recognized and bound by an AAV Rep molecule (e.g., an AAV Rep protein). In some instances, an AAV origin of replication comprises a terminal resolution site (TRS) (e.g., an AAV TRS, e.g., as described herein) and/or a Rep-binding motif (RBM) (e.g., an AAV RBM, e.g., as described herein). In some embodiments, the AAV origin of replication is situated in an AAV ITR.

As used herein, the term "proteinaceous exterior" refers to an exterior component that is predominantly (e.g., >50%, >60%, > 70%, >80%, >90%, >95%, >96%, >97%, >98%, or >99%) protein.

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 embodiments, the expression product comprises RNA or protein.

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.

As used herein, the term "Rep molecule" refers to a protein, e.g., a viral protein, that promotes viral genome replication. In some embodiments, the Rep molecule is a non-Anellovirus Rep protein (e.g., an AAV Rep protein), e.g., as described herein. In some embodiments, the Rep molecule is an Anellovirus Rep molecule, e.g., an Anellovirus ORF2 molecule, e.g., as described herein. An "AAV Rep molecule," as used herein, generally refers to a protein having the functionality of a wild-type AAV Rep protein, e.g., having the capacity to bind to an AAV RBM (e.g., a wild-type AAV RBM, e.g., as described herein, or an RBM having an RBM consensus sequence as described herein) and inducing replication of a nucleic acid molecule comprising the AAV RBM.

As used herein, the term "Rep-binding motif" ("RBM") refers to a nucleic acid sequence from a viral genome (e.g., a non-Anellovirus genome, e.g., an AAV genome), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto, which binds a Rep molecule. Generally, an RBM has at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an RBM sequence as described herein (e.g., an AAV RBM sequence as described herein). In some instances, an RBM is comprised in an origin of replication, e.g., in a genetic element of an anellovector. In some instances, an RBM is positioned within about 1, 2, 3, 4, 5, 10, 15, 20, 25, or 30 nucleotides of a terminal resolution site (TRS), e.g., as described herein. In some instances, an RBM is positioned about 13 nucleotides from a TRS. In some instances, an RBM is positioned 3' relative to a TRS. In some instances, an RBM recruits a Rep molecule to the origin of replication.

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 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.

As used herein, the term "non-pathogenic" refers to an organism or component thereof that does not cause or induce unacceptable disease or pathogenic condition, e.g., in a host organism, e.g., a mammal, e.g., a human.

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).

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 embodiments, the substantially non-immunogenic organism, particle, or component does not produce a clinically significant immune response. In 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 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 antibody) 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).

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 subseqence comprising an exogenous sequence or a sequence heterologous to the remainder of the larger sequence, e.g., a corresponding subsequence from a different *Anellovirus*).

As used herein, the term "terminal resolution site" ("TRS") refers to a nucleic acid sequence having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the TRS sequence of the genome of a virus, e.g., as described herein (e.g., an AAV TRS

sequence as described herein). In some instances, a TRS is cleaved by a Rep molecule (e.g., via endonuclease activity of the rep molecule). In some instances, cleavage of the TRS by a Rep molecule produces a 3' hydroxyl end for replication of the nucleic acid molecule comprising the TRS. In some instances, a TRS is comprised in an origin of replication, e.g., in a genetic element of an anellovector. In some instances, a TRS is positioned within about 1, 2, 3, 4, 5, 10, 15, 20, 25, or 30 nucleotides of a Rep-binding motif (RBM), e.g., as described herein. In some instances, a TRS is positioned about 13 nucleotides from an RBM. In some instances, a TRS is positioned 5' relative to an RBM.

As used herein, "treatment", "treating" and cognates thereof refer to the medical management of a subject with the intent to improve, ameliorate, stabilize, prevent or cure a disease, pathological condition, or disorder. This term includes 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 supportive treatment (treatment employed to supplement another therapy).

This invention relates generally to anellovectors, e.g., synthetic anellovectors, methods of administration of 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 will include a genetic element comprising a nucleic acid sequence (e.g., 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

i. Non-Anellovirus sequences (e.g., AAV sequences)

ii. Plasmids

iii. Circular nucleic acid constructs

iv. In vitro circularization

v. Tandem constructs

vi. Cis/trans constructsvii. Expression cassettes

viii. Design and production of a genetic element construct

C. Effectors

D. Host Cells

i. Introduction of genetic elements into host cells

ii. Methods for providing protein(s) in cis or trans

iii. Helpers, e.g., non-Anellovirus helpers

iv. Exemplary cell types

E. Culture Conditions

F. Harvest

G. In vitro assembly methods

H. Enrichment and Purification

II. Anellovectors

A. Anelloviruses

B. ORF1 molecules

C. ORF2 molecules

D. Genetic elements, e.g., genetic elements including non-Anellovirus sequences

E. Protein binding sequences

F. 5' UTR Regions

G. GC-rich regions

H. Effectors

I. Regulatory Sequences

J. Replication Proteins

K. Other Sequences

L. Proteinaceous exterior

III. Nucleic Acid Constructs

IV. Compositions

V. Methods of Use

VI. Administration/ Delivery

I. Compositions and Methods for Making Anellovectors

The present disclosure provides, in some aspects, anellovectors and methods thereof for delivering effectors. In some embodiments, the anellovectors or components thereof can be made as described below. 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 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. In some embodiments, the anellovectors or components thereof can be made using a tandem construct, e.g., as described in U.S. Provisional Application 63/038,483, which is incorporated herein by reference in its entirety. In some embodiments, the anellovectors or components thereof can be made using a bacmid/insect cell system, e.g., as described as described in U.S. Provisional Application Number 63/038,603, which is incorporated herein by reference in its entirety.

Without wishing to be bound by theory, rolling circle amplification may occur via Rep protein binding to a Rep binding site (e.g., comprising a 5' UTR, e.g., comprising a hairpin loop and/or an origin of replication, e.g., as described herein) positioned 5' relative to (or within the 5' region of) the genetic element region. The Rep protein may then proceed through the genetic element region, resulting in the synthesis of the genetic element. The genetic element may then be circularized and then enclosed within a proteinaceous exterior to form an anellovector.

Components and Assembly of Anellovectors

The compositions and methods herein can be used to produce anellovectors. As described herein, an anellovector generally comprises a genetic element (e.g., a single-stranded, circular DNA molecule, e.g., comprising a 5' UTR region as described herein) 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).

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., as described herein). 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.

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 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. 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. In other embodiments, the nucleic acid construct is a double-stranded minicircle produced by circularizing the nucleic acid sequence of the genetic element *in vitro*. In embodiments, the *in vitro*-circularized (IVC) minicircle is introduced into the host cell, where it is converted to a single-stranded genetic element suitable for enclosure in a proteinaceous exterior, as described herein.

ORF1 Molecules, e.g., for assembly of Anellovectors

An anellovector can be made, for example, by enclosing a genetic element within a proteinaceous exterior. 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).

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.

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, e.g., produced using a composition or construct 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.

In some embodiments, a first plurality of anellovectors comprising an ORF1 molecule as described herein is administered to a subject. In some embodiments, a second plurality of anellovectors comprising an ORF1 molecule described herein, is subsequently administered to the subject following administration of the first plurality. In some embodiments the second plurality of anellovectors comprises an ORF1 molecule having the same amino acid sequence as the ORF1 molecule comprised by the anellovectors of the first plurality. In some embodiments the second plurality of anellovectors comprises an ORF1 molecule having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% amino acid sequence identity to the ORF1 molecule comprised by the anellovectors of the first plurality.

ORF2 Molecules, e.g., for assembly of Anellovectors

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.

Production of protein components

Protein components of an anellovector, e.g., ORF1, can be produced in a variety of ways, e.g., as described herein. In some embodiments, the protein components of an anellovector, including, e.g., the proteinaceous exterior, are produced in the same host cell that packages the genetic elements into the proteinaceous exteriors, thereby producing the anellovectors. In some embodiments, the protein components of an anellovector, including, e.g., the proteinaceous exterior, are produced in a cell that does not comprise a genetic element and/or a genetic element construct (e.g., as described herein).

Baculovirus expression systems

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 GVs 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 GVs 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.

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 baculovirual 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 promoter 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.

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-transfecting 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.

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-*att*Tn7 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, recombinant bacmid DNA containing-colonies are identified and bacmid DNA is isolated to transfect insect cells.

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 US patent applications Nos. 14/447,341, 14/277,892, and 12/278,916, which are hereby incorporated by reference.

Insect cell systems

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 US Patent Application Nos. 14/564,512 and 14/775,154, each of which is hereby incorporated by reference.

Mammalian cell systems

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.

Genetic Element Constructs, e.g., for assembly of Anellovectors

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 vector backbone. Generally, 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 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, 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, 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.

In some aspects, the present disclosure provides a method for replication and propagation of the anellovector as described herein (e.g., in a cell culture system), which may comprise one or more of the following steps: (a) introducing (e.g., transfecting) a genetic element (e.g., linearized) into a cell line sensitive to anellovector infection; (b) harvesting the cells and optionally isolating cells showing the presence of the genetic element; (c) culturing the cells obtained in step (b) (e.g., for at least three days, such as at least one week or longer), depending on experimental conditions and gene expression; and (d) harvesting the cells of step (c), e.g., as described herein.

Non-Anellovirus Sequences

A genetic element construct as described herein may comprise a nucleic acid sequence (e.g., a sequence with a length of at least 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, or 4000 nucleotides) from the genome of a non-Anellovirus virus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. Examples of viruses from which the non-Anellovirus sequence can be derived include, without limitation, a Monodnavirus, e.g., a Shotokuvirus (e.g., a Cressdnaviricota [e.g., a redondovirus, circovirus {e.g., a porcine circovirus, e.g., PCV-1 or PCV-2; or beak-and-feather disease virus}, geminivirus {e.g., tomato golden mosaic virus}, or nanovirus {e.g., BBTV, MDV1, SCSVF, or FBNYV}]), or a Parvovirus (e.g., a dependoparavirus, e.g., a bocavirus or an Adeno-associated virus (AAV). In some instances, the genetic element construct comprises a sequence from a Monodnavirus, e.g., Shotokuvirus, e.g., Cossaviricota, e.g., Quintoviricetes, e.g., Piccovirales, e.g., Parvoviridae, e.g., Parvovirinae, e.g., Dependoparvovirus, e.g., an AAV. In some instances, the genetic element comprises a sequence from an AAV (e.g., AAV1, AAV2, or AAV5).

In some instances, the genetic element construct comprises a non-Anellovirus origin of replication, e.g., as described herein. A non-Anellovirus origin of replication may, in some instances, be comprised in an ITR from the non-Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. A non-Anellovirus origin of replication may, in some instances, comprise a Rep-binding motif (RBM) of the non-Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. A non-Anellovirus origin of replication may, in some instances, comprise a terminal resolution site (TRS) of the non-Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

Plasmids

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. In some embodiments, the genetic element can be excised from a plasmid (e.g., by *in vitro* circularization), for example, to form a minicircle, e.g., as described herein. In embodiments, excision of the genetic element separates the genetic element sequence from the plasmid backbone (e.g., separates the genetic element from a bacterial backbone).

Small circular nucleic acid constructs

In some embodiments, the genetic element construct is a circular nucleic acid construct, e.g., lacking a backbone (e.g., lacking a bacterial origin of replication and/or selectable marker). In embodiments, the genetic element is a double-stranded circular nucleic acid construct. In embodiments, the double-stranded 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.

In vitro circularization

In some instances, the genetic element to be packaged into a proteinaceous exterior is a single stranded circular DNA. The genetic element may, in some instances, be introduced into a host cell via a genetic element construct having a form other than a single stranded circular DNA. For example, the genetic element construct may be a double-stranded circular DNA. The double-stranded circular DNA may then be converted into a single-stranded circular DNA in the host cell (e.g., a host cell comprising a suitable enzyme for rolling circle replication, e.g., an *Anellovirus* Rep protein, e.g., Rep68/78, Rep60, RepA, RepB, Pre, MobM, TraX, TrwC, Mob02281, Mob02282, NikB, ORF50240, NikK, TecH, OrfJ, or

TraI, e.g., as described in Wawrzyniak et al. 2017, *Front. Microbiol.* 8: 2353; incorporated herein by reference with respect to the listed enzymes). In some embodiments, the double-stranded circular DNA is produced by *in vitro* circularization (IVC), e.g., as described in Example 15.

Generally, *in vitro* circularized DNA constructs can be produced by digesting a genetic element construct (e.g., a plasmid comprising the sequence of a genetic element) to be packaged, such that the genetic element sequence is excised as a linear DNA molecule. The resultant linear DNA can then be ligated, e.g., using a DNA ligase, to form a double-stranded circular DNA. In some instances, a double-stranded circular DNA produced by *in vitro* circularization can undergo rolling circle replication, e.g., as described herein. Without wishing to be bound by theory, it is contemplated that *in vitro* circularization results in a double-stranded DNA construct that can undergo rolling circle replication without further modification, thereby being capable of producing single-stranded circular DNA of a suitable size to be packaged into an anellovector, e.g., as described herein. In some embodiments, the double-stranded DNA construct is smaller than a plasmid (e.g., a bacterial plasmid). In some embodiments, the double-stranded DNA construct is excised from a plasmid (e.g., a bacterial plasmid) and then circularized, e.g., by *in vitro* circularization.

Tandem Constructs

In some embodiments, a genetic element construct comprises a first copy of a genetic element sequence (e.g., the nucleic acid sequence of a genetic element, e.g., as described herein) and at least a portion of a second copy of a genetic element sequence (e.g., the nucleic acid sequence of the same genetic element, or the nucleic acid sequence of a different genetic element), arranged in tandem. Genetic element constructs having such a structure are generally referred to herein as tandem constructs. Such tandem constructs are used for producing an anellovector genetic element. The first copy of the genetic element sequence and the second copy of the genetic element sequence may, in some instances, be immediately adjacent to each other on the genetic acid construct. In other instances, the first copy of the genetic element sequence and the second copy of the genetic element sequence may be separated, e.g., by a spacer sequence. In some embodiments, the second copy of the genetic element sequence, or the portion thereof, comprises an upstream replication-facilitating sequence (uRFS), e.g., as described herein. In some embodiments, the second copy of the genetic element sequence, or the portion thereof, comprises a downstream replication-facilitating sequence (dRFS), e.g., as described herein. In some embodiments, the uRFS and/or dRFS comprises an origin of replication (e.g., a mammalian origin of replication, an insect origin of replication, or a viral origin of replication, e.g., a non-Anellovirus origin of replication, e.g., as described herein) or portion thereof. In some embodiments, the uRFS and/or dRFS does not comprise an origin of replication. In some embodiments, the uRFS and/or dRFS comprises a hairpin loop

(e.g., in the 5' UTR). In some embodiments, a tandem construct produces higher levels of a genetic element than an otherwise similar construct lacking the second copy of the genetic element or portion thereof. Without being bound by theory, a tandem construct described herein may, in some embodiments, replicate by rolling circle replication. In some embodiments, a tandem construct is a plasmid. In some embodiments, a tandem construct is circular. In some embodiments, a tandem construct is linear. In some embodiments, a tandem construct is single-stranded. In some embodiments, a tandem construct is double-stranded. In some embodiments, a tandem construct is DNA.

A tandem construct may, in some instances, include a first copy of the sequence of the genetic element and a second copy of the sequence of the genetic element, or a portion thereof. It is understood that the second copy can be an identical copy of the first copy or a portion thereof, or can comprise one or more sequence differences, e.g., substitutions, additions, or deletions. In some instances, the second copy of the genetic element sequence or portion thereof is positioned 5' relative to the first copy of the genetic element sequence. In some instances, the second copy of the genetic element sequence or portion thereof is positioned 3' relative to the first copy of the genetic element sequence. In some instances, the second copy of the genetic element sequence or portion thereof and the first copy of the genetic element sequence are adjacent to each other in the tandem construct. In some instances, the second copy of the genetic element sequence or portion thereof and the first copy of the genetic element sequence are separated, e.g., by a spacer sequence.

In some embodiments, the tandem constructs described herein can be used to produce the genetic element of a vector (e.g., anellovector), vehicle, or particle (e.g., viral particle) comprising a capsid (e.g., a capsid comprising an Anellovirus ORF, e.g., an ORF1 molecule, e.g., as described herein) encapsulating a genetic element comprising a protein binding sequence that binds to the capsid and a heterologous (e.g., relative to the Anellovirus from which the ORF1 molecule was derived) sequence encoding a therapeutic effector. In embodiments, the vector is capable of delivering the genetic element into a mammalian, e.g., human, cell. In some embodiments, the genetic element has less than about 50% (e.g., less than 50%, 40%, 30%, 25%, 20%, 15%, 10%, 9%, 8%, 7%, 6%, 5.5%, 5%, 4.5%, 4%, 3.5%, 3%, 2.5%, 2%, 1.5%, or less) identity to a wild type Anellovirus genome sequence. In some embodiments, the genetic element has no more than 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, 5.5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, or 80% identity to a wild type Anellovirus genome sequence. In some embodiments, the genetic element has greater than about 2000, 3000, 4000, 4500, or 5000 contiguous nucleotides of non-Anellovirus genome sequence. In some embodiments, the genetic element has greater than about 2000 to 5000, 2500 to 4500, 3000 to 4500, 2500 to 4500, 3500, or 4000, 4500 (e.g., between about 3000 to 4500) nucleotides nucleotides of non-Anellovirus genome sequence.

In some embodiments of the systems and methods herein, a vector (e.g., an anellovector) is made by introducing into a cell a first nucleic acid molecule that is a genetic element or genetic element construct, e.g., a tandem construct, and a second nucleic acid molecule encoding one or more additional proteins (e.g., a Rep molecule and/or a capsid protein), e.g., as described herein. In some embodiments, the first nucleic acid molecule and the second nucleic acid molecule are attached to each other (e.g., in a genetic element construct described herein, e.g., in cis). In some embodiments, the first nucleic acid molecule and the second nucleic acid molecule are separate (e.g, in trans). In some embodiments, the first nucleic acid molecule is a plasmid, cosmid, bacmid, minicircle, or artificial chromosome. In some embodiments, the second nucleic acid molecule is a plasmid, cosmid, bacmid, minicircle, or artificial chromosome. In some embodiments, the second nucleic acid molecule is integrated into the genome of the host cell.

In some embodiments, the method further includes introducing the first nucleic acid molecule and/or the second nucleic acid molecule into the host cell. In some embodiments, the second nucleic acid molecule is introduced into the host cell prior to, concurrently with, or after the first nucleic acid molecule. In other embodiments, the second nucleic acid molecule is integrated into the genome of the host cell. In some embodiments, the second nucleic acid molecule is or comprises or is part of a helper construct, helper virus or other helper vector, e.g., as described herein.

Cis/Trans Constructs

In some embodiments, a genetic element construct 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, an ORF1 molecule is provided in *trans*, e.g., as described herein. In some embodiments,

an ORF2 molecule is provided in *trans*, e.g., as described herein. In some embodiments, an ORF1 molecule and an ORF1 molecule are both provided in *trans*, e.g., as described herein.

In some embodiments, the genetic element construct 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).

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

In some embodiments, a genetic element construct 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).

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 α-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 metallothionine (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.

Design and Production of a Genetic Element Construct

Various methods are available for synthesizing a genetic element construct. For instance, the genetic element construct sequence may be divided into smaller overlapping pieces (e.g., in the range of about 100 bp to about 10 kb segments or individual ORFs) that are easier to synthesize. These DNA segments are synthesized from a set of overlapping single-stranded oligonucleotides. The resulting overlapping synthons are then assembled into larger pieces of DNA, e.g., the genetic element construct. The segments or ORFs may be assembled into the genetic element construct, e.g., by in vitro recombination or unique restriction sites at 5' and 3' ends to enable ligation.

The genetic element construct can be synthesized with a design algorithm that parses the construct sequence into oligo-length fragments, creating suitable design conditions for synthesis that take into account the complexity of the sequence space. Oligos are then chemically synthesized on semiconductor-based, high-density chips, where over 200,000 individual oligos are synthesized per chip. The oligos are assembled with an assembly techniques, such as BioFab®, to build longer DNA segments from the smaller oligos. This is done in a parallel fashion, so hundreds to thousands of synthetic DNA segments are built at one time.

Each genetic element construct or segment of the genetic element construct may be sequence verified. In some embodiments, high-throughput sequencing of RNA or DNA can take place using AnyDot.chips (Genovoxx, Germany), which allows for the monitoring of biological processes (e.g., miRNA expression or allele variability (SNP detection). Other high-throughput sequencing systems include those disclosed in Venter, J., et al. Science 16 Feb. 2001; Adams, M. et al, Science 24 Mar. 2000; and M. J, Levene, et al. Science 299:682-686, January 2003; as well as US Publication Application No. 20030044781 and 2006/0078937. Overall such systems involve sequencing a target nucleic acid molecule having a plurality of bases by the temporal addition of bases via a polymerization reaction that is measured on a molecule of nucleic acid, i.e., the activity of a nucleic acid polymerizing enzyme on the template nucleic acid molecule to be sequenced is followed in real time. In some embodiments, shotgun sequencing is performed.

A genetic element construct can be designed such that factors for replicating or packaging may be supplied in *cis* or in *trans*, relative to the genetic element. For example, when supplied in *cis*, the genetic element may comprise one or more genes encoding an *Anellovirus* ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3, e.g., as described herein. In some embodiments, replication and/or packaging signals can be incorporated into a genetic element, for example, to induce amplification and/or encapsulation. In some embodiments, an effector is inserted into a specific site in the genome. In some embodiments, one or more viral ORFs are replaced with an effector.

In another example, when replication or packaging factors are supplied in *trans*, the genetic element may lack genes encoding one or more of an *Anellovirus* ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3, e.g., as described herein; this protein or proteins may be supplied, e.g., by another nucleic acid, e.g., a helper nucleic acid. In some embodiments, minimal *cis* signals (e.g., 5' UTR and/or GC-rich region) are present in the genetic element. In some embodiments, the genetic element does not encode replication or packaging factors (e.g., replicase and/or capsid proteins). Such factors may, in some embodiments, be supplied by one or more helper nucleic acids (e.g., a helper viral nucleic acid, a helper plasmid, or a helper nucleic acid integrated into the host cell genome). In some embodiments, the helper nucleic acids express proteins and/or RNAs sufficient to induce amplification and/or packaging, but may lack their own packaging signals. In some embodiments, the genetic element and the helper nucleic acid are introduced into the host cell (e.g., concurrently or separately), resulting in amplification and/or packaging of the genetic element but not of the helper nucleic acid.

In some embodiments, the genetic element construct may be designed using computer-aided design tools.

General methods of making constructs are described in, for example, Khudyakov & Fields, *Artificial DNA: Methods and Applications*, CRC Press (2002); in Zhao, *Synthetic Biology: Tools and*

Applications, (First Edition), Academic Press (2013); and Egli & Herdewijn, *Chemistry and Biology of Artificial Nucleic Acids*, (First Edition), Wiley-VCH (2012).

Effectors

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. 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 some embodiments, the effector-encoding sequence may be inserted into the genetic element e.g., at a non-coding region, e.g., a noncoding region disposed 3' of the open reading frames and 5' of the GC-rich region of the genetic element, in the 5' noncoding region upstream of the TATA box, in the 5' UTR, in the 3' noncoding region downstream of the poly-A signal, or upstream of the GC-rich region. In some embodiments, the effector-encoding sequence may be inserted into the genetic element, e.g., in a coding sequence (e.g., in a sequence encoding an Anellovirus ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3, e.g., as described herein). In some embodiments, the effector-encoding sequence replaces all or a part of the open reading frame. In some embodiments, the genetic element comprises a regulatory sequence (e.g., a promoter or enhancer, e.g., as described herein) operably linked to the effector-encoding sequence.

Host Cells

The anellovectors described herein can be produced, for example, in a host cell. Generally, a host cell is provided that comprises an anellovector genetic element and the components of an anellovector proteinaceous exterior (e.g., a polypeptide encoded by an Anellovirus ORF1 nucleic acid or an Anellovirus ORF1 molecule). The host cell is then incubated under conditions suitable for enclosure of the genetic element within the proteinaceous exterior (e.g., culture conditions as described herein). In some embodiments, the host cell is further incubated under conditions suitable for release of the anellovector from the host cell, e.g., into the surrounding supernatant. In some embodiments, the host cell is lysed for harvest of anellovectors from the cell lysate. In some embodiments, an anellovector may be introduced to a host cell line grown to a high cell density. In some embodiments, a host cell is an Expi-293 cell.

Introduction of genetic elements into host cells

The genetic element, or a nucleic acid construct comprising the sequence of a genetic element, may be introduced into a host cell. In some embodiments, the genetic element itself is introduced into the host cell. In some embodiments, a genetic element construct comprising the sequence of the genetic element (e.g., as described herein) is introduced into the host cell. A genetic element or genetic element construct can be introduced into a host cell, for example, using methods known in the art. For example, a genetic element or genetic element construct can be introduced into a host cell by transfection (e.g., stable transfection or transient transfection). In embodiments, the genetic element or genetic element construct is introduced into the host cell by lipofectamine transfection. In embodiments, the genetic element or genetic element construct is introduced into the host cell by calcium phosphate transfection. In some embodiments, the genetic element or genetic element construct is introduced into the host cell by electroporation. In some embodiments, the genetic element or genetic element construct is introduced into the host cell using a gene gun. In some embodiments, the genetic element or genetic element construct is introduced into the host cell by nucleofection. In some embodiments, the genetic element or genetic element construct is introduced into the host cell by PEI transfection. In some embodiments, the genetic element is introduced into the host cell by contacting the host cell with an anellovector comprising the genetic element

In embodiments, the genetic element construct is capable of replication once introduced into the host cell. In embodiments, the genetic element can be produced from the genetic element construct once introduced into the host cell. In some embodiments, the genetic element is produced in the host cell by a polymerase, e.g., using the genetic element construct as a template.

In some embodiments, the genetic elements or vectors comprising the genetic elements are introduced (e.g., transfected) into cell lines that express a viral polymerase protein in order to achieve expression of the anellovector. To this end, cell lines that express an anellovector polymerase protein may be utilized as appropriate host cells. Host cells may be similarly engineered to provide other viral functions or additional functions.

To prepare the anellovector disclosed herein, a genetic element construct may be used to transfect cells that provide anellovector proteins and functions required for replication and production. Alternatively, cells may be transfected with a second construct (e.g., a virus) providing anellovector proteins and functions before, during, or after transfection by the genetic element or vector comprising the genetic element disclosed herein. In some embodiments, the second construct may be useful to complement production of an incomplete viral particle. The second construct (e.g., virus) may have a conditional growth defect, such as host range restriction or temperature sensitivity, e.g., which allows the

subsequent selection of transfectant viruses. In some embodiments, the second construct may provide one or more replication proteins utilized by the host cells to achieve expression of the anellovector. In some embodiments, the host cells may be transfected with vectors encoding viral proteins such as the one or more replication proteins. In some embodiments, the second construct comprises an antiviral sensitivity.

The genetic element or vector comprising the genetic element disclosed herein can, in some instances, be replicated and produced into anellovectors using techniques known in the art. For example, various viral culture methods are described, e.g., in U.S. Pat. No. 4,650,764; U.S. Pat. No. 5,166,057; U.S. Pat. No. 5,854,037; European Patent Publication EP 0702085A1; U.S. patent application Ser. No. 09/152,845; International Patent Publications PCT WO97/12032; WO96/34625; European Patent Publication EP-A780475; WO 99/02657; WO 98/53078; WO 98/02530; WO 99/15672; WO 98/13501; WO 97/06270; and EPO 780 47SA1, each of which is incorporated by reference herein in its entirety.

Methods for providing protein(s) in cis or trans

In some embodiments (e.g., *cis* embodiments described herein), the genetic element construct further comprises one or more expression cassettes comprising a coding sequence for an Anellovirus ORF (e.g., an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, or a functional fragment thereof). In embodiments, the genetic element construct comprises an expression cassette comprising a coding sequence for an Anellovirus ORF1, or a splice variant or functional fragment thereof. Such genetic element constructs, which comprise expression cassettes for the effector as well as the one or more Anellovirus ORFs, may be introduced into host cells. Host cells comprising such genetic element constructs may, in some instances, be capable of producing the genetic elements and components for proteinaceous exteriors, and for enclosure of the genetic elements within proteinaceous exteriors, without requiring additional nucleic acid constructs or integration of expression cassettes into the host cell genome. In other words, such genetic element constructs may be used for *cis* anellovector production methods in host cells, e.g., as described herein.

In some embodiments (e.g., *trans* embodiments described herein), the genetic element does not comprise an expression cassette comprising a coding sequence for one or more Anellovirus ORFs (e.g., an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, or a functional fragment thereof). In embodiments, the genetic element construct does not comprise an expression cassette comprising a coding sequence for an Anellovirus ORF1, or a splice variant or functional fragment thereof. Such genetic element constructs, which comprise expression cassettes for the effector but lack expression cassettes for one or more Anellovirus ORFs (e.g., Anellovirus ORF1 or a splice variant or functional fragment thereof), may be introduced into host cells. Host cells comprising such genetic element constructs may, in some instances, require additional nucleic acid constructs or integration of expression

cassettes into the host cell genome for production of one or more components of the anellovector (e.g., the proteinaceous exterior proteins). In some embodiments, host cells comprising such genetic element constructs are incapable of enclosure of the genetic elements within proteinaceous exteriors in the absence of an additional nucleic construct encoding an Anellovirus ORF1 molecule. In other words, such genetic element constructs may be used for *trans* anellovector production methods in host cells, e.g., as described herein.

In some embodiments (e.g., *cis* embodiments described herein), the genetic element construct further comprises one or more expression cassettes comprising a coding sequence for one or more non-Anellovirus ORF (e.g., a non-Anellovirus Rep molecule, e.g., an AAV Rep molecule, e.g., an AAV Rep protein, e.g., an AAV Rep2 protein). Such genetic element constructs, which comprise expression cassettes for the effector as well as the one or more non-Anellovirus ORFs, may be introduced into host cells. Host cells comprising such genetic element constructs may, in some instances, be capable of producing the genetic elements and components for proteinaceous exteriors, and for enclosure of the genetic elements within proteinaceous exteriors, without requiring additional nucleic acid constructs or integration of expression cassettes into the host cell genome. In other words, such genetic element constructs may be used for *cis* anellovector production methods in host cells, e.g., as described herein.

In some embodiments (e.g., *trans* embodiments described herein), the genetic element does not comprise an expression cassette comprising a coding sequence for one or more non-Anellovirus ORFs (e.g., a non-Anellovirus Rep molecule, e.g., an AAV Rep molecule, e.g., an AAV Rep protein, e.g., an AAV Rep2 protein). Such genetic element constructs, which comprise expression cassettes for the effector but lack expression cassettes for one or more non-Anellovirus ORFs (e.g., a non-Anellovirus Rep molecule, e.g., an AAV Rep molecule, e.g., an AAV Rep protein, e.g., an AAV Rep2 protein), may be introduced into host cells. Host cells comprising such genetic element constructs may, in some instances, require additional nucleic acid constructs or integration of expression cassettes into the host cell genome for production of one or more components of the anellovector (e.g., for replication of the genetic element). In some embodiments, host cells comprising such genetic element constructs are incapable of replicating the genetic elements in the absence of an additional nucleic construct, e.g., encoding a non-Anellovirus Rep molecule, e.g., an AAV Rep molecule, e.g., an AAV Rep protein, e.g., an AAV Rep2 protein. In other words, such genetic element constructs may be used for *trans* anellovector production methods in host cells, e.g., as described herein.

Helpers and non-Anellovirus molecules

In some embodiments, a molecule (e.g., a nucleic acid molecule or a polypeptide) from a non-Anellovirus virus, or a molecule based thereon, is present in the host cell. The molecule from the non-

Anellovirus virus, or a molecule based thereon, may, in some embodiments, contribute to production of an anellovector as described herein. For example, the molecule from the non-Anellovirus virus, or a molecule based thereon, may comprise a non-Anellovirus Rep molecule (e.g., an AAV Rep molecule) that promotes replication of an anellovector genetic element comprising a cognate origin of replication (e.g., an AAV origin of replication).

In some embodiments, an AAV Rep protein comprises the amino acid sequence as listed in Table 60 below, or an amino acid sequence having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, an AAV Rep protein comprises the amino acid sequence of any of SEQ ID NO: 1030-1042, or an amino acid sequence having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

Table 60. Exemplary AAV Rep protein sequences

Name	Sequence	SEQ ID NO:
AAV2 Rep Sequences		
Rep coding region	ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGC GACCTTGACGAGCATCTGCCCGGCATTTCTGACAGCTTT GTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCA GATTCTGACATGGATCTGAATCTGATTGAGCAGGCACCC CTGACCGTGGCCGAGAAGCTGCAGCGCGACTTTCTGACG GAATGGCGCCGTGTGAGTAAGGCCCCGGAGGCCCTTTTC TTTGTGCAATTTGAGAAGGGAGAGAGCTACTTCCACATG CACGTGCTCGTGGAAACCACCGGGGTGAAATCCATGGTT TTGGGACGTTTCCTGAGTCAGATTCGCGAAAAACTGATT CAGAGAATTTACCGCGGGATCGAGCCGACTTTGCCAAAC TGGTTCGCGGTCACAAAGACCAGAAATGGCGCCGGAGGC GGGAACAAGGTGGTGGATGAGTGCTACATCCCCAATTAC TTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTGG ACTAATATGGAACAGTATTTAAGCGCCTGTTTGAATCTC ACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCAC GTGTCGCAGACGCAGGAGCAGAACAAAGAGAATCAGAAT CCCAATTCTGATGCGCCGGTGATCAGATCAAAAACTTCA GCCAGGTACATGGAGCTGGTCGGGTGGCTCGTGGACAAG GGGATTACCTCGGAGAAGCAGTGGATCCAGGAGGACCAG GCCTCATACATCTCCTTCAATGCGGCCTCCAACTCGCGG TCCCAAATCAAGGCTGCCTTGGACAATGCGGGAAAGATT ATGAGCCTGACTAAAACCGCCCCCGACTACCTGGTGGGC CAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTAT AAAATTTTGGAACTAAACGGGTACGATCCCCAATATGCG GCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTCGGC AAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACC GGGAAGACCAACATCGCGGAGGCCATAGCCCACACTGTG	1030

	CCCTTCTACGGGTGCGTAAACTGGACCAATGAGAACTTT CCCTTCAACGACTGTGTCGACAAGATGGTGATCTGGTGG GAGGAGGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCC AAAGCCATTCTCGGAGGAAGCAAGGTGCGCGTGGACCAG AAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTG ATCGTCACCTCCAACACCAACATGTGCGCCGTGATTGAC GGGAACTCAACGACCTTCGAACACCAGCAGCCGTTGCAA GACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGAT CATGACTTTGGGAAGGTCACCAAGCAGGAAGTCAAAGAC TTTTTCCGGTGGGCAAAGGATCACGTGGTTGAGGTGGAG CATGAATTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGA CCCGCCCCCAGTGACGCAGATATAAGTGAGCCCAAACGG GTGCGCGAGTCAGTTGCGCAGCCATCGACGTCAGACGCG GAAGCTTCGATCAACTACGCAGACAGGTACCAAAACAAA TGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCC TGCAGACAATGCGAGAGAATGAATCAGAATTCAAATATC TGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTT CCCGTGTCAGAATCTCAACCCGTTTCTGTCGTCAAAAAG GCGTATCAGAAACTGTGCTACATTCATCATATCATGGGA AAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAAT GTGGATTTGGATGACTGCATCTTTGAACAATAAATGATT TAAATCAGGTATGGCTGCCGATGGTTATCTTCCAGATTG GCTCGAGGACACTCTCTCTGA	
Rep78 AA	MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPP DSDMDLNLIEQAPLTVAEKLQRDFLTEWRRVSKAPEALF FVQFEKGESYFHMHVLVETTGVKSMVLGRFLSQIREKLI QRIYRGIEPTLPNWFAVTKTRNGAGGNKVVDECYIPNY LLPKTQPELQWAWTNMEQYLSACLNLTERKRLVAQHLTH VSQTQEQNKENQNPNSDAPVIRSKTSARYMELVGWLVDK GITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKI MSLTKTAPDYLVGQQPVEDISSNRIYKILELNGYDPQYA ASVFLGWATKKFGKRNTIWLFGPATTGKTNIAEAIAHTV PFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESA KAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVID GNSTTFEHQQPLQDRMFKFELTRRLDHDFGKVTKQEVKD FFRWAKDHVVEVEHEFYVKKGGAKKRPAPSDADISEPKR VRESVAQPSTSDAEASINYADRYQNKCSRHVGMNLMLFP CRQCERMNQNSNICFTHGQKDCLECFPVSESQPVSVVKK AYQKLCYIHHIMGKVPDACTACDLVNVDLDDCIFEQ	1031
Rep68 AA	MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPP DSDMDLNLIEQAPLTVAEKLQRDFLTEWRRVSKAPEALF FVQFEKGESYFHMHVLVETTGVKSMVLGRFLSQIREKLI QRIYRGIEPTLPNWFAVTKTRNGAGGNKVVDECYIPNY LLPKTQPELQWAWTNMEQYLSACLNLTERKRLVAQHLTH VSQTQEQNKENQNPNSDAPVIRSKTSARYMELVGWLVDK GITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKI MSLTKTAPDYLVGQQPVEDISSNRIYKILELNGYDPQYA	1032

	ASVFLGWATKKFGKRNTIWLFGPATTGKTNIAEAIAHTV PFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESA KAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVID GNSTTFEHQQPLQDRMFKFELTRRLDHDFGKVTKQEVKD FFRWAKDHVVEVEHEFYVKKGGAKKRPAPSDADISEPKR VRESVAQPSTSDAEASINYADRLARGHSL	
Rep52 AA	MELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQI KAALDNAGKIMSLTKTAPDYLVGQQPVEDISSNRIYKIL ELNGYDPQYAASVFLGWATKKFGKRNTIWLFGPATTGKT NIAEAIAHTVPFYGCVNWTNENFPFNDCVDKMVIWWEEG KMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVT SNTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLDHDF GKVTKQEVKDFFRWAKDHVVEVEHEFYVKKGGAKKRPAP SDADISEPKRVRESVAQPSTSDAEASINYADRYQNKCSR HVGMNLMLFPCRQCERMNQNSNICFTHGQKDCLECFPVS ESQPVSVVKKAYQKLCYIHHIMGKVPDACTACDLVNVDL DDCIFEQ	1033
Rep40 AA	MELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQI KAALDNAGKIMSLTKTAPDYLVGQQPVEDISSNRIYKIL ELNGYDPQYAASVFLGWATKKFGKRNTIWLFGPATTGKT NIAEAIAHTVPFYGCVNWTNENFPFNDCVDKMVIWWEEG KMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVT SNTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLDHDF GKVTKQEVKDFFRWAKDHVVEVEHEFYVKKGGAKKRPAP SDADISEPKRVRESVAQPSTSDAEASINYADRLARGHSL	1034
AAV3 Rep Sequences		
Rep coding region	ATGCCGGGGTTCTACGAGATTGTCCTGAAGGTCCCGAGT GACCTGGACGAGCGCCTGCCGGGCATTTCTAACTCGTTT GTTAACTGGGTGGCCGAGAAGGAATGGGACGTGCCGCCG GATTCTGACATGGATCCGAATCTGATTGAGCAGGCACCC CTGACCGTGGCCGAAAAGCTTCAGCGCGAGTTCCTGGTG GAGTGGCGCCGCGTGAGTAAGGCCCCGGAGGCCCTCTTT TTTGTCCAGTTCGAAAAGGGGGAGACCTACTTCCACCTG CACGTGCTGATTGAGACCATCGGGGTCAAATCCATGGTG GTCGGCCGCTACGTGAGCCAGATTAAAGAGAAGCTGGTG ACCCGCATCTACCGCGGGGTCGAGCCGCAGCTTCCGAAC TGGTTCGCGGTGACCAAAACGCGAAATGGCGCCGGGGGC GGGAACAAGGTGGTGGACGACTGCTACATCCCCAACTAC CTGCTCCCCAAGACCCAGCCCGAGCTCCAGTGGGCGTGG ACTAACATGGACCAGTATTTAAGCGCCTGTTTGAATCTC GCGGAGCGTAAACGGCTGGTGGCGCAGCATCTGACGCAC GTGTCGCAGACGCAGGAGCAGAACAAAGAGAATCAGAAC CCCAATTCTGACGCGCCGGTCATCAGGTCAAAAACCTCA GCCAGGTACATGGAGCTGGTCGGGTGGCTGGTGGACCGC GGGATCACGTCAGAAAAGCAATGGATTCAGGAGGACCAG GCCTCGTACATCTCCTTCAACGCCGCCTCCAACTCGCGG	1035

	TCCCAGATCAAGGCCGCGCTGGACAATGCCTCCAAGATC ATGAGCCTGACAAAGACGGCTCCGGACTACCTGGTGGG- CAGCAACCCGCCGGAGGACATTACCAAAAATCGGATCTA CCAAATCCTGGAGCTGAACGGGTACGATCCGCAGTACGC GGCCTCCGTCTTCCTGGGCTGGGCGCAAAAGAAGTTCGG GAAGAGGAACACCATCTGGCTCTTTGGGCCGGCCACGAC GGGTAAAACCAACATCGCGGAAGCCATCGCCCACGCCGT GCCCTTCTACGGCTGCGTAAACTGGACCAATGAGAACTT TCCCTTCAACGATTGCGTCGACAAGATGGTGATCTGGTG GGAGGAGGGCAAGATGACGGCCAAGGTCGTGGAGAGCGC CAAGGCCATTCTGGGCGGAAGCAAGGTGCGCGTGGACCA AAAGTGCAAGTCATCGGCCCAGATCGAACCCACTCCCGT GATCGTCACCTCCAACACCAACATGTGCGCCGTGATTGA CGGGAACAGCACCACCTTCGAGCATCAGCAGCCGCTGCA GGACCGGATGTTTGAATTTGAACTTACCCGCCGTTTGGA CCATGACTTTGGGAAGGTCACCAAACAGGAAGTAAAGGA CTTTTTCCGGTGGGCTTCCGATCACGTGACTGACGTGGC TCATGAGTTCTACGTCAGAAAGGGTGGAGCTAAGAAACG CCCCGCCTCCAATGACGCGGATGTAAGCGAGCCAAAACG GGAGTGCACGTCACTTGCGCAGCCGACAACGTCAGACGC GGAAGCACCGGCGGACTACGCGGACAGGTACCAAAACAA ATGTTCTCGTCACGTGGGCATGAATCTGATGCTTTTTCC CTGTAAAACATGCGAGAGAATGAATCAAATTTCCAATGT CTGTTTTACGCATGGTCAAAGAGACTGTGGGGAATGCTT CCCTGGAATGTCAGAATCTCAACCCGTTTCTGTCGTCAA AAAGAAGACTTATCAGAAACTGTGTCCAATTCATCATAT CCTGGGAAGGGCACCCGAGATTGCCTGTTCGGCCTGCGA TTTGGCCAATGTGGACTTGGATGACTGTGTTTCTGAGCA ATAAATGACTTAAACCAGGTATGGCTGCTGACGGTTATC TTCCAGATTGGCTCGAGGACAACCTTTCTGA	
Rep78 AA	MPGFYEIVLKVPSDLDERLPGISNSFVNWVAEKEWDVPP DSDMDPNLIEQAPLTVAEKLQREFLVEWRRVSKAPEALF FVQFEKGETYFHLHVLIETIGVKSMVVGRYVSQIKEKLV TRIYRGVEPQLPNWFAVTKTRNGAGGNKVVDDCYIPNY LLPKTQPELQWAWTNMDQYLSACLNLAERKRLVAQHLTH VSQTQEQNKENQNPNSDAPVIRSKTSARYMELVGWLVDR GITSEKQWIQEDQASYISFNAASNSRSQIKAALDNASKI MSLTKTAPDYLVGSNPPEDITKNRIYQILELNGYDPQYA ASVFLGWAQKKFGKRNTIWLFGPATTGKTNIAEAIAHAV PFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESA KAILGGSKVRVDQKCKSSAQIEPTPVIVTSNTNMCAVID GNSTTFEHQQPLQDRMFEFELTRRLDHDFGKVTKQEVKD FFRWASDHVTDVAHEFYVRKGGAKKRPASNDADVSEPKR ECTSLAQPTTSDAEAPADYADRYQNKCSRHVGMNLMLFP CKTCERMNQISNVCFTHGQRDCGECFPGMSESQPVSVVK KKTYQKLCPIHHILGRAPEIACSACDLANVDLDDCVSEQ	1036

Rep68 AA	MPGFYEIVLKVPSDLDERLPGISNSFVNWVAEKEWDVPP DSDMDPNLIEQAPLTVAEKLQREFLVEWRRVSKAPEALF FVQFEKGETYFHLHVLIETIGVKSMVVGRYVSQIKEKLV TRIYRGVEPQLPNWFAVTKTRNGAGGGNKVVDDCYIPNY LLPKTQPELQWAWTNMDQYLSACLNLAERKRLVAQHLTH VSQTQEQNKENQNPNSDAPVIRSKTSARYMELVGWLVDR GITSEKQWIQEDQASYISFNAASNSRSQIKAALDNASKI MSLTKTAPDYLVGSNPPEDITKNRIYQILELNGYDPQYA ASVFLGWAQKKFGKRNTIWLFGPATTGKTNIAEAIAHAV PFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESA KAILGGSKVRVDQKCKSSAQIEPTPVIVTSNTNMCAVID GNSTTFEHQQPLQDRMFEFELTRRLDHDFGKVTKQEVKD FFRWASDHVTDVAHEFYVRKGGAKKRPASNDADVSEPKR ECTSLAQPTTSDAEAPADYADRLARGQPF	1037
Rep52 AA	MELVGWLVDRGITSEKQWIQEDQASYISFNAASNSRSQI KAALDNASKIMSLTKTAPDYLVGSNPPEDITKNRIYQIL ELNGYDPQYAASVFLGWAQKKFGKRNTIWLFGPATTGKT NIAEAIAHAVPFYGCVNWTNENFPFNDCVDKMVIWWEEG KMTAKVVESAKAILGGSKVRVDQKCKSSAQIEPTPVIVT SNTNMCAVIDGNSTTFEHQQPLQDRMFEFELTRRLDHDF GKVTKQEVKDFFRWASDHVTDVAHEFYVRKGGAKKRPAS NDADVSEPKRECTSLAQPTTSDAEAPADYADRYQNKCSR HVGMNLMLFPCKTCERMNQISNVCFTHGQRDCGECFPGM SESQPVSVVKKKTYQKLCPIHHILGRAPEIACSACDLAN VDLDDCVSEQ	1038
Rep40 AA	MELVGWLVDRGITSEKQWIQEDQASYISFNAASNSRSQI KAALDNASKIMSLTKTAPDYLVGSNPPEDITKNRIYQIL ELNGYDPQYAASVFLGWAQKKFGKRNTIWLFGPATTGKT NIAEAIAHAVPFYGCVNWTNENFPFNDCVDKMVIWWEEG KMTAKVVESAKAILGGSKVRVDQKCKSSAQIEPTPVIVT SNTNMCAVIDGNSTTFEHQQPLQDRMFEFELTRRLDHDF GKVTKQEVKDFFRWASDHVTDVAHEFYVRKGGAKKRPAS NDADVSEPKRECTSLAQPTTSDAEAPADYADRLARGQPF	1039
AAV5 Rep Sequences		
Rep78 AA	MATFYEVIVRVPFDVEEHLPGISDSFVDWVTGQIWELPP ESDLNLTLVEQPQLTVADRIRRVFLYEWNKFSKQESKFF VQFEKGSEYFHLHTLVETSGISSMVLGRYVSQIRAQLVK VVFQGIEPQINDWVAITKVKKGGANKVVDSGYIPAYLLP KVQPELQWAWTNLDEYKLAALNLEERKRLVAQFLAESSQ RSQEAASQREFSADPVIKSKTSQKYMALVNWLVEHGITS EKQWIQENQESYLSFNSTGNSRSQIKAALDNATKIMSLT KSAVDYLVGSSVPEDISKNRIWQIFEMNGYDPAYAGSIL YGWCQRSFNKRNTVWLYGPATTGKTNIAEAIAHTVPFYG CVNWTNENFPFNDCVDKMLIWWEEGKMTNKVVESAKAIL GGSKVRVDQKCKSSVQIDSTPVIVTSNTNMCVVVDGNST TFEHQQPLEDRMFKFELTKRLPPDFGKITKQEVKDFFAW	1040

	AKVNQVPVTHEFKVPRELAGTKGAEKSLKRPLGDVTNTS YKSLEKRARLSFVPETPRSSDVTVDPAPLRPLNWNSRYD CKCDYHAQFDNISNKCDECEYLNRGKNGCICHNVTHCQI CHGIPPWEKENLSDFGDFDDANKEQ	
Rep52 AA	MALVNWLVEHGITSEKQWIQENQESYLSFNSTGNSRSQI KAALDNATKIMSLTKSAVDYLVGSSVPEDISKNRIWQIF EMNGYDPAYAGSILYGWCQRSFNKRNTVWLYGPATTGKT NIAEAIAHTVPFYGCVNWTNENFPFNDCVDKMLIWWEEG KMTNKVVESAKAILGGSKVRVDQKCKSSVQIDSTPVIVT SNTNMCVVVDGNSTTFEHQQPLEDRMFKFELTKRLPPDF GKITKQEVKDFFAWAKVNQVPVTHEFKVPRELAGTKGAE KSLKRPLGDVTNTSYKSLEKRARLSFVPETPRSSDVTVD PAPLRPLNWNSRYDCKCDYHAQFDNISNKCDECEYLNRG KNGCICHNVTHCQICHGIPPWEKENLSDFGDFDDANKEQ	1041
Rep40 AA	MSLTKSAVDYLVGSSVPEDISKNRIWQIFEMNGYDPAYA GSILYGWCQRSFNKRNTVWLYGPATTGKTNIAEAIAHTV PFYGCVNWTNENFPFNDCVDKMLIWWEEGKMTNKVVESA KAILGGSKVRVDQKCKSSVQIDSTPVIVTSNTNMCVVVD GNSTTFEHQQPLEDRMFKFELTKRLPPDFGKITKQEVKD FFAWAKVNQVPVTHEFKVPRELAGTKGAEKSLKRPLGDV TNTSYKSLEKRARLSFVPETPRSSDVTVDPAPLRPLNWN SRYDCKCDYHAQFDNISNKCDECEYLNRGKNGCICHNVT HCQICHGIPPWEKENLSDFGDFDDANKEQ	1042

In some embodiments, the molecule from the non-Anellovirus virus, or a molecule based thereon, is introduced into the host cell via a helper construct. In some embodiments, a method described herein comprises introducing a helper construct into a host cell (e.g., a host cell comprising a genetic element construct or a genetic element as described herein). In some embodiments, the helper construct is introduced into the host cell prior to introduction of the genetic element construct. In some embodiments, the helper construct is introduced into the host cell concurrently with the introduction of the genetic element construct. In some embodiments, the helper construct is introduced into the host cell after introduction of the genetic element construct.

In some embodiments, the helper construct comprises a sequence encoding a non-Anellovirus ORF. In some embodiments, the helper construct comprises a sequence encoding a non-Anellovirus Rep molecule, e.g., an AAV Rep molecule, e.g., an AAV Rep protein. In some embodiments, the helper construct comprises a sequence encoding an AAV REP2 molecule. In some embodiments, one or more helper constructs comprise a sequence encoding one or more of (e.g., 1, 2, or all 3 of) an Adenovirus E2A molecule, an Adenovirus E4 molecule, and an Adenovirus VARNA molecule. In embodiments, the AAV Rep molecule, Adenovirus E2A molecule, Adenovirus E4 molecule, and Adenovirus VARNA molecule

are encoded on the same construct. In embodiments, the AAV Rep molecule, Adenovirus E2A molecule, Adenovirus E4 molecule, and Adenovirus VARNA molecule are encoded on different constructs (e.g., at least 2, 3, or 4 separate constructs).

In some embodiments, the helper construct comprises a sequence encoding an Anellovirus ORF (e.g., one or more of an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and/or ORF1/2) , or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

Exemplary cell types

Exemplary host cells suitable for production of anellovectors include, without limitation, mammalian cells, e.g., human cells and insect cells. In some embodiments, the host cell is a human cell or cell line. In some embodiments, the cell is an immune cell or cell line, e.g., a T cell or cell line, a cancer cell line, a hepatic cell or cell line, a neuron, a glial cell, a skin cell, an epithelial cell, a mesenchymal cell, a blood cell, an endothelial cell, an eye cell, a gastrointestinal cell, a progenitor cell, a precursor cell, a stem cell, a lung cell, a cardiac cell, or a muscle cell. In some embodiments, the host cell is an animal cell (e.g., a mouse cell, rat cell, rabbit cell, or hamster cell, or insect cell).

In some embodiments, the host cell is a lymphoid cell. In some embodiments, the host cell is a T cell or an immortalized T cell. In embodiments, the host cell is a Jurkat cell. In embodiments, the host cell is a MOLT cell (e.g., a MOLT-4 or a MOLT-3 cell). In embodiments, the host cell is a MOLT-4 cell. In embodiments, the host cell is a MOLT-3 cell. In some embodiments, the host cell is an acute lymphoblastic leukemia (ALL) cell, e.g., a MOLT cell, e.g., a MOLT-4 or MOLT-3 cell. In some embodiments, the host cell is a B cell or an immortalized B cell. In some embodiments, the host cell comprises a genetic element construct (e.g., as described herein).

In some embodiments, the host cell is a MOLT cell (e.g., a MOLT-4 or a MOLT-3 cell).

In some embodiments, the host cell is an acute lymphoblastic leukemia (ALL) cell, e.g., a MOLT cell, e.g., a MOLT-4 or MOLT-3 cell.

In some embodiments, the host cell is an Expi-293 cell. In some embodiments, the host cell is an Expi-293F cell.

In an aspect, the present disclosure provides a method of manufacturing an anellovector comprising a genetic element enclosed in a proteinaceous exterior, the method comprising providing a MOLT-4 cell comprising an anellovector genetic element, and incubating the MOLT-4 cell under conditions that allow the anellovector genetic element to become enclosed in a proteinaceous exterior in the MOLT-4 cell. In some embodiments, the MOLT-4 cell further comprises one or more Anellovirus proteins (e.g., an Anellovirus ORF1 molecule) that form part or all of the proteinaceous exterior. In some

embodiments, the anellovector genetic element is produced in the MOLT-4 cell, e.g., from a genetic element construct (e.g., as described herein). In some embodiments, the method further comprises introducing the anellovector genetic element construct into the MOLT-4 cell.

In an aspect, the present disclosure provides a method of manufacturing an anellovector comprising a genetic element enclosed in a proteinaceous exterior, the method comprising providing a MOLT-3 cell comprising an anellovector genetic element, and incubating the MOLT-3 cell under conditions that allow the anellovector genetic element to become enclosed in a proteinaceous exterior in the MOLT-3 cell. In some embodiments, the MOLT-3 cell further comprises one or more Anellovirus proteins (e.g., an Anellovirus ORF1 molecule) that form part or all of the proteinaceous exterior. In some embodiments, the anellovector genetic element is produced in the MOLT-3 cell, e.g., from a genetic element construct (e.g., as described herein). In some embodiments, the method further comprises introducing the anellovector genetic element construct into the MOLT-3 cell.

In some embodiments, the host cell is a human cell. In embodiments, the host cell is a HEK293T cell, HEK293F cell, A549 cell, Jurkat cell, Raji cell, Chang cell, HeLa cell Phoenix cell, MRC-5 cell, NCI-H292 cell, or Wi38 cell. In some embodiments, the host cell is a non-human primate cell (e.g., a Vero cell, CV-1 cell, or LLCMK2 cell). In some embodiments, the host cell is a murine cell (e.g., a McCoy cell). In some embodiments, the host cell is a hamster cell (e.g., a CHO cell or BHK 21 cell). In some embodiments, the host cell is a MARC-145, MDBK, RK-13, or EEL cell. In some embodiments, the host cell is an epithelial cell (e.g., a cell line of epithelial lineage).

In some embodiments, the anellovector is cultivated in continuous animal cell line (e.g., immortalized cell lines that can be serially propagated). According to one embodiment of the invention, the cell lines may include porcine cell lines. The cell lines envisaged in the context of the present invention include 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.

Culture Conditions

Host cells comprising a genetic element and components of a proteinaceous exterior can be incubated under conditions suitable for enclosure of the genetic element within the proteinaceous exterior, thereby producing an anellovector. Suitable culture conditions include those described, e.g., in any of Examples 4, 5, 7, 8, 9, 10, 11, or 15. In some embodiments, the host cells are incubated in liquid media (e.g., Grace's Supplemented (TNM-FH), IPL-41, TC-100, Schneider's Drosophila, SF-900 II SFM, or and EXPRESS-FIVE™ SFM). In some embodiments, the host cells are incubated in adherent culture. In some embodiments, the host cells are incubated in suspension culture. In some embodiments, the host cells are incubated in a tube, bottle, microcarrier, or flask. In some embodiments, the host cells are

incubated in a dish or well (e.g., a well on a plate). In some embodiments, the host cells are incubated under conditions suitable for proliferation of the host cells. In some embodiments, the host cells are incubated under conditions suitable for the host cells to release anellovectors produced therein into the surrounding supernatant.

The production of anellovector-containing cell cultures according to the present invention can be carried out in different scales (e.g., in flasks, roller bottles or bioreactors). The media used for the cultivation of the cells to be infected generally comprise the standard nutrients required for cell viability, but may also comprise additional nutrients dependent on the cell type. Optionally, the medium can be protein-free and/or serum-free. Depending on the cell type the cells can be cultured in suspension or on a substrate. In some embodiments, different media is used for growth of the host cells and for production of anellovectors.

Harvest

Anellovectors produced by host cells can be harvested, e.g., according to methods known in the art. For example, anellovectors released into the surrounding supernatant by host cells in culture can be harvested from the supernatant (e.g., as described in Example 4). In some embodiments, the supernatant is separated from the host cells to obtain the anellovectors. In some embodiments, the host cells are lysed before or during harvest. In some embodiments, the anellovectors are harvested from the host cell lysates (e.g., as described in Example 10). In some embodiments, the anellovectors are harvested from both the host cell lysates and the supernatant. In some embodiments, the purification and isolation of anellovectors is performed according to known methods in virus production, for example, as described in Rinaldi, et al., DNA Vaccines: Methods and Protocols (Methods in Molecular Biology), 3rd ed. 2014, Humana Press (incorporated herein by reference in its entirety). In some embodiments, the anellovector may be harvested and/or purified by separation of solutes based on biophysical properties, e.g., ion exchange chromatography or tangential flow filtration, prior to formulation with a pharmaceutical excipient.

In vitro assembly methods

An anellovector may be produced, e.g., by *in vitro* assembly, e.g., in a cell-free suspension or in a supernatant. In some embodiments, the genetic element is contacted to an ORF1 molecule *in vitro,* e.g., under conditions that allow for assembly.

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.

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 encapisdated 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.

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.

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.

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.

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) 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 (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

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 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.

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

In some aspects, the invention described herein comprises compositions and methods of using and making an anellovector, anellovector preparations, and therapeutic compositions. In some embodiments, the anellovectors are made using compositions and methods as described herein. 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 and/or an exogenous nucleic acid sequence from a virus other than an Anellovirus (e.g., a Monodnavirus, e.g., a Shotokuvirus (e.g., a Cressdnaviricota [e.g., a redondovirus, circovirus {e.g., a porcine circovirus, e.g., PCV-1 or PCV-2; or beak-and-feather disease virus}, geminivirus {e.g., tomato golden mosaic virus}, or nanovirus {e.g., BBTV, MDV1, SCSVF, or FBNYV}]), or a Parvovirus (e.g., a dependoparavirus, e.g., a bocavirus or an AAV)). 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 heterologous 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.

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.

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.

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.

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.

In one aspect, the invention includes an anellovector comprising:

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

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

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. Parvoviridiae).

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.

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-1α promoter, an SV40 promoter, a CAGG promoter, or a UBC promoter, TTV viral promoters, Tissue specific, U6 (pollIII), 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.

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 2.8-4kb, about 2.8-3.2kb, about 3.6-3.9kb, or

about 2.8-2.9kb), less than about 5kb (e.g., less than about 2.9kb, 3.2 kb, 3.6kb, 3.9kb, or 4kb), or at least 100 nucleotides (e.g., at least 1kb).

The anellovectors, compositions comprising anellovectors, methods using such anellovectors, etc., as described herein are, in some instances, based in part on the examples which illustrate how different effectors, for example miRNAs (e.g. against IFN or miR-625), shRNA, etc and protein binding sequences, for example DNA sequences that bind to capsid protein such as Q99153, are combined with proteinaceious exteriors, for example a capsid disclosed in Arch Virol (2007) 152: 1961-1975, to produce anellovectors which can then be used to deliver an effector to cells (e.g., animal cells, e.g., human cells or non-human animal cells such as pig or mouse cells). In embodiments, the effector can silence expression of a factor such as an interferon. The examples further describe how anellovectors can be made by inserting effectors into sequences derived, e.g., from an *Anellovirus*. It is on the basis of these examples that the description hereinafter contemplates various variations of the specific findings and combinations considered in the examples. For example, the skilled person will understand from the examples that the specific miRNAs are used just as an example of an effector and that other effectors may be, e.g., other regulatory nucleic acids or therapeutic peptides. Similarly, the specific capsids used in the examples may be replaced by substantially non-pathogenic proteins described hereinafter. The specifc *Anellovirus* sequences described in the examples may also be replaced by the *Anellovirus* sequences described hereinafter. These considerations similarly apply to protein binding sequences, regulatory sequences such as promoters, and the like. Independent thereof, the person skilled in the art will in particular consider such embodiments which are closely related to the examples.

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).

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).

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., using a helper, e.g., a helper virus or helper plasmid, or 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*.

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).

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

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. *Anelloviruses* have not generally been linked to any human disease. However, attempts to link *Anellovirus* infection with human disease are confounded by the high incidence of asymptomatic *Anellovirus* viremia in control cohort population(s), the remarkable genomic diversity within the anellovirus viral family, the historical inability to propagate the agent in vitro, and the lack of animal model(s) of *Anellovirus* disease (Yzebe et al., Panminerva Med. (2002) 44:167-177; Biagini, P., Vet. Microbiol. (2004) 98:95-101).

Anelloviruses are generally transmitted by oronasal or fecal-oral infection, mother-to-infant and/or in utero transmission (Gerner et al., Ped. Infect. Dis. J. (2000) 19:1074-1077). Infected persons can, in some instances, be characterized by a prolonged (months to years) *Anellovirus* viremia. Humans may be co-infected with more than one genogroup or strain (Saback, et al., Scad. J. Infect. Dis. (2001) 33:121-125). There is a suggestion that these genogroups can recombine within infected humans (Rey et

al., Infect. (2003) 31:226-233). The double stranded isoform (replicative) intermediates have been found in several tissues, such as liver, peripheral blood mononuclear cells and bone marrow (Kikuchi et al., J. Med. Virol. (2000) 61:165-170; Okamoto et al., Biochem. Biophys. Res. Commun. (2002) 270:657-662; Rodriguez-lnigo et al., Am. J. Pathol. (2000) 156:1227-1234).

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.

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 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.

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* 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.

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* TATA box 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* initiator element 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* transcriptional start site 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* 5' UTR conserved domain 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* three open-reading frame region 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* poly(A) signal 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* GC-rich 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 nucleic acid sequence of Table B3. 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 B3. 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 B3. 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 B3. 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 B3. 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 B3. 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 B3. 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 B3. 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 B3. 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 B3. 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 B3. 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 B3. 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 B3.

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 C1. 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 C1. 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 C1. 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 C1. 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 C1. 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 C1. 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 C1. 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 C1. 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 C1. 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 C1. 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 C1. 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 C1. 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 C1. 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 C1.

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* 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.

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 F1. 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 F1. 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 F1. 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 F1. 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 F1. 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 F1. 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 F1. 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 F1. 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 F1. 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 F1. 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 F1. 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 F1. 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 F1.

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 F3. 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 F3. 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 F3. 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 F3. 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 F3. 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 F3. 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 F3. 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 F3. 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 F3. 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 F3. 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 F3. 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 F3. 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 F3.

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 F5. 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 F5. 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 F5. 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 F5. 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 F5. 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 F5. 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 F5. 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 F5. 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 F5. 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 F5. 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 F5. 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 F5. 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 F5.

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.

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 Tables A2-M2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

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).

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*.

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).

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).

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 a chimeric 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, 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.

Additional exemplary Anellovirus genomes, for which sequences or subsequences comprised therein can be utilized in the compositions and methods described herein (e.g., to form a genetic element of an anellovector, e.g., as 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.

5

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

Name	Ring1
Genus/Clade	*Alphatorquevirus*, Clade 3
Accession Number	AJ620231.1

Full Sequence: 3753 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TGCTACGTCACTAACCCACGTGTCCTCTACAGGCCAATCGCAGTCTATGT
CGTGCACTTCCTGGGCATGGTCTACATAATTATATAAATGCTTGCACTTC
CGAATGGCTGAGTTTTTGCTGCCCGTCCGCGGAGAGGAGCCACGGCAGGG
GATCCGAACGTCCTGAGGGCGGGTGCCGGAGGTGAGTTTACACACCGAAG
TCAAGGGGCAATTCGGGCTCAGGACTGGCCGGGCTTTGGGCAAGGCTCTT
AAAAATGCACTTTTCTCGAATAAGCAGAAAGAAAAGGAAAGTGCTACTGC
TTTGCGTGCCAGCAGCTAAGAAAAAACCAACTGCTATGAGCTTCTGGAAA
CCTCCGGTACACAATGTCACGGGGATCCAACGCATGTGGTATGAGTCCTT
TCACCGTGGCCACGCTTCTTTTTGTGGTTGTGGGAATCCTATACTTCACA
TTACTGCACTTGCTGAAACATATGGCCATCCAACAGGCCCGAGACCTTCT
GGGCCACCGGGAGTAGACCCCAACCCCCACATCCGTAGAGCCAGGCCTGC
CCCGGCCGCTCCGGAGCCCTCACAGGTTGATTCGAGACCAGCCCTGACAT
GGCATGGGGATGGTGGAAGCGACGGAGGCGCTGGTGGTTCCGGAAGCGGT
GGACCCGTGGCAGACTTCGCAGACGATGGCCTCGATCAGCTCGTCGCCGC
CCTAGACGACGAAGAGTAAGGAGGCGCAGACGGTGGAGGAGGGGGAGACG
AAAAACAAGGACTTACAGACGCAGGAGACGCTTTAGACGCAGGGGACGAA
AAGCAAAACTTATAATAAAACTGTGGCAACCTGCAGTAATTAAAAGATGC
AGAATAAAGGGATACATACCACTGATTATAAGTGGGAACGGTACCTTTGC
CACAAACTTTACCAGTCACATAAATGACAGAATAATGAAAGGCCCCTTCG
GGGGAGGACACAGCACTATGAGGTTCAGCCTCTACATTTTGTTTGAGGAG
CACCTCAGACACATGAACTTCTGGACCAGAAGCAACGATAACCTAGAGCT
AACCAGATACTTGGGGGCTTCAGTAAAAATATACAGGCACCCAGACCAAG
ACTTTATAGTAATATACAACAGAAGAACCCCTCTAGGAGGCAACATCTAC
ACAGCACCCTCTCTACACCCAGGCAATGCCATTTTAGCAAAACACAAAAT
```

ATTAGTACCAAGTTTACAGACAAGACCAAAGGGTAGAAAAGCAATTAGAC

TAAGAATAGCACCCCCCACACTCTTTACAGACAAGTGGTACTTTCAAAAG

GACATAGCCGACCTCACCCTTTTCAACATCATGGCAGTTGAGGCTGACTT

GCGGTTTCCGTTCTGCTCACCACAAACTGACAACACTTGCATCAGCTTCC

AGGTCCTTAGTTCCGTTTACAACAACTACCTCAGTATTAATACCTTTAAT

AATGACAACTCAGACTCAAAGTTAAAAGAATTTTTAAATAAAGCATTTCC

AACAACAGGCACAAAAGGAACAAGTTTAAATGCACTAAATACATTTAGAA

CAGAAGGATGCATAAGTCACCCACAACTAAAAAAACCAAACCCACAAATA

AACAAACCATTAGAGTCACAATACTTTGCACCTTTAGATGCCCTCTGGGG

AGACCCCATATACTATAATGATCTAAATGAAAACAAAAGTTTGAACGATA

TCATTGAGAAAATACTAATAAAAAACATGATTACATACCATGCAAAACTA

AGAGAATTTCCAAATTCATACCAAGGAAACAAGGCCTTTTGCCACCTAAC

AGGCATATACAGCCCACCATACCTAAACCAAGGCAGAATATCTCCAGAAA

TATTTGGACTGTACACAGAAATAATTTACAACCCTTACACAGACAAAGGA

ACTGGAAACAAAGTATGGATGGACCCACTAACTAAAGAGAACAACATATA

TAAAGAAGGACAGAGCAAATGCCTACTGACTGACATGCCCCTATGGACTT

TACTTTTTGGATATACAGACTGGTGTAAAAAGGACACTAATAACTGGGAC

TTACCACTAAACTACAGACTAGTACTAATATGCCCTTATACCTTTCCAAA

ATTGTACAATGAAAAAGTAAAAGACTATGGGTACATCCCGTACTCCTACA

AATTCGGAGCGGGTCAGATGCCAGACGGCAGCAACTACATACCCTTTCAG

TTTAGAGCAAAGTGGTACCCCACAGTACTACACCAGCAACAGGTAATGGA

GGACATAAGCAGGAGCGGGCCCTTTGCACCTAAGGTAGAAAAACCAAGCA

CTCAGCTGGTAATGAAGTACTGTTTTAACTTTAACTGGGGCGGTAACCCT

ATCATTGAACAGATTGTTAAAGACCCCAGCTTCCAGCCCACCTATGAAAT

ACCCGGTACCGGTAACATCCCTAGAAGAATACAAGTCATCGACCCGCGGG

TCCTGGGACCGCACTACTCGTTCCGGTCATGGGACATGCGCAGACACACA

TTTAGCAGAGCAAGTATTAAGAGAGTGTCAGAACAACAAGAAACTTCTGA

CCTTGTATTCTCAGGCCCAAAAAAGCCTCGGGTCGACATCCCAAAACAAG

AAACCCAAGAAGAAAGCTCACATTCACTCCAAAGAGAATCGAGACCGTGG

GAGACCGAGGAAGAAAGCGAGACAGAAGCCCTCTCGCAAGAGAGCCAAGA

GGTCCCCTTCCAACAGCAGTTGCAGCAGCAGTACCAAGAGCAGCTCAAGC

TCAGACAGGGAATCAAAGTCCTCTTCGAGCAGCTCATAAGGACCCAACAA

GGGGTCCATGTAAACCCATGCCTACGGTAGGTCCCAGGCAGTGGCTGTTT

CCAGAGAGAAAGCCAGCCCCAGCTCCTAGCAGTGGAGACTGGGCCATGGA

GTTTCTCGCAGCAAAAATATTTGATAGGCCAGTTAGAAGCAACCTTAAAG

ATACCCCTTACTACCCATATGTTAAAAACCAATACAATGTCTACTTTGAC

CTTAAATTTGAATAAACAGCAGCTTCAAACTTGCAAGGCCGTGGGAGTTT

CACTGGTCGGTGTCTACCTCTAAAGGTCACTAAGCACTCCGAGCGTAAGC

GAGGAGTGCGACCCTCCCCCCTGGAACAACTTCTTCGGAGTCCGGCGCTA
CGCCTTCGGCTGCGCCGGACACCTCAGACCCCCCCTCCACCCGAAACGCT
TGCGCGTTTCGGACCTTCGGCGTCGGGGGGGTCGGGAGCTTTATTAAACG
GACTCCGAAGTGCTCTTGGACACTGAGGGGGTGAACAGCAACGAAAGTGA
GTGGGGCCAGACTTCGCCATAAGGCCTTTATCTTCTTGCCATTTGTCAGT
GTCCGGGGTCGCCATAGGCTTCGGGCTCGTTTTTAGGCCTTCCGGACTAC
AAAAATCGCCATTTTGGTGACGTCACGGCCGCCATCTTAAGTAGTTGAGG
CGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACACCTACTCAAAA
TGGTGGACAATTTCTTCCGGGTCAAAGGTTACAGCCGCCATGTTAAAACA
CGTGACGTATGACGTCACGGCCGCCATTTTGTGACACAAGATGGCCGACT
TCCTTCCTCTTTTTCAAAAAAAAGCGGAAGTGCCGCCGCGGCGGCGGGGG
GCGGCGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGCGCCCCCCCCC
GCGCATGCGCGGGGCCCCCCCCCGCGGGGGGCTCCGCCCCCCGGCCCCCC
CCG (SEQ ID NO: 16)

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
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	MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAPAAPEPSQVDSRPALTWHGDGGSDGGAGGSGSGGPVADFA DDGLDQLVAALDDEE (SEQ ID NO: 17)
ORF2/2	MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAPAAPEPSQVDSRPALTWHGDGGSDGGAGGSGSGGPVADFA DDGLDQLVAALDDEELLKTPASSPPMKYPVPVTSLEEYKSSTRGSWDRTTRSGHGT CADTHLAEQVLRECQNNKKLLTLYSQAQKSLGSTSQNKKPKKKAHIHSKENRDRG RPRKKARQKPSRKRAKRSPSNSSCSSSTKSSSSSDRESKSSSSSS (SEQ ID NO: 18)
ORF2/3	MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAPAAPEPSQVDSRPALTWHGDGGSDGGAGGSGSGGPVADFA DDGLDQLVAALDDEEPKKASGRHPKTRNPRRKLTFTPKRIETVGDRGRKRDRSPLA REPRGPLPTAVAAAVPRAAQAQTGNQSPLRAAHKDPTRGPCKPMPTVGPRQWLFP ERKPAPAPSSGDWAMEFLAAKIFDRPVRSNLKDTPYYPYVKNQYNVYFDLKFE (SEQ ID NO: 19)
ORF2t/3	MSFWKPPVHNVTGIQRMWPKKASGRHPKTRNPRRKLTFTPKRIETVGDRGRKRDR SPLAREPRGPLPTAVAAAVPRAAQAQTGNQSPLRAAHKDPTRGPCKPMPTVGPRQ WLFPERKPAPAPSSGDWAMEFLAAKIFDRPVRSNLKDTPYYPYVKNQYNVYFDLK FE (SEQ ID NO: 20)
ORF1	MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRWRRGRRK TRTYRRRRRFRRRGRKAKLIIKLWQPAVIKRCRIKGYIPLIISGNGTFATNFTSHINDR IMKGPFGGGHSTMRFSLYILFEEHLRHMNFWTRSNDNLELTRYLGASVKIYRHPDQ DFIVIYNRRTPLGGNIYTAPSLHPGNAILAKHKILVPSLQTRPKGRKAIRLRIAPPTLFT DKWYFQKDIADLTLFNIMAVEADLRFPFCSPQTDNTCISFQVLSSVYNNYLSINTFN NDNSDSKLKEFLNKAFPTTGTKGTSLNALNTFRTEGCISHPQLKKPNPQINKPLESQ YFAPLDALWGDPIYYNDLNENKSLNDIIEKILIKNMITYHAKLREFPNSYQGNKAFC HLTGIYSPPYLNQGRISPEIFGLYTEIIYNPYTDKGTGNKVWMDPLTKENNIYKEGQS KCLLTDMPLWTLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFPKLYNEKVKDY GYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVMEDISRSGPFAPKVEKP STQLVMKYCFNFNWGGNPIIEQIVKDPSFQPTYEIPGTGNIPRRIQVIDPRVLGPHYSF RSWDMRRHTFSRASIKRVSEQQETSDLVFSGPKKPRVDIPKQETQEESSHSLQRESR PWETEEESETEALSQESQEVPFQQQLQQQYQEQLKLRQGIKVLFEQLIRTQQGVHV NPCLR (SEQ ID NO: 21)

ORF1/1	MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRIVKDPSFQPTYEIPG TGNIPRRIQVIDPRVLGPHYSFRSWDMRRHTFSRASIKRVSEQQETSDLVFSGPKKPR VDIPKQETQEESSHSLQRESRPWETEEESETEALSQESQEVPFQQQLQQQYQEQLKL RQGIKVLFEQLIRTQQGVHVNPCLR (SEQ ID NO: 22)
ORF1/2	MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRAQKSLGSTSQNKK PKKKAHIHSKENRDRGRPRKKARQKPSRKRAKRSPSNSSCSSSTKSSSSSDRESKSSS SSS (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
          |         |          |          |          |          |
          TAATAAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAA
          CCGTCACTTAGTTCCCCTTTTTGCAACAACTTCTGCTTTTTTCCAACTGC
          CGGAAAACCACATAATTTGCATGGCTAACCACAAACTGATATGCTAATTA
          ACTTCCACAAAACAACTTCCCCTTTTAAAACCACACCTACAAATTAATTA
          TTAAACACAGTCACATCCTGGGAGGTACTACCACACTATAATACCAAGTG
          CACTTCCGAATGGCTGAGTTTATGCCGCTAGACGGAGAACGCATCAGTTA
          CTGACTGCGGACTGAACTTGGGCGGGTGCCGAAGGTGAGTGAAACCACCG
          AAGTCAAGGGGCAATTCGGGCTAGTTCAGTCTAGCGGAACGGGCAAGAAA
          CTTAAAATTATTTTATTTTTCAGATGAGCGACTGCTTTAAACCAACATGC
          TACAACAACAAAACAAAGCAAACTCACTGGATTAATAACCTGCATTTAAC
          CCACGACCTGATCTGCTTCTGCCCAACACCAACTAGACACTTATTACTAG
          CTTTAGCAGAACAACAAGAAACAATTGAAGTGTCTAAACAAGAAAAAGAA
          AAAATAACAAGATGCCTTATTACTACAGAAGAAGACGGTACAACTACAGA
          CGTCCTAGATGGTATGGACGAGGTTGGATTAGACGCCCTTTTCGCAGAAG
          ATTTCGAAGAAAAAGAAGGGTAAGACCTACTTATACTACTATTCCTCTAA
          AGCAATGGCAACCGCCATATAAAAGAACATGCTATATAAAAGGACAAGAC
          TGTTTAATATACTATAGCAACTTAAGACTGGGAATGAATAGTACAATGTA
          TGAAAAAAGTATTGTACCTGTACATTGGCCGGGAGGGGGTTCTTTTTCTG
          TAAGCATGTTAACTTTAGATGCCTTGTATGATATACATAAACTTTGTAGA
          AACTGGTGGACATCCACAAACCAAGACTTACCACTAGTAAGATATAAAGG
```

ATGCAAAATAACATTTTATCAAAGCACATTTACAGACTACATAGTAAGAA

TACATACAGAACTACCAGCTAACAGTAACAAACTAACATACCCAAACACA

CATCCACTAATGATGATGATGTCTAAGTACAAACACATTATACCTAGTAG

ACAAACAAGAAGAAAAAAGAAACCATACACAAAAATATTTGTAAAACCAC

CTCCGCAATTTGAAAACAAATGGTACTTTGCTACAGACCTCTACAAAATT

CCATTACTACAAATACACTGCACAGCATGCAACTTACAAAACCCATTTGT

AAAACCAGACAAATTATCAAACAATGTTACATTATGGTCACTAAACACCA

TAAGCATACAAAATAGAAACATGTCAGTGGATCAAGGACAATCATGGCCA

TTTAAAATACTAGGAACACAAAGCTTTTATTTTTACTTTTACACCGGAGC

AAACCTACCAGGTGACACAACACAAATACCAGTAGCAGACCTATTACCAC

TAACAAACCCAAGAATAAACAGACCAGGACAATCACTAAATGAGGCAAAA

ATTACAGACCATATTACTTTCACAGAATACAAAAACAAATTTACAAATTA

TTGGGGTAACCCATTTAATAAACACATTCAAGAACACCTAGATATGATAC

TATACTCACTAAAAAGTCCAGAAGCAATAAAAAACGAATGGACAACAGAA

AACATGAAATGGAACCAATTAAACAATGCAGGAACAATGGCATTAACACC

ATTTAACGAGCCAATATTCACACAAATACAATATAACCCAGATAGAGACA

CAGGAGAAGACACTCAATTATACCTACTCTCTAACGCTACAGGAACAGGA

TGGGACCCACCAGGAATTCCAGAATTAATACTAGAAGGATTTCCACTATG

GTTAATATATTGGGGATTTGCAGACTTTCAAAAAAACCTAAAAAAAGTAA

CAAACATAGACACAAATTACATGTTAGTAGCAAAAACAAAATTTACACAA

AAACCTGGCACATTCTACTTAGTAATACTAAATGACACCTTTGTAGAAGG

CAATAGCCCATATGAAAAACAACCTTTACCTGAAGACAACATTAAATGGT

ACCCACAAGTACAATACCAATTAGAAGCACAAAACAAACTACTACAAACT

GGGCCATTTACACCAAACATACAAGGACAACTATCAGACAATATATCAAT

GTTTTATAAATTTTACTTTAAATGGGGAGGAAGCCCACCAAAAGCAATTA

ATGTTGAAAATCCTGCCCACCAGATTCAATATCCCATACCCCGTAACGAG

CATGAAACAACTTCGTTACAGAGTCCAGGGGAAGCCCCAGAATCCATCTT

ATACTCCTTCGACTATAGACACGGGAACTACACAACAACAGCTTTGTCAC

GAATTAGCCAAGACTGGGCACTTAAAGACACTGTTTCTAAAATTACAGAG

CCAGATCGACAGCAACTGCTCAAACAAGCCCTCGAATGCCTGCAAATCTC

GGAAGAAACGCAGGAGAAAAAAGAAAAAGAAGTACAGCAGCTCATCAGCA

ACCTCAGACAGCAGCAGCAGCTGTACAGAGAGCGAATAATATCATTATTA

AAGGACCAATAACTTTTAACTGTGTAAAAAAGGTGAAATTGTTTGATGAT

AAACCAAAAAAACCGTAGATTTACACCTGAGGAATTTGAAACTGAGTTACA

AATAGCAAAATGGTTAAAGAGACCCCCAAGATCCTTTGTAAATGATCCTC

CCTTTTACCCATGGTTACCACCTGAACCTGTTGTAAACTTTAAGCTTAAT

TTTACTGAATAAAGGCCAGCATTAATTCACTTAAGGAGTCTGTTTATTTA

AGTTAAACCTTAATAAACGGTCACCGCCTCCCTAATACGCAGGCGCAGAA

AGGGGGCTCCGCCCCCTTTAACCCCCAGGGGGCTCCGCCCCCTGAAACCC
CCAAGGGGGCTACGCCCCCTTACACCCCC (SEQ ID NO: 54)

Annotations:

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 (*Betatorquevirus*)

Ring2 (*Betatorquevirus*)	
ORF2	MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPTRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTDVLDGMDEVGLDALFAEDFEEKEG (SEQ ID NO: 55)
ORF2/2	MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPTRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTDVLDGMDEVGLDALFAEDFEEKEGFNIPYPVTSMKQLRY RVQGKPQNPSYTPSTIDTGTTQQQLCHELAKTGHLKTLFLKLQSQIDSNCSNKPSNA CKSRKKRRRKKKKKYSSSSATSDSSSSCTESE (SEQ ID NO: 56)
ORF2/3	MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPTRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTDVLDGMDEVGLDALFAEDFEEKEGARSTATAQTSPRMP ANLGRNAGEKRKRSTAAHQQPQTAAAAVQRANNIIIKGPITFNCVKKVKLFDDKPK NRRFTPEEFETELQIAKWLKRPPRSFVNDPPFYPWLPPEPVVNFKLNFTE (SEQ ID NO: 57)
ORF1	MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYTTIPLKQWQPPYKR TCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKL

	CRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLM MMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACN LQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTGA NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNK HIQEHLDMILYSLKSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNP DRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFADFQKNLKKVTNID TNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEA QNKLLQTGPFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNE HETTSLQSPGEAPESILYSFDYRHGNYTTTALSRISQDWALKDTVSKITEPDRQQLLK QALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 58)
ORF1/1	MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRIQYPIPRNEHETTSLQSPGE APESILYSFDYRHGNYTTTALSRISQDWALKDTVSKITEPDRQQLLKQALECLQISEE TQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 59)
ORF1/2	MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRSQIDSNCSNKPSNACKSRK KRRRKKKKKYSSSSATSDSSSSCTESE (SEQ ID NO: 60)

Table B3. Exemplary *Anellovirus* nucleic acid sequence (*Gammatorquevirus*)

Name Ring3.1

Genus/Clade *Gammatorquevirus*

Accession Number

Full Sequence: 3264 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAAAATGGCGGCAACCAATCATTTTATACTTTCACTTTCCAATTACAAGC
CGCCACGTCACAGAACAGGGGTGGAGACTTTAAAACTATATAACCAAGTG
ATGTGACGAATGGCTGAGTTTACCCCGCTAGACGGTGCAGGGACCGGATC
GAGCGCAGCGAGGAGGTCCCCGGCTGCCCGTGGGCGGGAGCCCGAGGTGA
GTGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCTAGCGGA
ACGGGCAAGAAACTTAAAATATGTTTTGTTTCAGATGCAGACACCTGCTT
CACAGATAAGCTCAGACGACTTCTTTGTACACACTCCATTTAATGCAGTA
ACTAAACAGCAAATATGGATGTCTCAAATTGCTGATGGACATGACAACAT
TTGTCACTGCCACCGTCCTTTTGCTCACCTGCTTGCTAATATTTTTCCTC
CTGGTCATAAAGACAGGGATCTTACCATTAATCAAATACTTGCTAGAGAT
CTTACAGAAACATGCCATTCTGGTGGAGACGAAGGAACAAGCGGTGGTGG
GGTCGCCGCTTCCGCTACCGCCGCTACAACAAATATAAAACCAGAAGGAG
ACGCAGAATACCCAGAAGACGAAATAGAAGATTTACTAAGACACGCAGGA
GAAGAAAAAGAAAGAAGGTAAGAAGAAAACTTAAAAAAATTACTATTAAA
CAATGGCAGCCAGATTCAGTGAAAAAATGTAAAATTAAAGGATATAGTAC
TTTAGTTATGGGTGCACAAGGAAAACAATACAACTGTTACACAAACCAAG
```

CAAGTGACTATGTTCAGCCTAAAGCACCACAAGGTGGGGGCTTTGGCTGT
GAAGTATTTAATTTAAAATGGCTATACCAAGAATATACTGCACACAGAAA
TATTTGGACAAAAACAAATGAATATACAGACCTTTGTAGATACACTGGAG
CTCAAATAATTTTATACAGGCACCCAGATGTTGATTTTATAGTCAGCTGG
GACAATCAGCCACCTTTTTTACTTAACAAATATACATATCCAGAACTGCA
ACCACAAAACCTTTTACTAGCTAGAAGGAAAAGAATTATTCTTAGTCAAA
AATCAAACCCCAAAGGAAAACTAAGAATTAAACTAAGAATACCACCACCA
AAACAAATGATAACAAAATGGTTTTTTCAAAGAGACTTTTGTGATGTGAA
TCTGTTTAAACTATGTGCTTCTGCTGCTTCTTTCCGCTACCCAGGTATCA
GTCATGGAGCTCAAAGTACTATTTTTTCTGCATATGCTTTAAACACTGAC
TTTTATCAATGCAGTGACTGGTGCCAAACTAACACAGAAACTGGCTACCT
AAACATTAAAACACAACAAATGCCACTATGGTTTCATTACAGAGAGGGTG
GCAAAGAGAAATGGTATAAATACACCAACAAAGAACACAGACCATATACA
AATACATATCTTAAAAGTATTAGCTATAATGATGGATTGTTTTCTCCTAA
AGCCATGTTTGCATTTGAAGTAAAAGCGGGGGGGTGAAGGAACAACAGAAC
CACCACAAGGCGCCCAATTAATTGCTAACCTTCCACTCATTGCACTAAGA
TATAATCCACATGAAGACACAGGCCATGGCAATGAAATTTACCTTACATC
AACTTTTAAAGGTACATATGACAAACCTAAAGTTACTGATGCTCTATACT
TTAACAATGTACCCCTGTGGATGGGATTTTATGGCTACTGGGACTTTATA
TTACAAGAAACAAAAAAACAAAGGTGTCTTTGATCAACATATGTTTGTTGT
TAAATGTCCTGCCTTAAGGCCCATATCACAAGTCACAAAACAAGTATACT
ACCCACTTGTAGACATGGACTTTTGTTCAGGGAGACTGCCATTTGATGAA
TATTTATCCAAAGACATTAAAAGTCATTGGTATCCCACTGCAGAAAGACA
AACAGTTACAATAAATAATTTTGTTACAGCAGGTCCATACATGCCTAAAT
TTGAACCCACAGACAAAGACAGTACATGGCAATTAAACTATCACTATAAA
TTTTTTTTTAAGTGGGGTGGTCCACAAGTCACAGACCCAACTGTTGAAGA
CCCATGCAGCAGAAACAAATATCCTGTCCCCGATACAATGCAACAAACAA
TACAAATTAAAAACCCTGAAAAGCTGCACCCAGCAACCCTCTTCCATGAC
TGGGACCTTAGAAGGGGCTTCATTACACAAGCAGCTATTAAAAGAATGTC
AGAAAACCTCCAAATTGATTCATCTTTCGAATCTGATGGCACAGAATCAC
CCAAAAAAAAGAAAAGATGCACCAAAGAAATCCCAACACAAAACCAAAAG
CAAGAAGAGATCCAAGAATGTCTCCTCTCACTCTGCGAAGAGCCTACATG
CCAAGAAGAAACAGAGGACCTCCAGCTCTTCATCCAGCAGCAGCAGCAGC
AGCAGTACAAGCTCAGAAAAAACCTCTTCAAACTCCTCACTCACCTGAAA
AAAGGACAGAGAATAAGTCAACTACAAACGGGACTTTTAGAGTAATACCA
TTTAAACCAGGTTTTGAACAAGAAACAGAAAAAGAACTTGCCATAGCTTT
CTGCAGACCACCTAGAAAATATAAAAATGATCCCCCTTTTTATCCCTGGT
TACCATGGACACCCCTTGTACACTTTAACCTTAATTACAAAGGCTAGGCC
AACACTGTTCACTTAGTGGTGTATGTTTAATAAAGTTTCACCCCCAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAATTGCAAAAATTCG
GCGCTCGCGCGCGCTGCGCGCGCGCGAGCGCCGTCACGCGCCGGCGCTCG
CGCGCCGCGCGTATGTGCTAACACACCACGCACCTAGATTGGGGTGCGCG
CGCTAGCGCGCGCACCCCAATGCGCCCCGCCCTCGTTCCGACCCGCTTGC
GCGGGTCGGACCACTTCGGGCTCGGGGGGGCGCGCCTGCGGCGCTTTTTT
ACTAAACAGACTCCGAGCCGCCATTTGGCCCCCCCTAAGCTCCGCCCCCC
TCATGAATATTCATAAAGGAAACCACATAATTAGAATTGCCGACCACAAA
CTGCCATATGCTAATTAGTTCCCCTTTTACACAGTAAAAAGGGGAAGTGG
GGGGGCATAGCCCCCCCACACCCCCCGCGGGGGGGGCAGAGCCCCCCCCC
GCACCCCCCCCCCTACGTCACAATCCACGCCCCCGCCGCCATCTTGGGTGC
GGCAGGGCGGGGGC (SEQ ID NO: 878)

Annotations:

Putative Domain	*Base range*
TATA Box	87– 93

Cap Site	110 – 117
Transcriptional Start Site	117
5’ UTR Conserved Domain	185 – 255
ORF2	285 – 671
ORF2/2	285 – 667 ; 2063 – 2498
ORF2/3	285 – 667 ; 2295 – 2697
TAIP	385 - 585
ORF1	512 – 2545
ORF1/1	512 – 667 ; 2063 – 2545
ORF1/2	512 – 667 ; 2295 – 2498
Three open-reading frame region	2295 – 2495
Poly(A) Signal	2729 – 2734
GC-rich region	3141 – 3264

Table B4. Exemplary *Anellovirus* amino acid sequences (*Gammatorquevirus*)

Ring 3.1 (*Gammatorquevirus*)	
ORF2	MQTPASQISSDDFFVHTPFNAVTKQQIWMSQIADGHDNICHCHRPFAHLLAN IFPPGHKDRDLTINQ ILARDLTETCHSGGDEGTSGGGVAASATAATTNIKPEGDAEYPEDEIEDLLR HAGEEKERR (SEQ ID NO: 879)
ORF2/2	MQTPASQISSDDFFVHTPFNAVTKQQIWMSQIADGHDNICHCHRPFAHLLAN IFPPGHKDRDLTINQ ILARDLTETCHSGGDEGTSGGGVAASATAATTNIKPEGDAEYPEDEIEDLLR HAGEEKERSGVVHKSQTQLLKTHAAETNILSPIQCNKQYKLKTLKSCTQQPS SMTGTLEGASLHKQLLKECQKTSKLIHLSNLMAQNHPKKRKDAPKKSQHK TKSKKRSKNVSSHSAKSLHAKKKQRTSSSSSSSSSSSTSSEKTSSNSSLT (SEQ ID NO: 880)
ORF2/3	MQTPASQISSDDFFVHTPFNAVTKQQIWMSQIADGHDNICHCHRPFAHLLAN IFPPGHKDRDLTINQ ILARDLTETCHSGGDEGTSGGGVAASATAATTNIKPEGDAEYPEDEIEDLLR HAGEEKERRITQKKEKMHQRNPNTKPKARRDPRMSPLTLRRAYMPRRNRG

	PPALHPAAAAAAAVQAQKKPLQTPHSPEKRTENKSTTNGTFRVIPFKPGFEQE TEKELAIAFCRPPRKYKNDPPFYPWLPWTPLVHFNLNYKG (SEQ ID NO: 881)
TAIP	MDMTTFVTATVLLLTCLLIFFLLVIKTGILPLIKYLLEILQKHAILVETKEQAV VGSPLPLPPLQQI (SEQ ID NO: 882)
ORF1	MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRRRRK RKKVRRKLKKITIKQWQP DSVKKCKIKGYSTLVMGAQGKQYNCYTNQASDYVQPKAPQGGGFGCEVF NLKWLYQEYTAHRNIWTKTNEYTDLCRYTGAQIILYRHPDVDFIVSWDNQP PFLLNKYTYPELQPQNLLLARRKRIILSQKSNPKGKLRIKLRIPPPKQMITKWF FQRDFCDVNLFKLCASAASFRYPGISHGAQSTIFSAYALNTDFYQCSDWCQT NTETGYLNIKTQQMPLWFHYREGGKEKWYKYTNKEHRPYTNTYLKSISYN DGLFSPKAMFAFEVKAGGEGTTEPPQGAQLIANLPLIALRYNPHEDTGHGNE IYLTSTFKGTYDKPKVTDALYFNNVPLWMGFYGYWDFILQETKNKGVFDQ HMFVVKCPALRPISQVTKQVYYPLVDMDFCSGRLPFDEYLSKDIKSHWYPT AERQTVTINNFVTAGPYMPKFEPTDKDSTWQLNYHYKFFFKWGGPQVTDP TVEDPCSRNKYPVPDTMQQTIQIKNPEKLHPATLFHDWDLRRGFITQAAIKR MSENLQIDSSFESDGTESPKKKKRCTKEIPTQNQKQEEIQECLLSLCEEPTCQ EETEDLQLFIQQQQQQQYKLRKNLFKLLTHLKKGQRISQLQTGLLE (SEQ ID NO: 883)
ORF1/1	MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRRRRK RKKWGGPQVTDPTVEDPCSRNKYPVPDTMQQTIQIKNPEKLHPATLFHDWD LRRGFITQAAIKRMSENLQIDSSFESDGTESPKKKKRCTKEIPTQNQKQEEIQE CLLSLCEEPTCQEETEDLQLFIQQQQQQQYKLRKNLFKLLTHLKKGQRISQL QTGLLE (SEQ ID NO: 884)
ORF1/2	MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRRRRK RKKNHPKKRKDAPKKSQHKTKSKKRSKNVSSHSAKSLHAKKKQRTSSSSSS SSSSSSTSSEKTSSNSSLT (SEQ ID NO: 885)

Table C1. Exemplary *Anellovirus* nucleic acid sequence (*Gammatorquevirus*)

Name	Ring4
Genus/Clade	*Gammatorquevirus*
Accession Number	

Full Sequence: 3176 bp

```
1         10        20        30        40        50
|         |         |         |         |         |
TAAAATGGCGGGAGCCAATCATTTTATACTTTCACTTTCCAATTAAAAAT
GGCCACGTCACAAACAAGGGGTGGAGCCATTTAAACTATATAACTAAGTG
GGGTGGCGAATGGCTGAGTTTACCCCGCTAGACGGTGCAGGGACCGGATC
GAGCGCAGCGAGGAGGTCCCCGGCTGCCCATGGGCGGGAGCCGAGGTGAG
TGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCTAGCGGAA
CGGGCAAGAAACTTAAAACAATATTTGTTTTACAGATGGTTAGTATATCC
TCAAGTGATTTTTTTAAGAAAACGAAATTTAATGAGGAGACGCAGAACCA
AGTATGGATGTCTCAAATTGCTGACTCTCATGATAATATCTGCAGTTGCT
GGCATCCATTTGCTCACCTTCTTGCTTCCATATTTCCTCCTGGCCACAAA
GATCGTGATCTTACTATTAACCAAATTCTTCTAAGAGATTATAAAGAAAA
ATGCCATTCTGGTGGAGAAGAAGGAGAAAATTCTGGACCAACAACAGGTT
TAATTACACCAAAAGAAGAAGATATAGAAAAAGATGGCCCAGAAGGCGCC
GCAGAAGAAGACCATACAGACGCCCTGTTCGCCGCCGCCGTAGAAAACTT
CGAAAGGTAAAGAGAAAAAAAAAATCTTTAATTGTTAGACAATGGCAACC
AGACAGTATAAGAACTTGTAAAATTATAGGACAGTCAGCTATAGTTGTTG
GGGCTGAAGGAAAGCAAATGTACTGTTATACTGTCAATAAGTTAATTAAT
GTGCCCCCAAAAACACCATATGGGGGAGGCTTTGGAGTAGACCAATACAC
ACTGAAATACTTATATGAAGAATACAGATTTGCACAAAACATTTGGACAC
AATCTAATGTACTGAAAGACTTATGCAGATACATAAATGTTAAGCTAATA
TTCTACAGAGACAACAAAACAGACTTTGTCCTTTCCTATGACAGAAACCC
ACCTTTTCAACTAACAAAATTTACATACCCAGGAGCACACCCACAACAAA
TCATGCTTCAAAAACACCACAAATTCATACTATCACAAATGACAAAGCCT
AATGGAAGACTAACAAAAAAACTCAAAATTAAACCTCCTAAACAAATGCT
TTCTAAATGGTTCTTTTCAAAACAATTCTGTAAATACCCTTTACTATCTC
TTAAAGCTTCTGCACTAGACCTTAGGCACTCTTACCTAGGCTGCTGTAAT
GAAAATCCACAGGTATTTTTTTATTATTTAAACCATGGATACTACACAAT
AACAAACTGGGGAGCACAATCCTCAACAGCATACAGACCTAACTCCAAGG
TGACAGACACAACATACTACAGATACAAAAATGACAGAAAAAATATTAAC
ATTAAAAGCCATGAATACGAAAAAAGTATATCATATGAAAACGGTTATTT
TCAATCTAGTTTCTTACAAACACAGTGCATATATACCAGTGAGCGTGGTG
AAGCCTGTATAGCAGAAAAACCACTAGGAATAGCTATTTACAATCCAGTA
AAAGACAATGGAGATGGTAATATGATATACCTTGTAAGCACTCTAGCAAA
CACTTGGGACCAGCCTCCAAAAGACAGTGCTATTTTAATACAAGGAGTAC
CCATATGGCTAGGCTTATTTGGATATTTAGACTACTGTAGACAAATTAAA
GCTGACAAAACATGGCTAGACAGTCATGTACTAGTAATTCAAAGTCCTGC
TATTTTTACTTACCCAAATCCAGGAGCAGGCAAATGGTATTGTCCACTAT
CACAAAGTTTTATAAATGGCAATGGTCCGTTTAATCAACCACCTACACTG
CTACAAAAAGCAAAGTGGTTTCCACAAATACAATACCAACAAGAAATTAT
TAATAGCTTTGTAGAATCAGGACCATTTGTTCCCAAATATGCAAATCAAA
CTGAAAGCAACTGGGAACTAAAATATAAATATGTTTTTACATTTAAGTGG
GGTGGACCACAATTCCATGAACCAGAAATTGCTGACCCTAGCAAACAAGA
GCAGTATGATGTCCCCGATACTTTCTACCAAACAATACAAATTGAAGATC
CAGAAGGACAAGACCCCAGATCTCTCATCCATGATTGGGACTACAGACGA
GGCTTTATTAAAGAAAGATCTCTTAAAAGAATGTCAACTTACTTCTCAAC
TCATACAGATCAGCAAGCAACTTCAGAGGAAGACATTCCCAAAAAGAAAA
```

AGAGAATTGGACCCCAACTCACAGTCCCACAACAAAAAGAAGAGGAGACA
CTGTCATGTCTCCTCTCTCTCTGCAAAAAAGATACCTTCCAAGAAACAGA
GACACAAGAAGACCTCCAGCAGCTCATCAAGCAGCAGCAGGAGCAGCAGC
TCCTCCTCAAGAGAAACATCCTCCAGCTCATCCACAAACTAAAAGAGAAT
CAACAAATGCTTCAGCTTCACACAGGCATGTTACCTTAACCAGATTTAAA
CCTGGATTTGAAGAGCAAACAGAGAGAGAATTAGCAATTATATTTCATAG
GCCCCCTAGAACCTACAAAGAGGACCTTCCATTCTATCCCTGGCTACCAC
CTGCACCCCTTGTACAATTTAACCTTAACTTCAAAGGCTAGGCCAACAAT
GTACACTTAGTAAAGCATGTTTATTAAAGCACAACCCCCAAAATAAATGT
AAAAATAAAAAAAAAAAAAAAAAAATAAAAAATTGCAAAAATTCGGCGCT
CGCGCGCATGTGCGCCTCTGGCGCAAATCACGCAACGCTCGCGCGCCCGC
GTATGTCTCTTTACCACGCACCTAGATTGGGGTGCGCGCGCTAGCGCGCG
CACCCCAATGCGCCCCGCCCTCGTTCCGACCCGCTTGCGCGGGTCGGACC
ACTTCGGGCTCGGGGGGGCGCGCCTGCGGCGCTTTTTTACTAAACAGACT
CCGAGCCGCCATTTGGCCCCCTAAGCTCCGCCCCCCTCATGAATATTCAT
AAAGGAAACCACATAATTAGAATTGCCGACCACAAACTGCCATATGCTAA
TTAGTTCCCCTTTTACAAAGTAAAAGGGGAAGTGAACATAGCCCCACACC
CGCAGGGGCAAGGCCCCGCACCCCTACGTCACTAACCACGCCCCCGCCGC
CATCTTGGGTGCGGCAGGGCGGGGGC (SEQ ID NO: 886)

Annotations:

Putative Domain	*Base range*
TATA Box	87– 93
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

Table C2. Exemplary *Anellovirus* amino acid sequences (*Gammatorquevirus*)

Ring4 (*Gammatorquevirus*)	
ORF2	MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPP GHKDRDLTINQILLR DYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGAAEEDHTDALFAAAV ENFER (SEQ ID NO: 887)
ORF2/2	MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPP GHKDRDLTINQILLRDYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGA AEEDHTDALFAAAVENFESGVDHNSMNQKLLTLANKSSMMSPILSTKQYKL KIQKDKTPDLSSMIGTTDEALLKKDLLKECQLTSQLIQISKQLQRKTFPKRKR ELDPNSQSHNKKKRRHCHVSSLSAKKIPSKKQRHKKTSSSSSSSSRSSSSSR ETSSSSSTN (SEQ ID NO: 888)
ORF2/3	MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPP GHKDRDLTINQILLRDYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGA AEEDHTDALFAAAVENFERSASNFRGRHSQKEKENWTPTHSPTTKRRGDTV MSPLSLQKRYLPRNRDTRRPPAAHQAAAGAAAPPQEKHPPAHPQTKRESTN ASASHRHVTLTRFKPGFEEQTERELAIIFHRPPRTYKEDLPFYPWLPPAPLVQ FNLNFKG (SEQ ID NO: 889)
TAIP	MRRRRTKYGCLKLLTLMIISAVAGIHLLTFLLPYFLLATKIVILLLTKFF (SEQ ID NO: 890)
ORF1	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRK LRKVKRKKKSLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLINV PPKTPYGGGFGVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLIFY RDNKTDFVLSYDRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKPNGR LTKKLKIKPPKQMLSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCNENPQ VFFYYLNHGYYTITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNINIKSHE YEKSISYENGYFQSSFLQTQCIYTSERGEACIAEKPLGIAIYNPVKDNGDGNM IYLVSTLANTWDQPPKDSAILIQGVPIWLGLFGYLDYCRQIKADKTWLDSHV LVIQSPAIFTYPNPGAGKWYCPLSQSFINGNGPFNQPPTLLQKAKWFPQIQYQ QEIINSFVESGPFVPKYANQTESNWELKYKYVFTFKWGGPQFHEPEIADPSK QEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDYRRGFIKERSLKRMSTYFST

	HTDQQATSEEDIPKKKKRIGPQLTVPQQKEEETLSCLLSLCKKDTFQETETQE DLQQLIKQQQEQQLLLKRNILQLIHKLKENQQMLQLHTGMLP (SEQ ID NO: 891)
ORF1/1	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRK LRKWGGPQFHEPEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDY RRGFIKERSLKRMSTYFSTHTDQQATSEEDIPKKKKRIGPQLTVPQQKEEETL SCLLSLCKKDTFQETETQEDLQQLIKQQQEQQLLLKRNILQLIHKLKENQQM LQLHTGMLP (SEQ ID NO: 892)
ORF1/2	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRK LRKISKQLQRKTFPKRKR ELDPNSQSHNKKKRRHCHVSSLSAKKIPSKKQRHKKTSSSSSSSSRSSSSSSR ETSSSSSTN (SEQ ID NO: 893)

Table E1. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus*) – Clade 1

Name Ring5.2

Genus/Clade *Alphaatorquevirus* Clade 1

Accession Number

Full Sequence: 3696 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
ATTTTGTTCAGCCCGCCAATTTCTCTTTCAAACAGGCCAATCAGCTACTA
CTTCGTGCACTTCCTGGGGCGTGTCCTGCCGCTCTATATAAGCAGAGGCG
GTGACGAATGGTAGAGTTTTTCTTGGCCCGTCCGCGGCGAGAGCGCGAGC
GAAGCGAGCGATCGAGCGTCCCGAGGGCGGGTGCCGGAGGTGAGTTTACA
CACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTATGGGCA
AGATTCTTAAAAAATTCCCCCGATCCCTTTGCCGCCAGGACATAAAAACA
TGCCGTGGAGACCGCCGGTCCATAGTGTCCAGGGGCGAGAGGATCAGTGG
TTCGCAAGCTTTTTTCACGGCCACGATTCGTTTTGCGGCTGCGGTGACCC
TCTTGGCCATATTAATAGCATTGCTCATCGCTTTCCTCGCGCCGGTCCAC
CAAGGCCCCCTCCGGGGCTAGATCAGCCTAACCCCCGGGAGCAGGGCCCG
GCCGGACCCGGAGGGCCGCCCGCCATCTTGGCCCTGCCGGCTCCGCCCGC
GGAGCCTGACGACCCGCAGCCACGGCGTGGTGGTGGGGACGGTGGCGCCG
CCGCTGGCGCCGCAGACGACCATACACAACGAGACTACGACGAAGAAGAG
CTAGACGAGCTTTTCCGCGCCGCCGCCGAAGACGATTTGTAAGTAGGAGA
TGGCGCCGGCCTTACAGGCGCAGGAGGAGACGCGGGCGACGCAGACGCAG
ACGCAGACGCAGACATAAGCCCACCCTAATACTCAGACAGTGGCAACCTG
ACTGTATCAGACACTGTAAAATAACAGGATGGATGCCCCTCATTATCTGT
GGAAAGGGGTCCACCCAGTTCAACTACATCACCCACGCGGACGATATCAC
CCCCAGGGGAGCCTCCTACGGAGGCAATTTCACAAACATGACTTTCTCCC
```

TGGAGGCCATATATGAACAGTTCCTATACCACAGAAACAGGTGGTCGGCC
TCTAACCACGACCTAGAACTGTGCAGATACAAGGGGACCACCTTAAAACT
CTACAGACACCCAGAAGTAGACTACATAGTTACCTACAGCAGAACAGGAC
CCTTTGAAATCAGCCACATGACCTACCTCAGCACTCACCCCATGCTAATG
CTGCTAAACAAGCACCACATTGTGGTGCCCAGCTTAAAGACTAAGCCCAG
AGGCAGAAAGGCCATAAAAGTCAGGATAAGGCCCCCAAAACTCATGAACA
ACAAGTGGTACTTCACCAGAGACTTCTGTAACATAGGCCTCTTCCAGCTC
TGGGCCACAGGCTTAGAACTCAGAAACCCCTGGCTCAGAATGAGCACCCT
GAGCCCCTGCATAGGCTTTAATGTCCTCAAAAACAGCATTTACACAAACC
TCAGCAACCTGCCACAATACAAAAACGAAAGACTAAACATCATTAACAAC
ATACTTCACCCACAAGAAATTACAGGTACAAACAACAAAAAGTGGCAGTA
CACATACACCAAACTCATGGCCCCTATTTACTATTCAGCAAACAGGGCCA
GCACCTATGACTGGGAAAATTACAGCAAAGAAACAAACTACAATAATACA
TATGTTAAATTTACCCAGAAAAGACAGGAAAAACTAACTAAAATTAGAAA
AGAGTGGCAGATGCTTTATCCACAACAACCCACAGCACTGCCAGACTCCT
ATGACCTCCTACAAGAGTATGGCCTCTACAGTCCATACTACCTAAACCCC
ACAAGAATAAACCTAGACTGGATGACCCCATACACACACGTCAGATACAA
TCCCCTAGTAGACAAGGGCTTTGGAAACAGAATATACATCCAGTGGTGCT
CAGAAGCAGATGTTAGCTACAACAGGACAAAATCCAAGTGTCTGCTACAA
GACATGCCCCTGTTTTTCATGTGCTATGGCTACATAGACTGGGCAATAAA
AAACACTGGAGTGTCATCTCTAGTGAAGGACGCCAGAATCTGCATCAGGT
GTCCCTACACAGAGCCACAACTAGTTGGCTCCACAGAAGACATAGGCTTT
GTACCCATCTCAGAAACCTTCATGAGGGGCGACATGCCGGTACTTGCACC
ATACATACCGTTAAGCTGGTTTTGCAAGTGGTATCCCAACATAGCTCACC
AAAAGGAAGTCCTTGAGTCAATCATTTCCTGCAGCCCCTTCATGCCCCGT
GACCAAGACATGAACGGTTGGGATATCACAATCGGTTACAAAATGGACTT
CTTATGGGGCGGTTCCCCTCTCCCCTCACAGCCAATCGACGACCCCTGCC
AGCAGGGAACCCACCCGATTCCCGACCCCGATAAACACCCTCGCCTCCTA
CAAGTCTCGAACCCGAAACTACTCGGACCGAGGACAGTGTTCCACAAGTG
GGACATCAGACGTGGGCAGTTTAGCAAAAGAAGTATTAAGAGAGTGTCAG
AATACTCAAGCGATGATGAATCTCTTGCGCCAGGTCTCCCATCAAAGCGA
AACAAGCTCGACTCGGCGTTCCGAGGAGAAAATCGAGAGCAAAAAGAATG
CTATTCTCTCCTCAAAGCGCTCGAGGAAGAAGAGACCCCAGAAGAAGAAG
AACCAGCACCCCAAGAAAAAGCCCAGAAAGAGGAGCTACTCCACCAGCTC
CAGCTCCAGAGACGCCACCAGCGAGTCCTCAGACGAGGGCTCAAGCTCGT
CTTTACAGACATCCTCCGACTCCGCCAGGGAGTCCACTGGAACCCGGAGC
TCACATAGCGCCCCCACCTTACATACCAGACCTGCTTTTTCCCAATACTG
GTAAAAAAAAAAAATTCTCTCCCTTCGATTGGGAGACAGAGGCGCAAATA
GCGGGGTGGATGCGGCGGCCCATGCGCTTCTATCCCTCAGACACCCCTCA
CTACCCGTGGCTACCCCCCGAGCGAGATATCCCGAAAATATGTAACATAA
ACTTCAAAATAAAGCTTCAAGAGTGAGTGATTCGAGGCCCTCCTCTGTTC
ACTTAGCGGTGTCTACCTCTTAAGGTCACTAAGCACTCCGAGCGTAAGCG
AGGAGTGCGACCCTCTACCAAGGGGCAACTTCCTCGGGGTCCGGCGCTAC
GCGCTTCGCGCTGCGCCGGACATCTCGGACCCCTCGACCCGAATCGCTTG
CGCGATTCGGACCTGCGGCCTCGGGGGGGTCGGGGGCTTTACTAAACAGA
CTCCGAGGTGCCATTGGACACTGTAGGGGGTGAACAGCAACGAAAGTGAG
TGGGGCCAGACTTCGCCATAAGGCCTTTATCTTCTTGCCATTGGATAGTG
ACTTCCGGGTCCGCCTGGGGGCCGCCATTTTAGCTTCGGCCGCCATTTTA
GGCCCTCGCGGGCCTCCGTAGGCGCGCTTTAGTGACGTCACGGCAGCCAT
TTTGTCGTGACGTTTGAGACACGTGATGGGGGCGTGCCTAAACCCGGAAG
CATCCCTGGTCACGTGACTCTGACGTCACGGCGGCCATCTTGTGCTGTCC
GCCATCTTGTAACTTCCTTCCGCTTTTTCAAAAAAAAAGAGGAAGTGTGA
CGTAGCGGCGGGGGGGCGGCGCGCTTCGCGCGCCGCCCACCAGGGGGCGC
TGCGCGCCCCCCGCGCATGCGCAGGGGCCTCTCGAGGGGCTCCGCCCCCC
CCCCGTGCTAAATTTACCGCGCATGCGCGACCACGCCCCCGCCGCC (SEQ ID NO: 894)

Annotations:

Putative Domain	*Base range*
TATA Box	85– 91
Cap Site	108 – 115
Transcriptional Start Site	115
5’ UTR Conserved Domain	178 – 248
ORF2	300 – 692
ORF2/2	300 – 688 ; 2282 – 2804
ORF2/3	300 – 688 ; 2484 – 2976
ORF2t/3	300 – 349 : 2484 - 2976
TAIP	322 - 471
ORF1	572 – 2758
ORF1/1	572 – 688 ; 2282 – 2758
ORF1/2	572 – 688 ; 2484 – 2804
Three open-reading frame region	2484 – 2755
Poly(A) Signal	3018 –3023
GC-rich region	3555 – 3696

Table D2. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*) Clade 1

Ring 5.2 (*Alphaatorquevirus*) Clade 1	
ORF2	MPWRPPVHSVQGREDQWFASFFHGHDSFCGCGDPLGHINSIAHRFPRAGPP RPPPGLDQPNPREQGPAGPGGPPAILALPAPPAEPDDPQPRRGGGDGGAAAG AADDHTQRDYDEEELDELFRAAAEDDL (SEQ ID NO: 895)
ORF2/2	MPWRPPVHSVQGREDQWFASFFHGHDSFCGCGDPLGHINSIAHRFPRAGPP RPPPGLDQPNPREQGPAGPGGPPAILALPAPPAEPDDPQPRRGGGDGGAAAG AADDHTQRDYDEEELDELFRAAAEDDFQSTTPASREPTRFPTPINTLASYKS RTRNYSDRGQCSTSGTSDVGSLAKEVLRECQNTQAMMNLLRQVSHQSETSS TRRSEEKIESKKNAILSSKRSRKKRPQKKKNQHPKKKPRKRSYSTSSSSRDAT SESSDEGSSSSLQTSSDSARESTGTRSSHSAPTLHTRPAFSQYW (SEQ ID NO: 896)
ORF2/3	MPWRPPVHSVQGREDQWFASFFHGHDSFCGCGDPLGHINSIAHRFPRAGPP RPPPGLDQPNPREQGPAGPGGPPAILALPAPPAEPDDPQPRRGGGDGGAAAG

	AADDHTQRDYDEEELDELFRAAAEDDLSPIKAKQARLGVPRRKSRAKRMLF SPQSARGRRDPRRRRTSTPRKSPERGATPPAPAPETPPASPQTRAQARLYRHP PTPPGSPLEPGAHIAPPPYIPDLLFPNTGKKKKFSPFDWETEAQIAGWMRRPM RFYPSDTPHYPWLPPERDIPKICNINFKIKLQ (SEQ ID NO: 897)
ORF2t/3	MPWRPPVHSVQGREDQWSPIKAKQARLGVPRRKSRAKRMLFSPQSARGRR DPRRRRTSTPRKSPERGATPPAPAPETPPASPQTRAQARLYRHPPTPPGSPLEP GAHIAPPPYIPDLLFPNTGKKKKFSPFDWETEAQIAGWMRRPMRFYPSDTPH YPWLPPERDIPKICNINFKIKLQE (SEQ ID NO: 898)
TAIP	IVSRGERISGSQAFFTATIRFAAAVTLLAILIALLIAFLAPVHQGPLRG (SEQ ID NO: 899)
ORF1	TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFVSRRWRRPYR RRRRRGRRRRRRRRRHKPTLILRQWQPDCIRHCKITGWMPLIICGKGSTQFN YITHADDITPRGASYGGNFTNMTFSLEAIYEQFLYHRNRWSASNHDLELCRY KGTTLKLYRHPEVDYIVTYSRTGPFEISHMTYLSTHPMLMLLNKHHIVVPSL KTKPRGRKAIKVRIRPPKLMNNKWYFTRDFCNIGLFQLWATGLELRNPWLR MSTLSPCIGFNVLKNSIYTNLSNLPQYKNERLNIINNILHPQEITGTNNKKWQ YTYTKLMAPIYYSANRASTYDWENYSKETNYNNTYVKFTQKRQEKLTKIR KEWQMLYPQQPTALPDSYDLLQEYGLYSPYYLNPTRINLDWMTPYTHVRY NPLVDKGFGNRIYIQWCSEADVSYNRTKSKCLLQDMPLFFMCYGYIDWAIK NTGVSSLVKDARICIRCPYTEPQLVGSTEDIGFVPISETFMRGDMPVLAPYIPL SWFCKWYPNIAHQKEVLESIISCSPFMPRDQDMNGWDITIGYKMDFLWGGS PLPSQPIDDPCQQGTHPIPDPDKHPRLLQVSNPKLLGPRTVFHKWDIRRGQFS KRSIKRVSEYSSDDESLAPGLPSKRNKLDSAFRGENREQKECYSLLKALEEE ETPEEEEPAPQEKAQKEELLHQLQLQRRHQRVLRRGLKLVFTDILRLRQGVH WNPELT (SEQ ID NO: 900)
ORF1/1	TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFPIDDPCQQGT HPIPDPDKHPRLLQVSNPKLLGPRTVFHKWDIRRGQFSKRSIKRVSEYSSDDE SLAPGLPSKRNKLDSAFRGENREQKECYSLLKALEEEETPEEEEPAPQEKAQ KEELLHQLQLQRRHQRVLRRGLKLVFTDILRLRQGVHWNPELT (SEQ ID NO: 901)

ORF1/2	TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFVSHQSETSSTR RSEEKIESKKNAILSSK RSRKKRPQKKKNQHPKKKPRKRSYSTSSSSRDATSESSDEGSSSSLQTSSDSA RESTGTRSSHSAPTLHTRPAFSQYW (SEQ ID NO: 902)

Table F1. 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
|         |         |         |         |         |
TTATTAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAA
CCGTCACTAACTTCCTTATTTAACATTACTTCCCTTTTAACCAATGAATA
TTCATACAACACATCACACTTCCTGGGAGGAGACATAAAACTATATAACT
AACTACACAGACGAATGGCTGAGTTTATGCCGCTAGACGGAGGACGCACA
GCTACTGCTGCGACCTGAACTTGGGCGGGTGCCGAAGGTGAGTGTAACCA
CCGTAGTCAAGGGGCAATTCGGGCTAGTTCAGTCTAGCGGAACGGGCAAG
ATTATTAATACAAACTTATTTTTACAGATGAGCAAACAACTAAAACCAAC
TTTATACAAAGACAAATCATTGGAATTACAATGGCTAAACAACATTTTTA
GCTCTCACGACCTGTGCTGCGGCTGCAACGATCCAGTTTTACATTTACTG
ATTTTAATTAACAAAACCGGAGAAGCACCTAAACCAGAAGAAGACATTAA
AAATATAAAATGCCTCCTTACTGGCGCCAAAAATACTACCGAAGAAGATA
TAGACCTTTCTCCTGGAGAACTAGAAGAATTATTCAAAGAAGAAAAAGAT
GGAGATACCGCAAACCAAGAAAAACATACTGGAGAAGAAAACTGCGGGTA
AGAAAACGTTTTTATAAAAGAAAGTTAAAAAAAATTGTACTTAAACAGTT
TCAACCAAAAATTATTAGAAGATGTACAATATTTGGAACAATCTGCCTAT
TTCAAGGCTCTCCAGAAAGAGCCAACAATAATTATATTCAAACAATCTAC
TCCTACGTACCAGATAAAGAACCAGGAGGAGGGGGATGGACTTTAATAAC
TGAAAGCTTAAGTAGTTTATGGGAAGACTGGGAACATTTAAAAAATGTAT
GGACTCAAAGTAACGCTGGTTTACCACTTGTAAGATACGGGGGAGTAACA
TTATACTTTTATCAATCTGCCTATACTGACTATATTGCTCAAGTTTTCAA
CTGTTATCCTATGACAGACACAAAATACACACATGCAGACTCAGCACCAA
ACAGAATGTTATTAAAAAAACATGTAATAAGAGTACCTAGCAGAGAAACA
CGCAAAAAAAGAAAGCCATACAAAAGAGTTAGAGTAGGACCTCCTTCTCA
AATGCAAAACAAATGGTACTTTCAAAGAGACATATGTGAAATACCATTAA
TAATGATTGCAGCCACAGCCGTTGACTTTAGATATCCCTTTTGTGCAAGC
GACTGTGCTAGTAACAACTTAACTCTAACATGTTTAAACCCACTATTGTT
TCAAAACCAAGACTTTGACCACCCATCCGATACACAAGGCTACTTTCCAA
AACCTGGAGTATATCTATACTCAACACAAAGAAGTAACAAGCCAAGTTCT
TCAGACTGTATATACTTAGGAAACACAAAAGACAATCAAGAAGGTAAATC
TGCAAGTAGTCTAATGACTCTAAAAACACAAAAAATAACAGATTGGGGAA
ATCCATTTTGGCATTATTATATAGACGGTTCTAAAAAAATATTTTCTTAC
TTTAAACCCCCATCACAATTAGACAGCAGCGACTTTGAACACATGACAGA
ATTAGCAGAACCAATGTTTATACAAGTTAGATACAACCCAGAAAGAGACA
CAGGACAAGGAAACTTAATATACGTAACAGAAAACTTTAGAGGACAACAC
```

TGGGACCCTCCATCTAGTGACAACCTAAAATTAGATGGATTTCCCTTATA
TGACATGTGCTGGGGTTTCATAGACTGGATAGAAAAAGTTCATGAAACAG
AAAACTTACTTACCAACTACTGCTTCTGTATTAGAAGCAGCGCTTTCAAT
GAAAAAAAAACAGTTTTTATACCTGTAGATCATTCATTTTTAACAGGTTT
TAGCCCATATGAAACTCCAGTTAAATCATCAGACCAAGCTCACTGGCACC
CACAAATAAGATTTCAAACAAAATCAATAAATGACATTTGTTTAACAGGC
CCCGGTTGTGCTAGGTCCCCATATGGCAATTACATGCAGGCAAAAATGAG
TTATAAATTTCATGTAAAATGGGGAGGATGTCCAAAAACTTATGAAAAAC
CATATGATCCTTGTTCACAGCCCAATTGGACTATTCCCCATAACCTCAAT
GAAACAATACAAATCCAGAATCCAAACACATGCCCACAAACAGAACTCCA
AGAATGGGACTGGCGACGTGATATTGTTACAAAAAAAGCTATCGAAAGAA
TTAGACAACACACGGAACCTCATGAAACTTTGCAAATCTCTACAGGTTCC
AAACACAACCCACCAGTACACAGACAAACATCACCGTGGACGGACTCAGA
AACGGACTCGGAAGAGGAAAAAGACCAAACACAAGAGATCCAGATCCAGC
TCAACAAGCTCAGAAAGCATCAACAGCATCTCAAGCAGCAGCTCAAGCAG
TACCTGAAACCCCAAAATATAGAATAGTTGCAAGCAACATAAAAGTTGAA
CTTTTTCCTACTAAAAAACCTTTTAAAAACAGACGCTTTACTCCTTCTGA
AAGAGAAACAGAAAGACAATGTGCTAAAGCTTTTTGTAGACCAGAAAGAC
ATTTCTTTTATGATCCTCCTTTTTACCCTTACTGTGTACCTGAACCTATT
GTAAACTTTGCTTTGGGATATAAAATTTAAGGCCAACAAATTTCACTTAG
TGGTGTCTGTTTATTAAAGTTTAACCTTAATAAGCATACTCCGCCTCCCT
ACATTAAGGCGCCAAAAGGGGGCTCCGCCCCCTTAAACCCCAAGGGGGCT
CCGCCCCCTTAAACCCCCAAGGGGGCTCCGCCCCCTTACACCCCC (SEQ ID NO: 1001)

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 F2. Exemplary *Anellovirus* amino acid sequences (*Betatorquevirus*)

Ring9 (*Betatorquevirus*)

ORF2	MSKQLKPTLYKDKSLELQWLNNIFSSHDLCCGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGDTANQEKHTGEEN CG (SEQ ID NO: 1002)
ORF2/2	MSKQLKPTLYKDKSLELQWLNNIFSSHDLCCGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGDTANQEKHTGEEN CGPIGLFPITSMKQYKSRIQTHAHKQNSKNGTGDVILLQKKLSKELDNTRNL MKLCKSLQVPNTTHQYTDKHHRGRTQKRTRKRKKTKHKRSRSSSTSSESIN SISSSSSSST (SEQ ID NO: 1003)
ORF2/3	MSKQLKPTLYKDKSLELQWLNNIFSSHDLCCGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGDTANQEKHTGEEN CGFQTQPTSTQTNITVDGLRNGLGRGKRPNTRDPDPAQQAQKASTASQAAA QAVPETPKYRIVASNIKVELFPTKKPFKNRRFTPSERETERQCAKAFCRPERH FFYDPPFYPYCVPEPIVNFALGYKI (SEQ ID NO: 1004)
ORF1	MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTYWRRKLRVRKR FYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQGSPERANNNYIQTIYSYVPDKE PGGGGWTLITESLSSLWEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSA YTDYIAQVFNCYPMTDTKYTHADSAPNRMLLKKHVIRVPSRETRKKRKPYK RVRVGPPSQMQNKWYFQRDICEIPLIMIAATAVDFRYPFCASDCASNNLTLT CLNPLLFQNQDFDHPSDTQGYFPKPGVYLYSTQRSNKPSSSDCIYLGNTKDN QEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSYFKPPSQLDSSDFEH MTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFP LYDMCWGFIDWIEKVHETENLLTNYCFCIRSSAFNEKKTVFIPVDHSFLTGFS PYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSPYGNYMQAKMSYK FHVKWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWD WRRDIVTKKAIERIRQHTEPHETLQISTGSKHNPPVHRQTSPWTDSETDSEEE KDQTQEIQIQLNKLRKHQQHLKQQLKQYLKPQNIE (SEQ ID NO: 1005)
ORF1/1	MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTYWRRKLRPNW TIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAIERIRQHTEPHETLQIS TGSKHNPPVHRQTSPWTDSETDSEEEKDQTQEIQIQLNKLRKHQQHLKQQL KQYLKPQNIE (SEQ ID NO: 1006)

ORF1/2	MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTYWRRKLRVPNT THQYTDKHHRGRTQKRTRKRKKTKHKRSRSSSTSSESINSISSSSSSST (SEQ ID NO: 1007)

Table F3. Exemplary *Anellovirus* nucleic acid sequence (*Betatorquevirus*)

Name	Ring10
Genus/Clade	*Betatorquevirus*
Accession Number	JX134044.1

Full Sequence: 2912 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAATAAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAA
CCGTCACTTAGTTCCTCTTTTTCCACAACTTCCTCTTTTACTAATGAATA
TTCATGTAATTAATTAATAATCACCGTAATTCCGGGGAGGAGCCTTTAAA
CTATAAAACTAACTACACATTCGAATGGCTGAGTTTATGCCGCCAGACGG
AGACGGGATCACTTCAGTGACTCCAGGCTGATCAAGGGCGGGTGCCGAAG
GTGAGTGAAACCACCGTAGTCAAGGGGCAATTCGGGCTAGATCAGTCTGG
CGGAACGGGCAAGAAACTTAAAATGTACTTTATTTTACAGAAATGTTCAA
ATCTCCAACATACTTAACAACTAAAGGCAAAAACAATGCCTTAATCAACT
GCTTCGTTGGAGACCACGATCTTCTGTGCAGCTGTAACAATCCTGCCTAC
CATTGCCTCCAAATACTTGCAACTACCTTAGCACCTCAACTAAAACAAGA
AGAAAAACAACAAATAATACAATGCCTTGGTGGTACAGACGCCGTAGCTA
CAACCCGTGGAGACGAAGAAATTGGTTTAGAAGACCTAGAAAAACTATTT
ACAGAAGATACAGAAGAAGACGCCGCTGGGTAAGAAGAAAACCTTTTTAC
AAACGTAAAATTAAGAGACTAAATATAGTAGAATGGCAACCTAAATCAAT
TAGAAAATGTAGAATAAAAGGAATGCTATGCTTGTTTCAAACGACAGAAG
ACAGACTGTCATATAACTTTGATATGTATGAAGAGTCTATTATACCAGAA
AAACTGCCGGGAGGGGGGGGATTTAGCATTAAGAATATAAGCTTATATGC
CTTATACCAAGAACACATACATGCACACAACATATTTACACACACAAACA
CAGACAGACCACTAGCAAGATACACAGGCTGTTCTTTAAAATTCTACCAA
AGCAAAGACATAGACTACGTAGTAACATATTCTACATCACTCCCACTAAG
AAGCTCAATGGGAATGTACAACTCCATGCAACCATCCATACATCTAATGC
AACAAAACAAACTAATTGTACCAAGCAAACAAACACAAAAAAGAAGAAAA
CCATATATTAAAAAACATATATCACCACCAACACAAATGAAATCTCAATG
GTACTTTCAACATAACATTGCAAACATACCGCTACTAATGATAAGAACCA
CAGCATTAACATTAGATAATTACTATATAGGAAGCAGACAATTAAGTACA
AATGTCACTATACATACACTTAACACAACATACATCCAAAACAGAGACTG
GGGAGACAGAAATAAAACTTACTACTGCCAAACATTAGGAACACAAAGAT
ACTTCCTATATGGAACACATTCAACTGCACAAAATATTAATGACATAAAG
CTACAAGAACTAATACCTTTAACAAACACACAAGACTATGTACAAGGCTT
TGATTGGACAGAAAAAGACAAACATAACATAACAACCTACAAAGAATTCT
TAACTAAAGGAGCAGGAAATCCATTTCACGCAGAATGGATAACAGCACAA
AACCCAGTAATACACACAGCAAACAGTCCTACACAAATAGAACAAATATA
CACCGCTTCAACAACAACATTCCAAAACAAAAAACTAACAGACCTACCAA
CGCCAGGATATATATTTATAACTCCAACAGTAAGCTTAAGATACAACCCA
TACAAAGACCTAGCAGAAAGAAACAAATGCTACTTTGTAAGAAGCAAAAT
AAATGCACACGGGTGGGACCCAGAACAACACCAAGAATTAATAAACAGTG
```

ACCTACCACAATGGTTACTATTATTTGGCTACCCAGACTACATAAAAAGA
ACACAAAACTTTGCATTAGTAGACACAAATTACATACTAGTAGACCACTG
CCCATACACAAATCCAGAAAAAACACCATTTATACCTTTAAGCACATCAT
TTATAGAAGGTAGAAGCCCATACAGTCCTTCAGACACACATGAACCAGAT
GAAGAAGACCAAAACAGGTGGTACCCATGCTACCAATATCAACAAGAATC
AATAAATTCAATATGTCTTAGCGGTCCAGGCACACCAAAAATACCAAAAG
GAATAACAGCAGAAGCAAAAGTAAAATATTCCTTTAATTTTAAGTGGGGT
GGTGACCTACCACCAATGTCTACAATTACAAACCCGACAGACCAGCCAAC
ATATGTTGTTCCCAATAACTTCAATGAAACAACTTCGTTACAGAATCCAA
CCACCAGACCAGAGCACTTCTTGTACTCCTTTGACGAAAGGAGGGGACAA
CTTACAGAAAAAGCTACAAAACGCTTGCTTAAAGACTGGGAAACTAAAGA
AACTTCTTTATTGTCTACAGAATACAGATTCGCGGAGCCAACACAAACAC
AAGCCCCACAAGAGGACCCGTCCTCGGAAGAAGAAGAAGAGAGCAACCTC
TTCGAGCGACTCCTCCGACAGCGAACCAAGCAGCTCCAGCTCAAGCGCAG
AATAATACAAACATTGAAAGACCTACAAAAATTAGAATAACTAACAGCAA
AAACACCGTTTACCTATTTCCACCTGAACAAAAGAACAGAAGACTAACAC
CATGGGAAATACAAGAAGACAAAGAAATAGCCAATTTATTTGGCAGACCA
CATAGATACTTTTTAAAAGACATTCCTTTCTATTGGGATATACCCCCAGA
GCCTAAAGTAAACTTTGATTTAAATTTTCAATAAAGAAATAAAGGGCAAG
GCCCCATTAACTCAAAGTCGGTGTCTACCTCTTTAAGTTTAACTTTACTA
AACGGACTCCGCCTCCCTAAATTTGGGCGCCAAAAGGGGGCTCCGCCCCC
TTAAACCCCAGGGGGCTCCGCCCCCTAAAACCCCCAAGGGGGCTACGCCC
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 F4. Exemplary *Anellovirus* amino acid sequences (*Betatorquevirus*)

Ring10 (*Betatorquevirus*)

ORF2	MFKSPTYLTTKGKNNALINCFVGDHDLLCSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAG (SEQ ID NO: 1009)
ORF2/2	MFKSPTYLTTKGKNNALINCFVGDHDLLCSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGQHMLFPI TSMKQLRYRIQPPDQSTSCTPLTKGGDNLQKKLQNACLKTGKLKKLLYCLQ NTDSRSQHKHKPHKRTRPRKKKKRATSSSDSSDSEPSSSSSSAE (SEQ ID NO: 1010)
ORF2/3	MFKSPTYLTTKGKNNALINCFVGDHDLLCSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGIQIRGANT NTSPTRGPVLGRRRREQPLRATPPTANQAAPAQAQNNTNIERPTKIRITNSKN TVYLFPPEQKNRRLTPWEIQEDKEIANLFGRPHRYFLKDIPFYWDIPPEPKVN FDLNFQ (SEQ ID NO: 1011)
ORF1	MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWVRRKPFYKRKIKR LNIVEWQPKSIRKCRIKGMLCLFQTTEDRLSYNFDMYEESIIPEKLPGGGGFSI KNISLYALYQEHIHAHNIFTHTNTDRPLARYTGCSLKFYQSKDIDYVVTYSTS LPLRSSMGMYNSMQPSIHLMQQNKLIVPSKQTQKRRKPYIKKHISPPTQMKS QWYFQHNIANIPLLMIRTTALTLDNYYIGSRQLSTNVTIHTLNTTYIQNRDW GDRNKTYYCQTLGTQRYFLYGTHSTAQNINDIKLQELIPLTNTQDYVQGFD WTEKDKHNITTYKEFLTKGAGNPFHAEWITAQNPVIHTANSPTQIEQIYTAS TTTFQNKKLTDLPTPGYIFITPTVSLRYNPYKDLAERNKCYFVRSKINAHGW DPEQHQELINSDLPQWLLLFGYPDYIKRTQNFALVDTNYILVDHCPYTNPEK TPFIPLSTSFIEGRSPYSPSDTHEPDEEDQNRWYPCYQYQQESINSICLSGPGTP KIPKGITAEAKVKYSFNFKWGGDLPPMSTITNPTDQPTYVVPNNFNETTSLQ NPTTRPEHFLYSFDERRGQLTEKATKRLLKDWETKETSLLSTEYRFAEPTQT QAPQEDPSSEEEEESNLFERLLRQRTKQLQLKRRIIQTLKDLQKLE (SEQ ID NO: 1012)
ORF1/1	MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWPTYVVPNNFNETT SLQNPTTRPEHFLYSFDERRGQLTEKATKRLLKDWETKETSLLSTEYRFAEP TQTQAPQEDPSSEEEEESNLFERLLRQRTKQLQLKRRIIQTLKDLQKLE

ORF1/2	MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWNTDSRSQHKHKPH KRTRPRKKKKRATSSSDSSDSEPSSSSSSAE (SEQ ID NO: 1013)

Table F5. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus*, Clade 4)

Name	Ring20
Genus/Clade	*Alphatorquevirus* Clade 4
Accession Number	AF122914.3

Full Sequence: 3853 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
GGCTTAGTGCGTCACCACCCACGTGACCCGCCTCCGCCAATTAACAGGTA
CTTCGTACACTTCCTGGGCGGGCTTATAAGACTAATATAAGTAGCTGCAC
TTCCGAATGGCTGAGTTTTCCACGCCCGTCCGCAGCGGTGAAGCCACGGA
GGGAGCTCAGCGCGTCCCGAGGGCGGGTGCCGGAGGTGAGTTTACACACC
GCAGTCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTTTGGGCAAGGC
TCTTAAAAAAGCTATGTTTATTGGCAGGCACTACCGAAAGAAAAGGGCGC
TGCTACTGCTATCTGTGCATTCTACAAAGACAAAAGGGAAACTTCTAATA
GCTATGTGGACTCCCCCACGCAATGATCAACAATACCTTAACTGGCAATG
GTACACTTCTGTACTTAGCTCCCACTCTGCTATGTGCGGGTGTTCCGACG
CTATCGCTCATCTTAATCATCTTGCTAATCTGCTTCGTGCCCCGCAAAAT
CCGCCCCCGCCTGATAATCCAAGACCCCTACCCGTGCGAGCACTGCCTGC
TCCCCCGGCTGCCCACGAGGCAGCCGGTGATCGAGCACCATGGCCTATGG
GTGGTGGAGGAGACGCCGGAGGCGCTGGCGCAGGTGGAGACGCCGACCAT
GGAGGCGCCGCTGGAGGACCCGCAGACGCAGACCTGCTAGACGCCGTGGC
CGCCGCAGAAACGTAAGGAGACGGCGCAGAGGGAGGTGGAGAAGGAGGTA
CAGGAGGTGGAAAAGAAAGGGCAGACGTAGAAGAAAAGCAAAAATAATAA
TAAGACAGTGGCAGCCAAACTACAGAAGAAGATGTAATATAGTGGGCTAC
CTCCCTATACTTATCTGTGGTGGAAATACTGTTTCTAGAAACTATGCCAC
ACACTCAGACGATACTAACTATCCAGGACCCTTTGGGGGAGGCATGACCA
CAGACAAATTCAGCCTTAGAATACTATATGATGAATACAAAAGATTTATG
AACTACTGGACAGCCTCAAATGAGGACCTAGATCTCTGTAGATATCTAGG
ATGCACTTTTTACTTCTTTAGACACCCTGAAGTAGACTTTATTATAAAAA
TAAACACCATGCCCCCATTCTTAGATACAACCATAACAGCACCTAGCATA
CACCCAGGCCTCATGGCCCTAGACAAAAGAGCCAGATGGATTCCTTCTCT
TAAAAATAGACCAGGTAAAAAACACTATATAAAAATTAGAGTAGGGGCTC
CTAAAATGTTCACAGATAAATGGTACCCTCAAACAGACCTCTGTGACATG
ACACTGCTAACTATCTATGCAACCGCAGCGGATATGCAATATCCGTTCGG
CTCACCACTAACTGACACTGTGGTTGTTAACTCCCAAGTTCTGCAATCCA
TGTATGATGAAACAATTAGCATATTACCTGATGAAAAAACTAAAAGAAAT
AGCCTTCTTACTTCTATAAGAAGCTACATACCTTTTTATAATACTACACA
AACAATAGCTCAATTAAAACCATTTGTAGATGCAGGAGGACACACAACAG
GCTCAACAACAACTACATGGGGACAACTATTAAACACAACTAAATTTACC
ACTACCACAACAACCACATACACATACCCTGGCACCACAAATACAGCAGT
AACATTTATAACAGCCAATGATACCTGGTACAGGGGAACAGCATATAAAG
ATAACATTAAAGATGTACCACAAAAAGCAGCACAATTATACTTTCAAACA
ACACAAAAACTACTAGGAAACACATTCCATGGCTCAGATGAAACACTTGA
```

ATACCATGCAGGCCTATACAGCTCTATCTGGCTATCACCAGGTAGATCCT
ACTTTGAAACACCAGGTGCATACACAGACATTAAATATAACCCTTTTACA
GACAGAGGAGAAGGCAACATGCTGTGGATAGACTGGCTAAGTAAAAAAAA
CATGAAATATGACAAAGTGCAAAGTAAGTGCCTAGTAGCAGACCTACCAC
TGTGGGCAGCAGCATATGGTTATGTAGAATTCTGCTCTAAAAGCACAGGA
GACACAAACATACACATGAATGCCAGACTACTAATAAGAAGTCCTTTTAC
AGACCCCCAGCTAATAGTACACACAGACCCCACTAAAGGCTTTGTACCCT
ATTCTTTAAACTTTGGAAATGGTAAAATGCCAGGAGGTAGCAGCAATGTT
CCCATAAGAATGAGAGCTAAGTGGTACCCCACTTTATCCCACCAACAAGA
AGTTCTAGAGGCCTTAGCACAGTCAGGACCCTTTGCTTATCACTCAGACA
TTAAAAAAGTATCTCTAGGCATAAAATACCGTTTTAAGTGGATCTGGGGT
GGAAACCCCGTTCGCCAACAGGTTGTTAGAAATCCCTGCAAGGAACCCCA
CTCCTCGGGCAATAGAGTCCCTAGAAGCATACAAATCGTTGACCCGAGAT
ACAACTCACCGGAACTTACCATCCATGCCTGGGACTTCAGACGTGGCTTC
TTTGGCCCGAAAGCTATTCAAAGAATGCAACAACAACCAACTGCTACTGA
ATTTTTTTCAGCAGGCCGCAAGAGACCCAGAAGGGACACAGAAGTGTATC
AGTCCGACCAAGAAAAGGAGCAAAAAGAAAGCTCGCTTTTCCCCCCAGTC
AAGCTCCTCCGAAGAGTCCCCCCGTGGGAGGACTCGGAACAGGAGCAAAG
CGGGTCGCAAAGCTCAGAGGAAGAGACGGCGACCCTCTCCCAGCAGCTCA
AACAGCAGCTGCAGCAGCAGCGAGTCTTGGGAGTCAAACTCAGACTCCTG
TTCAACCAAGTCCAAAAAATCCAACAAAATCAAGATATCAACCCTACCTT
GTTACCAAGGGGGGGGGATCTAGTATCCTTCTTTCAGGCTGTACCATAAA
TATGTTTCCAGACCCTAAACCTTACTGCCCCTCCAGCAATGACTGGAAAG
AAGAGTATGAGGCCTGTAAATATTGGGATAGACCTCCCAGACACAACCTT
AGAGACCCCCCCTTTTACCCCTGGGCCCCTAAAAACAATCCTTGCAATGT
AAGCTTTAAACTTGGCTTCAAATAAACTAGGCCGTGGGAGTTTCACTTGT
CGGTGTCTACCTCTATAAGTCACTAAGCACTCCGAGCGCAGCGAGGAGTG
CGACCCTTCCCCCTGGTGCAACGCCCTCGGCGGCCGCGCGCTACGCCTTC
GGCTGCGCGCGGCACCTCGGACCCCCGCTCGTGCTGACACGCTTGCGCGT
GTCAGACCACTTCGGGCTCGCGGGGGTCGGGAAATTTGCTAAACAGACTC
CGAGTTGCCATTGGACACTGTAGCTATGAATCAGTAACGAAAGTGAGTGG
GGCCAGACTTCGCCATAAGGCCTTTATCTTCTTGCCATTTGTCAGTATTG
GGGGTCGCCATAAACTTTGGGCTCCATTTTAGGCCTTCCGGACTACAAAA
ATCGCCATATTTGTGACGTCAGAGCCGCCATTTTAAGTCAGCTCTGGGGA
GGCGTGACTTCCAGTTCAAAGGTCATCCTCACCATAACTGGCACAAAATG
GCCGCCAACTTCTTCCGGGTCAAAGGTCACTGCTACGTCATAGGTGACGT
GGGGGGGGACCTACTTAAACACGGAAGTAGGCCCCGACACGTCACTGTCA
CGTGACAGTACGTCACAGCCGCCATTTTGTTTTACAAAATAGCCGACTTC
CTTCCTCTTTTTTAAAAAAAGGCGCCAAAAAACCGTCGGCGGGGGGGCCG
CGCGCTGCGCGCGCGGCCCCCGGGGGAGGCACAGCCTCCCCCCCCCGCGC
GCATGCGCGCGGGTCCCCCCCCCTCCGGGGGGCTCCGCCCCCCGGCCCCC
CCC (SEQ ID NO: 1014)

Annotations:

Putative Domain	*Base range*
TATA Box	86 – 90
Initiation Element	104 – 119
Transcriptional Start Site	114
5’ UTR Conserved Domain	174 – 244
ORF2	354 – 716
ORF2/2	354 – 712; 2372 – 2873

ORF2/3	354 – 712; 2565 – 3075
ORF2t/3	354 – 400; 2565 – 3075
TAIP	373 – 690
ORF1	590 – 2899
ORF1/1	590 – 712; 2372 – 2899
ORF1/2	590 – 712; 2565 – 2873
Three open-reading frame region	2551 – 2870
Poly(A)-Signal	3071 – 3076
GC-rich region	3733 – 3853

Table F6. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*)

Ring20 (*Alphatorquevirus* Clade 4)	
ORF2	MWTPPRNDQQYLNWQWYTSVLSSHSAMCGCSDAIAHLNHLANLLRAPQN PPPPDNPRPLPVRALPAPPAAHEAAGDRAPWPMGGGGDAGGAGAGGDADH GGAAGGPADADLLDAVAAAET (SEQ ID NO: 1015)
ORF2/2	MWTPPRNDQQYLNWQWYTSVLSSHSAMCGCSDAIAHLNHLANLLRAPQN PPPPDNPRPLPVRALPAPPAAHEAAGDRAPWPMGGGGDAGGAGAGGDADH GGAAGGPADADLLDAVAAAETLLEIPARNPTPRAIESLEAYKSLTRDTTHRN LPSMPGTSDVASLARKLFKECNNNQLLLNFFQQAARDPEGTQKCISPTKKRS KKKARFSPQSSSSEESPRGRTRNRSKAGRKAQRKRRRPSPSSSNSSCSSSESW ESNSDSCSTKSKKSNKIKISTLPCYQGGGI (SEQ ID NO: 1016)
ORF2/3	MWTPPRNDQQYLNWQWYTSVLSSHSAMCGCSDAIAHLNHLANLLRAPQN PPPPDNPRPLPVRALPAPPAAHEAAGDRAPWPMGGGGDAGGAGAGGDADH GGAAGGPADADLLDAVAAAETPQETQKGHRSVSVRPRKGAKRKLAFPPSQ APPKSPPVGGLGTGAKRVAKLRGRDGDPLPAAQTAAAAAAASLGSQTQTPV QPSPKNPTKSRYQPYLVTKGGGSSILLSGCTINMFPDPKPYCPSSNDWKEEY EACKYWDRPPRHNLRDPPFYPWAPKNNPCNVSFKLGFK (SEQ ID NO: 1017)
ORF2t/3	MWTPPRNDQQYLNWQWPQETQKGHRSVSVRPRKGAKRKLAFPPSQAPPKS PPVGGLGTGAKRVAKLRGRDGDPLPAAQTAAAAAAASLGSQTQTPVQPSPK NPTKSRYQPYLVTKGGGSSILLSGCTINMFPDPKPYCPSSNDWKEEYEACKY WDRPPRHNLRDPPFYPWAPKNNPCNVSFKLGFK (SEQ ID NO: 1018)

TAIP	MINNTLTGNGTLLYLAPTLLCAGVPTLSLILIILLICFVPRKIRPRLIIQDPYPCE HCLLPRLPTRQPVIEHHGLWVVEETPEALAQVETPTMEAPLEDPQTQTC (SEQ ID NO: 1019)
ORF1	MAYGWWRRRRRRWRRWRRRPWRRRWRTRRRRPARRRGRRRNVRRRRR GRWRRRYRRWKRKGRRRRKAKIIIRQWQPNYRRRCNIVGYLPILICGGNTV SRNYATHSDDTNYPGPFGGGMTTDKFSLRILYDEYKRFMNYWTASNEDLD LCRYLGCTFYFFRHPEVDFIIKINTMPPFLDTTITAPSIHPGLMALDKRARWIP SLKNRPGKKHYIKIRVGAPKMFTDKWYPQTDLCDMTLLTIYATAADMQYP FGSPLTDTVVVNSQVLQSMYDETISILPDEKTKRNSLLTSIRSYIPFYNTTQTI AQLKPFVDAGGHTTGSTTTTWGQLLNTTKFTTTTTTTYTYPGTTNTAVTFIT ANDTWYRGTAYKDNIKDVPQKAAQLYFQTTQKLLGNTFHGSDETLEYHAG LYSSIWLSPGRSYFETPGAYTDIKYNPFTDRGEGNMLWIDWLSKKNMKYDK VQSKCLVADLPLWAAAYGYVEFCSKSTGDTNIHMNARLLIRSPFTDPQLIVH TDPTKGFVPYSLNFGNGKMPGGSSNVPIRMRAKWYPTLSHQQEVLEALAQS GPFAYHSDIKKVSLGIKYRFKWIWGGNPVRQQVVRNPCKEPHSSGNRVPRSI QIVDPRYNSPELTIHAWDFRRGFFGPKAIQRMQQQPTATEFFSAGRKRPRRD TEVYQSDQEKEQKESSLFPPVKLLRRVPPWEDSEQEQSGSQSSEEETATLSQ QLKQQLQQQRVLGVKLRLLFNQVQKIQQNQDINPTLLPRGGDLVSFFQAVP (SEQ ID NO: 1020)
ORF1/1	MAYGWWRRRRRRWRRWRRRPWRRRWRTRRRRPARRRGRRRNVVRNPC KEPHSSGNRVPRSIQIVDPRYNSPELTIHAWDFRRGFFGPKAIQRMQQQPTAT EFFSAGRKRPRRDTEVYQSDQEKEQKESSLFPPVKLLRRVPPWEDSEQEQSG SQSSEEETATLSQQLKQQLQQQRVLGVKLRLLFNQVQKIQQNQDINPTLLPR GGDLVSFFQAVP (SEQ ID NO: 1021)
ORF1/2	MAYGWWRRRRRRWRRWRRRPWRRRWRTRRRRPARRRGRRRNAARDPEG TQKCISPTKKRSKKKARFSPQSSSSEESPRGRTRNRSKAGRKAQRKRRRPSPS SSNSSCSSSESWESNSDSCSTKSKKSNKIKISTLPCYQGGGI (SEQ ID NO: 1022)

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

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.

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).

Arginine-rich region

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

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:

(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 β-sheets;

(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) β-strands; and/or

(iii) the tertiary structure of the jelly-roll domain comprises at least two (e.g., at least 2, 3, or 4) β-sheets; and/or

(iv) the jelly-roll domain comprises a ratio of β-sheets to α-helices of at least 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1.

In certain embodiments, a jelly-roll domain comprises two β-sheets.

In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises about eight (e.g., 4, 5, 6, 7, 8, 9, 10, 11, or 12) β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises eight β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises seven β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises six β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises five β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises four β-strands.

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

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

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)

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 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

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.

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.

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	Ring1
Genus/Clade	*Alphatorquevirus*, Clade 3
Accession Number	AJ620231.1
Protein Accession Number	CAF05750.1

Full Sequence: 743 AA

```
1         10        20        30        40        50
|         |         |         |         |         |
MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRWRRGR
RKTRTYRRRRRFRRRGRKAKLIIKLWQPAVIKRCRIKGYIPLIISGNGTF
ATNFTSHINDRIMKGPFGGGHSTMRFSLYILFEEHLRHMNFWTRSNDNLE
LTRYLGASVKIYRHPDQDFIVIYNRRTPLGGNIYTAPSLHPGNAILAKHK
ILVPSLQTRPKGRKAIRLRIAPPTLFTDKWYFQKDIADLTLFNIMAVEAD
LRFPFCSPQTDNTCISFQVLSSVYNNYLSINTFNNDNSDSKLKEFLNKAF
PTTGTKGTSLNALNTFRTEGCISHPQLKKPNPQINKPLESQYFAPLDALW
GDPIYYNDLNENKSLNDIIEKILIKNMITYHAKLREFPNSYQGNKAFCHL
```

TGIYSPPYLNQGRISPEIFGLYTEIIYNPYTDKGTGNKVWMDPLTKENNI
YKEGQSKCLLTDMPLWTLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFP
KLYNEKVKDYGYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVM
EDISRSGPFAPKVEKPSTQLVMKYCFNFNWGGNPIIEQIVKDPSFQPTYE
IPGTGNIPRRIQVIDPRVLGPHYSFRSWDMRRHTFSRASIKRVSEQQETS
DLVFSGPKKPRVDIPKQETQEESSHSLQRESRPWETEEESETEALSQESQ
EVPFQQQLQQQYQEQLKLRQGIKVLFEQLIRTQQGVHVNPCLR
(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	MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRR RWRRGRRKTRTYRRRRRFRRRGRK (SEQ ID NO: 186)
Jelly-roll Domain	AKLIIKLWQPAVIKRCRIKGYIPLIISGNGTFATNFTSHINDRIMKGPFGG GHSTMRFSLYILFEEHLRHMNFWTRSNDNLELTRYLGASVKIYRHPDQ DFIVIYNRRTPLGGNIYTAPSLHPGNAILAKHKILVPSLQTRPKGRKAIRL RIAPPTLFTDKWYFQKDIADLTLFNIMAVEADLRFPFCSPQTDNTCISFQ VLSSVYNNYLSI (SEQ ID NO: 187)
Hypervariable domain	NTFNNDNSDSKLKEFLNKAFPTTGTKGTSLNALNTFRTEGCISHPQLKK PNPQINKPLESQYFAPLDALWGDPIYYNDLNENKSLNDIIEKILIKNMIT YHAKLREFPNSYQGNKAFCHLTGIYSPPYLNQGR (SEQ ID NO: 188)
N22	ISPEIFGLYTEIIYNPYTDKGTGNKVWMDPLTKENNIYKEGQSKCLLTD MPLWTLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFPKLYNEKVKD

	YGYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVMEDISRSG PFAPKVEKPSTQLVMKYCFNFN (SEQ ID NO: 189)
C-terminal domain	WGGNPIIEQIVKDPSFQPTYEIPGTGNIPRRIQVIDPRVLGPHYSFRSWD MRRHTFSRASIKRVSEQQETSDLVFSGPKKPRVDIPKQETQEESSHSLQR ESRPWETEEESETEALSQESQEVPFQQQLQQQYQEQLKLRQGIKVLFEQ LIRTQQGVHVNPCLR (SEQ ID NO: 190)

Table P. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Betatorquevirus*)

Name	Ring2
Genus/Clade	*Betatorquevirus*
Accession Number	JX134045.1
Protein Accession Number	AGG91484.1

Full Sequence: 666 AA

```
        1         10        20        30        40        50
        |         |         |         |         |         |
MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYTTIPLKQWQ
PPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSML
TLDALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTE
LPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQF
ENKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQ
NRNMSVDQGQSWPFKILGTQSFYFYFYTGANLPGDTTQIPVADLLPLTNP
RINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSL
KSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGED
TQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFADFQKNLKKVTNID
TNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQV
QYQLEAQNKLLQTGPFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVEN
PAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRISQ
DWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQ
QQQLYRERIISLLKDQ (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	MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVR (SEQ ID NO: 216)
Jelly-roll Domain	PTYTTIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPV HWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTNQDLPLVRYKGCKIT FYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTR RKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACNLQNPFVKP DKLSNNVTLWSLNT (SEQ ID NO: 217)
Hypervariable domain	ISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTGANLPGDTTQIPVADLL PLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLD MILYSLKSPEAIKNEWTTENMKWNQLNNAG (SEQ ID NO: 218)
N22	TMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELIL EGFPLWLIYWGFADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVI LNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFTPNI QGQLSDNISMFYKFYFK (SEQ ID NO: 219)
C-terminal domain	WGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRH GNYTTTALSRISQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEK KEKEVQQLISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 220)

Table D1. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Gammatorquevirus*)

Name	Ring 3.1

Genus/Clade *Gammatorquevirus*

Accession Number

Protein Accession Number

Full Sequence: 677 AA

```
          1         10         20         30         40         50
          |          |          |          |          |          |
MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRRRRKR
KKVRRKLKKITIKQWQPDSVKKCKIKGYSTLVMGAQGKQYNCYTNQASDY
VQPKAPQGGGFGCEVFNLKWLYQEYTAHRNIWTKTNEYTDLCRYTGAQII
LYRHPDVDFIVSWDNQPPFLLNKYTYPELQPQNLLLARRKRIILSQKSNP
KGKLRIKLRIPPPKQMITKWFFQRDFCDVNLFKLCASAASFRYPGISHGA
QSTIFSAYALNTDFYQCSDWCQTNTETGYLNIKTQQMPLWFHYREGGKEK
WYKYTNKEHRPYTNTYLKSISYNDGLFSPKAMFAFEVKAGGEGTTEPPQG
AQLIANLPLIALRYNPHEDTGHGNEIYLTSTFKGTYDKPKVTDALYFNNV
PLWMGFYGYWDFILQETKNKGVFDQHMFVVKCPALRPISQVTKQVYYPLV
DMDFCSGRLPFDEYLSKDIKSHWYPTAERQTVTINNFVTAGPYMPKFEPT
DKDSTWQLNYHYKFFFKWGGPQVTDPTVEDPCSRNKYPVPDTMQQTIQIK
NPEKLHPATLFHDWDLRRGFITQAAIKRMSENLQIDSSFESDGTESPKKK
KRCTKEIPTQNQKQEEIQECLLSLCEEPTCQEETEDLQLFIQQQQQQQYK
LRKNLFKLLTHLKKGQRISQLQTGLLE (SEQ ID NO: 919)
```

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 59
Jelly-roll domain	60 - 260
Hypervariable Region	261 - 356
N22	357 – 517
C-terminal Domain	518 – 677

Table D2. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Gammatorquevirus*)

Ring3.1 (*Gammatorquevirus*)	
Arg-Rich Region	MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRR RRKRKKVRRKLKK (SEQ ID NO: 920)
Jelly-roll Domain	ITIKQWQPDSVKKCKIKGYSTLVMGAQGKQYNCYTNQASDYVQPKAP QGGGFGCEVFNLKWLYQEYTAHRNIWTKTNEYTDLCRYTGAQIILYR HPDVDFIVSWDNQPPFLLNKYTYPELQPQNLLLARRKRIILSQKSNPKG

	KLRIKLRIPPPKQMITKWFFQRDFCDVNLFKLCASAASFRYPGISHGAQS TIFSAYAL (SEQ ID NO: 921)
Hypervariable domain	NTDFYQCSDWCQTNTETGYLNIKTQQMPLWFHYREGGKEKWYKYTN KEHRPYTNTYLKSISYNDGLFSPKAMFAFEVKAGGEGTTEPPQGAQLIA N (SEQ ID NO: 922)
N22	LPLIALRYNPHEDTGHGNEIYLTSTFKGTYDKPKVTDALYFNNVPLWM GFYGYWDFILQETKNKGVFDQHMFVVKCPALRPISQVTKQVYYPLVD MDFCSGRLPFDEYLSKDIKSHWYPTAERQTVTINNFVTAGPYMPKFEPT DKDSTWQLNYHYKFFFK (SEQ ID NO: 923)
C-terminal domain	WGGPQVTDPTVEDPCSRNKYPVPDTMQQTIQIKNPEKLHPATLFHDWD LRRGFITQAAIKRMSENLQIDSSFESDGTESPKKKKRCTKEIPTQNQKQE EIQECLLSLCEEPTCQEETEDLQLFIQQQQQQQYKLRKNLFKLLTHLKK GQRISQLQTGLLE (SEQ ID NO: 924)

Table R. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Gammatorquevirus*)

Name Ring4

Genus/Clade *Gammatorquevirus*

Accession Number

Protein Accession Number

Full Sequence: 662 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRKL
RKVKRKKKSLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLIN
VPPKTPYGGGFGVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLI
FYRDNKTDFVLSYDRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKP
NGRLTKKLKIKPPKQMLSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCN
ENPQVFFYYLNHGYYTITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNIN
IKSHEYEKSISYENGYFQSSFLQTQCIYTSERGEACIAEKPLGIAIYNPV
KDNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWLGLFGYLDYCRQIK
ADKTWLDSHVLVIQSPAIFTYPNPGAGKWYCPLSQSFINGNGPFNQPPTL
LQKAKWFPQIQYQQEIINSFVESGPFVPKYANQTESNWELKYKYVFTFKW
GGPQFHEPEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDYRR
GFIKERSLKRMSTYFSTHTDQQATSEEDIPKKKKRIGPQLTVPQQKEEET
LSCLLSLCKKDTFQETETQEDLQQLIKQQQEQQLLLKRNILQLIHKLKEN
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 (*Gammatorquevirus*)

Ring4 (*Gammatorquevirus*)	
Arg-Rich Region	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRR RRKLRKVKRKKK (SEQ ID NO: 926)
Jelly-roll Domain	SLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLINVPPKTPY GGGFGVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLIFYRDN KTDFVLSYDRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKPNGR LTKKLKIKPPKQMLSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCNE NPQVFFYYL (SEQ ID NO: 927)
Hypervariable domain	NHGYYTITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNINIKSHEYEK SISYENGYFQSSFLQTQCIYTSERGEACIAE (SEQ ID NO: 928)
N22	KPLGIAIYNPVKDNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWLG LFGYLDYCRQIKADKTWLDSHVLVIQSPAIFTYPNPGAGKWYCPLSQSF INGNGPFNQPPTLLQKAKWFPQIQYQQEIINSFVESGPFVPKYANQTESN WELKYKYVFTFK (SEQ ID NO: 929)
C-terminal domain	WGGPQFHEPEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDY RRGFIKERSLKRMSTYFSTHTDQQATSEEDIPKKKKRIGPQLTVPQQKE EETLSCLLSLCKKDTFQETETQEDLQQLIKQQQEQQLLLKRNILQLIHKL KENQQMLQLHTGMLP (SEQ ID NO: 930)

Table D5. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*) Clade 1

Name	Ring 5.2
Genus/Clade	*Alphatorquevirus* Clade 1
Accession Number	
Protein Accession Number	

Full Sequence: 728 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFVSRRWRRPYRR
RRRRGRRRRRRRRRHKPTLILRQWQPDCIRHCKITGWMPLIICGKGSTQF
NYITHADDITPRGASYGGNFTNMTFSLEAIYEQFLYHRNRWSASNHDLEL
CRYKGTTLKLYRHPEVDYIVTYSRTGPFEISHMTYLSTHPMLMLLNKHHI
VVPSLKTKPRGRKAIKVRIRPPKLMNNKWYFTRDFCNIGLFQLWATGLEL
RNPWLRMSTLSPCIGFNVLKNSIYTNLSNLPQYKNERLNIINNILHPQEI
TGTNNKKWQYTYTKLMAPIYYSANRASTYDWENYSKETNYNNTYVKFTQK
RQEKLTKIRKEWQMLYPQQPTALPDSYDLLQEYGLYSPYYLNPTRINLDW
MTPYTHVRYNPLVDKGFGNRIYIQWCSEADVSYNRTKSKCLLQDMPLFFM
CYGYIDWAIKNTGVSSLVKDARICIRCPYTEPQLVGSTEDIGFVPISETF
MRGDMPVLAPYIPLSWFCKWYPNIAHQKEVLESIISCSPFMPRDQDMNGW
DITIGYKMDFLWGGSPLPSQPIDDPCQQGTHPIPDPDKHPRLLQVSNPKL
LGPRTVFHKWDIRRGQFSKRSIKRVSEYSSDDESLAPGLPSKRNKLDSAF
RGENREQKECYSLLKALEEEETPEEEEPAPQEKAQKEELLHQLQLQRRHQ
RVLRRGLKLVFTDILRLRQGVHWNPELT (SEQ ID NO: 931)
```

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 66
Jelly-roll domain	67 - 277
Hypervariable Region	278 - 395
N22	396 – 561
C-terminal Domain	562 – 728

Table D6. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*) Clade 1

Ring5.2 (*Alphatorquevirus*) Clade 1	
Arg-Rich Region	TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFVSRRWRR PYRRRRRRGRRRRRRRRRHK (SEQ ID NO: 932)

Jelly-roll Domain	PTLILRQWQPDCIRHCKITGWMPLIICGKGSTQFNYITHADDITPRGASY GGNFTNMTFSLEAIYEQFLYHRNRWSASNHDLELCRYKGTTLKLYRHP EVDYIVTYSRTGPFEISHMTYLSTHPMLMLLNKHHIVVPSLKTKPRGRK AIKVRIRPPKLMNNKWYFTRDFCNIGLFQLWATGLELRNPWLRMSTLS PCIGFNVLKNSIYTNL (SEQ ID NO: 933)
Hypervariable domain	SNLPQYKNERLNIINNILHPQEITGTNNKKWQYTYTKLMAPIYYSANRA STYDWENYSKETNYNNTYVKFTQKRQEKLTKIRKEWQMLYPQQPTAL PDSYDLLQEYGLYSPYYLNPTR (SEQ ID NO: 934)
N22	INLDWMTPYTHVRYNPLVDKGFGNRIYIQWCSEADVSYNRTKSKCLL QDMPLFFMCYGYIDWAIKNTGVSSLVKDARICIRCPYTEPQLVGSTEDI GFVPISETFMRGDMPVLAPYIPLSWFCKWYPNIAHQKEVLESIISCSPFM PRDQDMNGWDITIGYKMDFL (SEQ ID NO: 935)
C-terminal domain	WGGSPLPSQPIDDPCQQGTHPIPDPDKHPRLLQVSNPKLLGPRTVFHKW DIRRGQFSKRSIKRVSEYSSDDESLAPGLPSKRNKLDSAFRGENREQKE CYSLLKALEEEETPEEEEPAPQEKAQKEELLHQLQLQRRHQRVLRRGL KLVFTDILRLRQGVHWNPELT (SEQ ID NO: 936)

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 sigal.

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.

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).

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.

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.

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 Figure 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.

In some embodiments, an ORF1 molecule 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

In some embodiments, a polypeptide (e.g., an ORF1 molecule) described herein comprises the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829), wherein X^n is a contiguous sequence of any *n*

amino acids. For example, X^2 indicates a contiguous sequence of any two amino acids. In some embodiments, the YNPX^{2}DXGX^{2}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^{2}DXGX^{2}N (SEQ ID NO: 829), wherein X^n is a contiguous sequence of any *n* amino acids.

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^{2}DXGX^{2}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^{2}DXGX^{2}N (SEQ ID NO: 829) motif. In some embodiments, the YNPX^{2}DXGX^{2}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^{2}DXGX^{2}N (SEQ ID NO: 829) motif.

Exemplary YNPX^{2}DXGX^{2}N (SEQ ID NO: 829) motif-flanking secondary structures are described in Example 47 and Figure 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 Figure 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 Figure 48 of PCT/US19/65995, flanking a YNPX^{2}DXGX^{2}N (SEQ ID NO: 829) motif (e.g., as described herein).

Conserved Secondary Structural Motif in ORF1 Jelly-Roll Domain

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 emboiments, an ORF1 molecule comprises one or more secondary structural elements comprised by the jelly-roll domain of an Anellovius 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/or a ninth beta strand.

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.

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.

Exemplary jelly-roll domain secondary structures are described in Example 47 of PCT/US19/65995 and FIG. 25 herein. 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. 25 herein.

Consensus ORF1 Domain Sequences

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_{a-b})" 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	LVLTQWQPNTVRRCYIRGYLPLIICGEN(X_{0-3})TTSRNYATHS DDTIQKGPFGGGMSTTTFSLRVLYDEYQRFMNRWTYSNED LDLARYLGCKFTFYRHPDXDFIVQYNTNPPFKDTKLTAPSIH P(X_{1-5})GMLMLSKRKILIPSLKTRPKGKHYVKVRIGPPKLFED KWYTQSDLCDVPLVXLYATAADLQHPFGSPQTDNPCVTFQ VLGSXYNKHLSISP; wherein X = any amino acid.	227
N22	SNFEFPGAYTDITYNPLTDKGVGNMVWIQYLTKPDTIXDKT QS(X_{0-3})KCLIEDLPLWAALYGYVDFCEKETGDSAIIXNXGRV LIRCPYTKPPLYDKT(X_{0-4})NKGFVPYSTNFGNGKMPGGSGY VPIYWRARWYPTLFHQKEVLEDIVQSGPFAYKDEKPSTQLV MKYCFNFN;	228

	wherein X = any amino acid.	
CTD	WGGNPISQQVVRNPCKDSG(X_{0-3})SGXGRQPRSVQVVDPKY MGPEYTFHSWDWRRGLFGEKAIKRMSEQPTDDEIFTGGXPK RPRRDPPTXQXPEE(X_{1-4})QKESSSFR(X_{2-14})PWESSSQEXESES QEEEE(X_{0-30})EQTVQQQLRQQLREQRRLRVQLQLLFQQLLKT (X_{0-4})QAGLHINPLLLSQA(X_{0-40})*; wherein X = any amino acid.	229

Table 37B. Betatorquevius ORF1 domain consensus sequences

Domain	Sequence	SEQ ID NO:
Jelly-Roll	LKQWQPSTIRKCKIKGYLPLFQCGKGRISNNYTQYKESIVPH HEPGGGGWSIQQFTLGALYEEHLKLRNWWTKSNDGLPLVR YLGCTIKLYRSEDTDYIVTYQRCYPMTATKLTYLSTQPSRM LMNKHKIIVPSKXT(X_{1-4})NKKKKPYKKIFIKPPSQMQNKWYF QQDIANTPLLQLTXTACSLDRMYLSSDSISNNITFTSLNTNFF QNPNFQ; wherein X = any amino acid.	230
N22	(X_{4-10})TPLYFECRYNPFKDKGTGNKVYLVSNN(X_{1-8})TGWDPP TDPDLIIEGFPLWLLLWGWLDWQKKLGKIQNIDTDYILVIQS XYYIPP(X_{1-3})KLPYYVPLDXD(X_{0-2})FLHGRSPY(X_{3-16})PSDKQH WHPKVRFQXETINNIALTGPGTPKLPNQKSIQAHMKYKFYF K; wherein X = any amino acid.	231
CTD	WGGCPAPMETITDPCKQPKYPIPNNLLQTTSLQXPTTPIETYL YKFDERRGLLTKKAAKRIKKDXTTETTLFTDTGXXTSTTLPT XXQTETTQEEXTSEEE(X_{0-5})ETLLQQLQQLRRKQKQLRXRIL QLLQLLXLL(X_{0-26})*; wherein X = any amino acid.	232

Table 37C. Gammatorquevius ORF1 domain consensus sequences

Domain	Sequence	SEQ ID NO:
Jelly-Roll	TIPLKQWQPESIRKCKIKGYGTLVLGAEGRQFYCYTNEKDE YTPPKAPGGGGFGVELFSLEYLYEQWKARNNIWTKSNXYK DLCRYTGCKITFYRHPTTDFIVXYSRQPPFEIDKXTYMXXHP QXLLLRKHKKIILSKATNPKGKLKKKIKIKPPKQMLNKWFF QKQFAXYGLVQLQAAACBLRYPRLGCCNENRLITLYYLN; wherein X = any amino acid.	233
N22	LPIVVARYNPAXDTGKGNKXWLXSTLNGSXWAPPTTDKDL IIEGLPLWLALYGYWSYJKKVKKDKGILQSHMFVVKSPAIQP LXTATTQXTFYPXIDNSFIQGKXPYDEPJTXNQKKLWYPTLE HQQETINAIVESGPYVPKLDNQKNSTWELXYXYTFYFK; wherein X = any amino acid.	234
CTD	WGGPQIPDQPVEDPKXQGTYPVPDTXQQTIQIXNPLKQKPE TMFHDWDYRRGIITSTALKRMQENLETDSSFXSDSEETP(X_{0-2})KKKKRLTXELPXPQEETEEIQSCLLSLCEESTCQEE(X_{1-6})ENL QQLIHQQQQQQQQLKHNILKLLSDLKZKQRLLQLQTGILE(X_{1-10})*; wherein X = any amino acid.	235

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

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:

(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.

(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. In some embodiments, a putative Anellovirus ORF1 sequence comprises the sequence Y[NCS]P*XX*D*X*[GASKR]*XX*[NTSVAK].

(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.

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.

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 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

In some embodiments, a polypeptide (e.g., an ORF2 molecule) described herein comprises the amino acid sequence [W/F]X^7HX^3CX^1CX^5H (SEQ ID NO: 949), wherein X^n is a contiguous sequence of any *n* amino acids. In embodiments, X^7 indicates a contiguous sequence of any seven amino acids. In embodiments, X^3 indicates a contiguous sequence of any three amino acids. In embodiments, X^1 indicates any single amino acid. In embodiments, X^5 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^7HX^3CX^1CX^5H (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^7HX^3CX^1CX^5H (SEQ ID NO: 949), wherein X^n is a contiguous sequence of any *n* amino acids.

Genetic Elements, e.g., genetic elements including non-Anellovirus sequences

In some embodiments, the anellovector comprises a genetic element. In some embodiments, the genetic element comprises a nucleic acid sequence (e.g., a contiguous nucleic acid sequence having a length of at least 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, or 4000 nucleotides) from a virus other than an Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the virus other than an Anellovirus is a Monodnavirus, e.g., a Shotokuvirus (e.g., a Cressdnaviricota [e.g., a redondovirus, circovirus {e.g., a porcine circovirus, e.g., PCV-1 or PCV-2; or beak-and-feather disease virus}, geminivirus {e.g., tomato golden mosaic virus}, or nanovirus {e.g., BBTV, MDV1, SCSVF, or FBNYV}]), or a Parvovirus (e.g., a dependoparavirus, e.g., a bocavirus or an AAV). In some embodiments, the virus other than Anellovirus is an AAV (e.g., AAV1, AAV2, or AAV5). In some embodiments, the nucleic acid sequence from the virus other than an Anellovirus comprises a non-Anellovirus origin of replication (e.g., an origin of replication derived from an AAV, e.g., AAV1, AAV2, or AAV5). In some embodiments, the non-Anellovirus origin of replication comprises an AAV Rep-binding motif (RBM), e.g., as described herein, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the non-Anellovirus origin of replication comprises an AAV terminal resolution site (TRS), e.g., as described herein, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the non-Anellovirus origin of

replication is derived from a virus that replicates by rolling circle replication. In some embodiments, the non-Anellovirus origin of replication is derived from a virus that replicates by rolling hairpin replication.

In some embodiments, the genetic element comprises one or more inverted terminal repeats (ITR). In some embodiments, the genetic element comprises one ITR. In some embodiments, the genetic element comprises an ITR positioned 5' relative to an effector or an effector-encoding sequence as described herein. In some embodiments, the genetic element comprises an ITR positioned 3' relative to an effector or an effector-encoding sequence as described herein. In some embodiments, the genetic element comprises two ITRs, e.g., flanking an effector or an effector-encoding sequence as described herein. In some embodiments, the non-Anellovirus origin of replication is comprised in an ITR, e.g., an AAV ITR, e.g., as described herein.

In some embodiments, a genetic element comprises an ITR sequence from an AAV (e.g., AAV1, AAV2, AAV3, AAV4, AAV5, or AAV6), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the AAV ITR has a sequence, e.g., as described in Grimm et al. (2005, *J. Virol.*, DOI: 10.1128/JVI.80.1.426-439.2006; incorporated herein by reference in its entirety), e.g., as shown in Figure 1A of Grimm et al., *supra*. In embodiments, the AAV ITR has a sequence as described herein Chiorini et al. (1999, *J. Virol* 73(5): 4293-4298; incorporated herein by reference in its entirety).

In some embodiments, a genetic element comprises a subsequence of an ITR sequence (e.g., from an AAV, e.g., as described herein), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the genetic element comprises the sequence of AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCC (SEQ ID NO: 1051), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the genetic element comprises the sequence of CGGGCGGGTGGTGGCGGCGGTTGGGGCTCGGCGCTCGCTCGCTCGCTGGGCGGGCGGGCGG T (SEQ ID NO: 1052, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

In some embodiments, a genetic element comprises an RBM sequence (e.g., from an AAV, e.g., as described herein), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the genetic element comprises the sequence of (GMGY)x4 (SEQ ID NO: 1053), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the genetic element comprises the sequence of (GMGY)x5 (SEQ ID NO: 1054), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the genetic element

comprises the sequence of GCGCGCTCGCTCGCTC (SEQ ID NO: 1055, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the genetic element comprises the sequence of GCTCGCTCGCTCGCTG (SEQ ID NO: 1056, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

In some embodiments, a genetic element comprises a TRS sequence (e.g., from an AAV, e.g., as described herein), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the genetic element comprises the sequence of XGTTGG (SEQ ID NO: 1057 (wherein X is selected from G, C, T, or A), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the genetic element comprises the sequence of AGTTGG (SEQ ID NO: 1058, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the genetic element comprises the sequence of GGTTGG (SEQ ID NO: 1059, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

In some embodiments, genetic element construct (e.g., as described herein) comprises a nucleic acid sequence having a structure as shown in Table 61 below, or as diagrammed in Figure 10.

In some embodiments, a genetic element (e.g., as described herein) comprises a nucleic acid sequence having a structure as shown in Table 61 below, or as diagrammed in Figure 10. In embodiments, a genetic element comprises 1, 2, or all of: (i) one or more (e.g., one or two) non-Anellovirus (e.g., AAV) ITR sequences; (ii) a sequence encoding an exogenous effector; and/or (iii) a sequence (e.g., a contiguous or non-contiguous sequence) from an Anellovirus genome (or a sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto), or a contiguous portion thereof having a length of at least 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, or 4000 nucleotides.

In embodiments, the genetic element comprises a non-Anellovirus (e.g., AAV) ITR sequence positioned within the Anellovirus genome, or the portion thereof. In an embodiment, the non-Anellovirus ITR sequence is positioned closer to the 5' end of the Anellovirus genome sequence, or the portion thereof, than to the 3' end of the Anellovirus genome sequence, or the portion thereof. In an embodiment, the non-Anellovirus ITR sequence is positioned closer to the 3' end of the Anellovirus genome sequence, or the portion thereof, than to the 5' end of the Anellovirus genome sequence, or the portion thereof.

In embodiments, the genetic element comprises a non-Anellovirus (e.g., AAV) ITR sequence positioned at the 5' end of the Anellovirus genome sequence, or the portion thereof. In embodiments, the

genetic element comprises a non-Anellovirus (e.g., AAV) ITR sequence positioned at the 3' end of the Anellovirus genome sequence, or the portion thereof.

In embodiments, the non-Anellovirus ITR sequence shares the same orientation as the Anellovirus genome sequence, or the portion thereof. In embodiments, the non-Anellovirus ITR sequence has the reverse orientation from the Anellovirus genome sequence, or the portion thereof.

In embodiments, the genetic element comprises a sequence encoding an effector (e.g., an endogenous effector or an exogenous effector). In embodiments, the sequence encoding the effect is positioned upstream of the non-Anellovirus ITR sequence. In embodiments, the sequence encoding the effect is positioned downsteam of the non-Anellovirus ITR sequence.

In embodiments, the genetic element comprises a plurality of (e.g., two) non-Anellovirus ITR sequences. In embodiments, the plurality of non-Anellovirus ITR sequences share the same sequence. In embodiments, the plurality of non-Anellovirus ITR sequences have different sequences. In embodiments, the plurality of non-Anellovirus ITR sequences share at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% nucleic acid sequence identity. In embodiments, the genetic element comprises two non-Anellovirus ITR sequences that share the same orientation. In embodiments, the genetic element comprises two non-Anellovirus ITR sequences that have opposite orientations. In embodiments, the genetic element comprises a sequence encoding an effector (e.g., an endogenous effector or an exogenous effector), wherein the sequence encoding the effector shares the same orientation as one or more of the non-Anellovirus ITR sequences. In embodiments, the genetic element comprises a sequence encoding an effector (e.g., an endogenous effector or an exogenous effector), wherein the sequence encoding the effector is in the opposite orientation as one or more of the non-Anellovirus ITR sequences.

Table 61. Exemplary AAV-Anellovirus genetic element structures

Plasmid Number	***Plasmid Name***	***Description***	***Schematic***	***Replication /Packaging?***
pRTx-1260	pRing2-5'NCR-AAV-Ori-FWD	Ring2 with single AAV-TRS-RBM in the 5'NCR in FWD orientation, for IVC to make positive strand	FIG. 10A	Yes
pRTx-1261	pRing2-5'NCR-AAV-Ori-Rev	Ring2 with single AAV-TRS-RBM in the 5'NCR in Rev orientation, for IVC to make negative strand	FIG. 10B	Yes

pRTx-1262	pRing2-3'NCR-AAV-Ori-FWD	Ring2 with single AAV-TRS-RBM in the 3'NCR in FWD orientation, for IVC to make positive strand	FIG. 10C	Yes
pRTx-1263	pRing2-3'NCR-AAV-Ori-Rev	Ring2 with single AAV-TRS-RBM in the 3'NCR in Rev orientation, for IVC to make negative strand	FIG. 10D	Yes
In progress	pRing2-5'AAV-Ori	Ring2 with single AAV-TRS-RBM before the 5'NCR in FWD orientation, for IVC to make positive strand	FIG. 10E	
In progress	pRing2-5'AAV-Ori-Rev	Ring2 with single AAV-TRS-RBM before the 5'NCR in Rev orientation, for IVC to make negative strand	FIG. 10F	
In progress	pRing2ΔORF::hEF1a_EGFP-5'AAV-Ori	Ring2 vector with ORFs replaced by a hEF1a-EGFP gene, with single AAV-TRS-RBM before the 5'NCR in FWD orientation, for IVC to make positive strand	FIG. 10G	
In progress	pRing2ΔORF::hEF1a_EGFP-5'AAV-Ori-Rev	Ring2 vector with ORFs replaced by a hEF1a-EGFP gene, with single AAV-TRS-RBM before the 5'NCR in Rev orientation, for IVC to make negative strand	FIG. 10H	
In progress	pRing2-2xAAV-Ori	Ring2 flanked by AAV-TRS-RBM in FWD orientations, to make	FIG. 10I	

		positive strand off of a plasmid		
In progress	pRing2-2xAAV-Ori-Rev	Ring2 flanked by AAV-TRS-RBM in Rev orientations, to make negative strand off of a plasmid	FIG. 10J	
pRTx-1472	pRing2ΔORF::hEF1a_EGFP-2xAAV-Ori	Ring2 vector with ORFs replaced by a hEF1a-EGFP gene flanked by AAV-TRS-RBM in FWD orientations, to make positive strand off of a plasmid	FIG. 10K	Yes
In progress	pRing2ΔORF::hEF1a_EGFP-2xAAV-Ori-Rev	Ring2 vector with ORFs replaced by a hEF1a-EGFP gene flanked by AAV-TRS-RBM in FWD orientations, to make negative strand off of a plasmid	FIG. 10L	

In some embodiments, the genetic element is capable of undergoing replication in the presence of a non-Anellovirus Rep molecule, e.g., a Rep protein from a Monodnavirus, e.g., a Shotokuvirus (e.g., a Cressdnaviricota [e.g., a redondovirus, circovirus {e.g., a porcine circovirus, e.g., PCV-1 or PCV-2; or beak-and-feather disease virus}, geminivirus {e.g., tomato golden mosaic virus}, or nanovirus {e.g., BBTV, MDV1, SCSVF, or FBNYV}]), or a Parvovirus (e.g., a dependoparavirus, e.g., a bocavirus or an AAV); or a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the genetic element is capable of undergoing replication in the presence of an AAV Rep molecule, e.g., an AAV Rep protein (e.g., an AAV1, AAV2, or AAV5 Rep protein), or a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

In some embodiments, the genetic element is linear. In some embodiments, the genetic element is circular. In some embodiments, the genetic element is single-stranded. In some embodiments, the genetic element is double-stranded. In some embodiments, the genetic element consists at least of

75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% DNA. In some embodiments, the genetic element is 100% DNA.

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 DNA, 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 DNA. 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. In certain embodiments, the genetic element is a circular, single stranded DNA that comprises a promoter sequence, a sequence encoding a therapeutic effector, and a capsid binding protein.

In some embodiments, the genetic element has a length less than 20kb (e.g., less than about 19kb, 18kb, 17kb, 16kb, 15kb, 14kb, 13kb, 12kb, 11kb, 10kb, 9kb, 8kb, 7kb, 6kb, 5kb, 4kb, 3kb, 2kb, 1kb, or less). In some embodiments, the genetic element has, independently or in addition to, a length greater than 1000b (e.g., at least about 1.1kb, 1.2kb, 1.3kb, 1.4kb, 1.5kb, 1.6kb, 1.7kb, 1.8kb, 1.9kb, 2kb, 2.1kb, 2.2kb, 2.3kb, 2.4kb, 2.5kb, 2.6kb, 2.7kb, 2.8kb, 2.9kb, 3kb, 3.1kb, 3.2kb, 3.3kb, 3.4kb, 3.5kb, 3.6kb, 3.7kb, 3.8kb, 3.9kb, 4kb, 4.1kb, 4.2kb, 4.3kb, 4.4kb, 4.5kb, 4.6kb, 4.7kb, 4.8kb, 4.9kb, 5kb, 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.

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.

In embodiments, the genetic element was produced from a double-stranded circular DNA (e.g., produced by in vitro circularization). In some embodiments, the genetic element was produced by rolling circle replication from the double-stranded circular DNA. In embodiments, the rolling circle replication occurs in a cell (e.g., a host cell, e.g., a mammalian cell, e.g., a human cell, e.g., a HEK293T cell, an A549 cell, or a Jurkat cell). In embodiments, the genetic element can be amplified exponentially by rolling circle replication in the cell. In embodiments, the genetic element can be amplified linearly by rolling circle replication in the cell. In embodiments, the double-stranded circular DNA or genetic element is capable of yielding at least 2, 4, 8, 16, 32, 64, 128, 256, 518, 1024 or more times the original quantity by rolling circle replication in the cell. In embodiments, the double-stranded circular DNA was introduced into the cell, e.g., as described herein.

In some embodiments, the double-stranded circular DNA and/or 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 double-stranded circular DNA and/or the genetic element does not comprise a bacterial plasmid backbone.

In one embodiment, the invention includes a genetic element comprising a nucleic acid sequence (e.g., a DNA 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

A strategy employed by many viruses is that the viral capsid protein recognizes a specific protein binding sequence in its genome. For example, in viruses with unsegmented genomes, such as the L-A virus of yeast, there is a secondary structure (stem-loop) and a specific sequence at the 5' end of the genome that are both used to bind the viral capsid protein. However, viruses with segmented genomes, such as *Reoviridae*, *Orthomyxoviridae* (influenza), *Bunyaviruses* and *Arenaviruses,* need to package each of the genomic segments. Some viruses utilize a complementarity region of the segments to aid the virus in including one of each of the genomic molecules. Other viruses have specific binding sites for each of

the different segments. See for example, Curr Opin Struct Biol. 2010 Feb; 20(1): 114–120; and Journal of Virology (2003), 77(24), 13036-13041.

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.

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.

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

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.

In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGGGX$_1$CAGTCT, 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 AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGGGX$_1$CAGTCT, 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$_1$ is A. In embodiments, X$_1$ is absent.

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

AGGTGAGTTTACACACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGC, 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 AGGTGAGTTTACACACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGC, 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 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 AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGATCAGTCT, 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 AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGATCAGTCT, 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 some embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR of an 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 nucleic acid molecule comprises a nucleic acid sequence having at least 97.183% sequence identity to the 5' UTR conserved domain listed in Table C1. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCT, 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 AGGTGAGTGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCT, 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 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_1, X_2, X_3, X_4, and X_5 are each independently any nucleotide, e.g., wherein X_1 = G or T, X_2 = C or A, X_3 = G or A, X_4 = T or C, and X_5 = 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-tth8 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 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	CGGGTGCCGX_1AGGTGAGTTTACACACCGX_2AGT CAAGGGGCAATTCGGGCTCX_3GGACTGGCCGGG CX_4X_5TGGG X_1 = G or T X_2 = C or A X_3 = G or A X_4 = T or C X_5 = A, C, or T	105
Exemplary TTV Sequence	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT WTGGG	106
TTV-CT30F	CGGGTGCCGTAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	107

TTV-HD23a	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CTGGG	108
TTV-JA20	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT TTGGG	109
TTV-TJN02	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	110
TTV-tth8	CGGGTGCCGGAGGTGAGTTTACACACCGAAGTC AAGGGGCAATTCGGGCTCAGGACTGGCCGGGCT TTGGG	111
Alphatorquevirus Consensus 5' UTR	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGC X_1X_2TGGG; wherein X_1 comprises T or C, and wherein X_2 comprises A, C, or T.	112
Alphatorquevirus Clade 1 5' UTR (e.g., TTV-CT30F)	CGGGTGCCGTAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	113
Alphatorquevirus Clade 2 5' UTR (e.g., TTV-P13-1)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CGGG	114
Alphatorquevirus Clade 3 5' UTR (e.g., TTV-tth8)	CGGGTGCCGGAGGTGAGTTTACACACCGAAGTC AAGGGGCAATTCGGGCTCAGGACTGGCCGGGCT TTGGG	115
Alphatorquevirus Clade 4 5' UTR (e.g., TTV-HD20a)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGAGGCCGGGCCAT GGG	116
Alphatorquevirus Clade 5 5' UTR (e.g., TTV-16)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CCGGG	117

Alphatorquevirus Clade 6 5' UTR (e.g., TTV-TJN02)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	118
Alphatorquevirus Clade 7 5' UTR (e.g., TTV-HD16d)	CGGGTGCCGAAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	119

Identification of 5' UTR sequences

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:

(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).

(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

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.

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).

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-tth8, 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-tth8, TTV-HD20a, TTV-16, TTV-TJN02, or TTV-HD16d, e.g., as listed in Table 39.

In embodiments, the 36-nucleotide GC-rich sequence is selected from:

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

(ii) GCGCTX$_1$CGCGCGCGCGCCGGGGGGCTGCGCCCCCCC (SEQ ID NO: 164),

wherein X_1 is selected from T, G, or A;

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

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

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

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

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

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

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

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

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

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_1, X_4, X_5, X_6, X_7, X_{12}, X_{13}, X_{14}, X_{15}, X_{20}, X_{21}, X_{22}, X_{26}, X_{29}, X_{30}, and X_{33} are each independently any nucleotide and wherein X_2, X_3, X_8, X_9, X_{10}, X_{11}, X_{16}, X_{17}, X_{18}, X_{19}, X_{23}, X_{24}, X_{25}, X_{27}, X_{28}, X_{31}, X_{32}, and X_{34} are each independently absent or any nucleotide. In some embodiments, one or more of (e.g., all of) X_1 through X_{34} 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	$CGGCGGX_1GGX_2GX_3X_4X_5CGCGCTX_6CGCGC$ $GCX_7X_8X_9X_{10}CX_{11}X_{12}X_{13}X_{14}GGGGX_{15}X_{16}X_{17}X_{18}$ $X_{19}X_{20}X_{21}GCX_{22}X_{23}X_{24}X_{25}CCCCCCCCX_{26}CGCGC$ $ATX_{27}X_{28}GCX_{29}CGGGX_{30}CCCCCCCCCCX_{31}X_{32}X$ $_{33}GGGGGGCTCCGX_{34}CCCCCCGGCCCCCC$ X_1 = G or C X_2 = G, C, or absent X_3 = C or absent X_4 = G or C X_5 = G or C X_6 = T, G, or A	120

		X_7 = G or C X_8 = G or absent X_9 = C or absent X_{10} = C or absent X_{11} = G, A, or absent X_{12} = G or C X_{13} = C or T X_{14} = G or A X_{15} = G or A X_{16} = A, G, T, or absent X_{17} = G, C, or absent X_{18} = G, C, or absent X_{19} = C, A, or absent X_{20} = C or A X_{21} = T or A X_{22} = G or C X_{23} = G, T, or absent X_{24} = C or absent X_{25} = G, C, or absent X_{26} = G or C X_{27} = G or absent X_{28} = C or absent X_{29} = G or A X_{30} = G or T X_{31} = C, T, or absent X_{32} = G, C, A, or absent X_{33} = G or C X_{34} = C or absent	
Exemplary TTV Sequence	Full sequence	GCCGCCGCGGCGGCGGSGGNGNSGCGCGCT DCGCGCGCSNNNCRCCRGGGGGNNNNCWG CSNCNCCCCCCCCCGCGCATGCGCGGGKCC CCCCCCCNNCGGGGGGCTCCGCCCCCCGGC CCCCCCCCGTGCTAAACCCACCGCGCATGC GCGACCACGCCCCCGCCGCC	121

	Fragment 1	GCCGCCGCGGCGGCGGSGGNGNSGCGCGCT DCGCGCGCSNNNCRCCRGGGGGNNNNCWG CSNCNCCCCCCCCCGCGCAT	122
	Fragment 2	GCGCGGGKCCCCCCCCCNNCGGGGGGCTC CG	123
	Fragment 3	CCCCCCGGCCCCCCCCCGTGCTAAACCCAC CGCGCATGCGCGACCACGCCCCCGCCGCC	124
TTV-CT30F	Full sequence	GCGGCGG-GGGGGCG-GCCGCG-TTCGCGCGCCGCCCACCAGGGGGTG--CTGCG-CGCCCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCCC--GGGGGGGCTCCGCCCCCCCCGGCCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCCCGCCGCC	125
	Fragment 1	GCGGCGG	126
	Fragment 2	GGGGGCG	127
	Fragment 3	GCCGCG	128
	Fragment 4	TTCGCGCGCCGCCCACCAGGGGGTG	129
	Fragment 5	CTGCG	130
	Fragment 6	CGCCCCCCCCCCGCGCAT	131
	Fragment 7	GCGCGGGGCCCCCCCCCC	132
	Fragment 8	GGGGGGGCTCCGCCCCCCCCGGCCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCCCGCCGCC	133
TTV-HD23a	Full sequence	CGGCGGCGGCGGCG-CGCGCGCTGCGCGCGCG---CGCCGGGGGGGCGCCAGCG-CCCCCCCCCCCCGCGCAT GCACGGGTCCCCCCCCCCCACGGGGGGCTCC G CCCCCCCGGCCCCCCCCC	134
	Fragment 1	CGGCGGCGGCGGCG	135
	Fragment 2	CGCGCGCTGCGCGCGCG	136
	Fragment 3	CGCCGGGGGGGCGCCAGCG	137

	Fragment 4	CCCCCCCCCCCGCGCAT	138
	Fragment 5	GCACGGGTCCCCCCCCCCACGGGGGGCTCCG	139
	Fragment 6	CCCCCCGGCCCCCCCCC	140
TTV-JA20	Full sequence	CCGTCGGCGGGGGGGCCGCGCGCTGCGCGCGGCCC-CCGGGGGAGGCACAGCCTCCCCCCCCCGCGCGCATGCGCGCGGGTCCCCCCCCCCTCCGGGGGGCTCCGCCCCCCGGCCCCCCCCC	141
	Fragment 1	CCGTCGGCGGGGGGGCCGCGCGCTGCGCGCGGCCC	142
	Fragment 2	CCGGGGGAGGCACAGCCTCCCCCCCCCGCGCGCATGCGCGCGGGTCCCCCCCCCCTCCGGGGGGCTCCGCCCCCCGGCCCCCCCCC	143
TTV-TJN02	Full sequence	CGGCGGCGGCG-CGCGCGCTACGCGCGCG---CGCCGGGGGG----CTGCCGC-CCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCC-GCGGGGGGCTCCG CCCCCCGGCCCCCC	144
	Fragment 1	CGGCGGCGGCG	145
	Fragment 2	CGCGCGCTACGCGCGCG	146
	Fragment 3	CGCCGGGGGG	147
	Fragment 4	CTGCCGC	148
	Fragment 5	CCCCCCCCCGCGCAT	149
	Fragment 6	GCGCGGGGCCCCCCCCC	150
	Fragment 7	GCGGGGGGCTCCG	151
	Fragment 8	CCCCCCGGCCCCCC	152
TTV-tth8	Full sequence	GCCGCCGCGGCGGCGGGGG-GCGGCGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGCG---CCCCCCCCCCGCGCATGCGCGGGGCCCCCCCCCC-GCGGGGGGCTCCG CCCCCCGGCCCCCCCCG	153

	Fragment 1	GCCGCCGCGGCGGCGGGGG	154
	Fragment 2	GCGGCGCGCTGCGCGCGCCGCCCAGTAGG GGGAGCCATGCG	155
	Fragment 3	CCCCCCCCCGCGCAT	156
	Fragment 4	GCGCGGGGCCCCCCCCC	157
	Fragment 5	GCGGGGGGCTCCG	158
	Fragment 6	CCCCCCGGCCCCCCCG	159
	Fragment 7	CGCGCTGCGCGCGCCGCCCAGTAGGGGGA GCCATGC	160
	Fragment 8	CCGCCATCTTAAGTAGTTGAGGCGGACGGT GGCGTGAGTTCAAAGGTCACCATCAGCCAC ACCTACTCAAAATGGTGG	161
	Fragment 9	CTTAAGTAGTTGAGGCGGACGGTGGCGTGA GTTCAAAGGTCACCATCAGCCACACCTACT CAAAATGGTGGACAATTTCTTCCGGGTCAA AGGTTACAGCCGCCATGTTAAAACACGTGA CGTATGACGTCACGGCCGCCATTTTGTGAC ACAAGATGGCCGACTTCCTTCC	162
Additional GC-rich Sequences	36-nucleotide consensus GC-rich region sequence 1	CGCGCTGCGCGCGCCGCCCAGTAGGGGGA GCCATGC	163
	36-nucleotide region consensus sequence 2	GCGCTX$_1$CGCGCGCGCGCCGGGGGGCTGCG CCCCCCC, wherein X$_1$ is selected from T, G, or A	164
	TTV Clade 1 36-nucleotide region	GCGCTTCGCGCGCCGCCCACTAGGGGGCGT TGCGCG	165
	TTV Clade 3 36-nucleotide region	GCGCTGCGCGCGCCGCCCAGTAGGGGGCG CAATGCG	166

	TTV Clade 3 isolate GH1 36-nucleotide region	GCGCTGCGCGCGCGGCCCCCGGGGGAGGC ATTGCCT	167
	TTV Clade 3 sle1932 36-nucleotide region	GCGCTGCGCGCGCGCCGGGGGGGCGCC AGCGCCC	168
	TTV Clade 4 ctdc002 36-nucleotide region	GCGCTTCGCGCGCGCCGGGGGGCTCCGC CCCCCC	169
	TTV Clade 5 36-nucleotide region	GCGCTTCGCGCGCGCCGGGGGGCTGCGC CCCCCC	170
	TTV Clade 6 36-nucleotide region	GCGCTACGCGCGCGCCGGGGGGCTGCG CCCCCCC	171
	TTV Clade 7 36-nucleotide region	GCGCTACGCGCGCGCCGGGGGGCTCTGC CCCCCC	172
Additional Alphatorquevirus GC-rich region sequences	TTV-CT30F	GCGGCGGGGGGGCGGCCGCGTTCGCGCGC CGCCCACCAGGGGGTGCTGCGCGCCCCCCC CCGCGCATGCGCGGGGCCCCCCCCCGGGG GGGCTCCGCCCCCCCGGCCCCCCCCCGTGC TAAACCCACCGCGCATGCGCGACCACGCCC CCGCCGCC	801
	TTV-P13-1	CCGAGCGTTAGCGAGGAGTGCGACCCTACC CCCTGGGCCCACTTCTTCGGAGCCGCGCGC TACGCCTTCGGCTGCGCGCGGCACCTCAGA CCCCCGCTCGTGCTGACACGCTTGCGCGTG TCAGACCACTTCGGGCTCGCGGGGGTCGGG	802

	TTV-tth8	GCCGCCGCGGCGGCGGGGGGCGGCGCGCT GCGCGCGCCGCCCAGTAGGGGGAGCCATG CGCCCCCCCCCCGCGCATGCGCGGGGCCCCC CCCCGCGGGGGGCTCCGCCCCCCGGCCCCC CCCG	803
	TTV-HD20a	CGGCCCAGCGGCGGCGCGCGCGCTTCGCGC GCGCGCCGGGGGGCTCCGCCCCCCCCCGCG CATGCGCGGGGCCCCCCCCCGCGGGGGGCT CCGCCCCCCGGTCCCCCCCCG	804
	TTV-16	CGGCCGTGCGGCGGCGCGCGCGCTTCGCGC GCGCGCCGGGGGCTGCCGCCCCCCCCCGCG CATGCGCGCGGGGCCCCCCCCCCGCGGGGG GCTCCGCCCCCCCGGCCCCCCCCCCCG	805
	TTV-TJN02	CGGCGGCGGCGCGCGCGCTACGCGCGCGC GCCGGGGGGCTGCCGCCCCCCCCCCGCGCA TGCGCGGGGCCCCCCCCCCGCGGGGGGCTCC GCCCCCCCGGCCCCCC	806
	TTV-HD16d	GGCGGCGGCGCGCGCGCTACGCGCGCGCG CCGGGGAGCTCTGCCCCCCCCCCGCGCATGC GCGCGGGTCCCCCCCCCCGCGGGGGGCTCCG CCCCCCGGTCCCCCCCCCG	807

Effectors

In some embodiments, the genetic element may include one or more sequences that 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 degradation. 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.

In some embodiments, the sequence encoding an effector is part of the genetic element, e.g., it can be inserted at an insert site as described herein. In embodiments, the sequence encoding an effector is inserted into the genetic element at a noncoding region, e.g., a noncoding region disposed 3' of the open reading frames and 5' of the GC-rich region of the genetic element, in the 5' noncoding region upstream of the TATA box, in the 5' UTR, in the 3' noncoding region downstream of the poly-A signal, or upstream of the GC-rich region. In embodiments, the sequence encoding an effector is inserted into the genetic element at about nucleotide 3588 of a TTV-tth8 plasmid, e.g., as described herein or at about nucleotide 2843 of a TTMV-LY2 plasmid, e.g., as described herein. In embodiments, the sequence encoding an effector is inserted into the genetic element at or within nucleotides 336-3015 of a TTV-tth8 plasmid, e.g., as described herein, or at or within nucleotides 242-2812 of a TTV-LY2 plasmid, e.g., as described herein. In some embodiments, the sequence encoding an effector replaces part or all of an open reading frame (e.g., an ORF as described herein, e.g., an ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3).

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

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.

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.

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.

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).

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 about18-25 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).

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.

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.

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%.

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.

In some embodiments, the regulatory nucleic acid is at least one miRNA, e.g., 2, 3, 4, 5, 6, or more. In some embodiments, the genetic element comprises a sequence that encodes an miRNA at least about 75%, 80%, 85%, 90% 95%, 96%, 97%, 98%, 99% or 100% nucleotide sequence identity to any one of the nucleotide sequences or a sequence that is complementary to a sequence described herein, e.g., in Table 40.

Table 40: Examples of regulatory nucleic acids, e.g., miRNAs.

Accession number of strain	Exemplary subsequence nucleotides	Pre_miRNA	SEQ ID NO:	miRNA_5prime_per_MiRdup	SEQ ID NO:	miRNA_3prime_per_MiRdup	SEQ ID NO:
AB008394.1	AB008394_3475_3551	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUAAAGGUUAAAGGUCAUCC	300	AGUAGCUGACGUCAAGGAUUGAC(5')	395	CAUCCUCGGCGGAAGCUACACAA(3')	490

		UCGGCGGAAGCUA CACAAAAUGGU					
AB008394.1	AB008394_3579_3657	GCGUACGUCACAA GUCACGUGGAGGG GACCCGCUGUAAC CCGGAAGUAGGCC CCGUCACGUGACU UACCACGUGUGUA	301	CAAGUCACGU GGAGGGGACC CG(5')	396	GGCCCCGUCA CGUGACUUAC CAC(3')	491
AB017613.1	AB017613_3462_3539	GCCAUUUUAAGUA GCUGACGUCAAGG AUUGACGUGAAGG UUAAAGGUCAUCC UCGGCGGAAGCUA CACAAAAUGGUG	302	AAGUAGCUGA CGUCAAGGAU UGACG(5')	397	UCAUCCUCGG CGGAAGCUAC ACAA(3')	492
AB017613.1	AB017613_3566_3644	GCACACGUCAUAA GUCACGUGGUGGG GACCCGCUGUAAC CCGGAAGUAGGCC CCGUCACGUGAUU UGUCACGUGUGUA	303	AUAAGUCACG UGGUGGGGAC CCG(5')	398	GGCCCCGUCA CGUGAUUUGU CAC(3')	493
AB025946.1	AB025946_3534_3600	CUUCCGGGUCAUA GGUCACACCUACG UCACAAGUCACGU GGGGAGGGUUGGC GUAUAGCCCGGAA G	304	UGGGGAGGGU UGGCGUAUAG CCCGGA(3')	399	CCGGGUCAUA GGUCACACCU ACGUCAC(5')	494
AB025946.1	AB025946_3730_3798	GCCGGGGGGCUGC CGCCCCCCCCGGG GAAAGGGGGGGGC CCCCCCCGGGGGG GGGUUUGCCCCCC GGC	305	CCCCCCCCGG GGGGGGGUUU GCCC(3')	400	GGCUGCCGCC CCCCCCGGGG AAAGGGGG(5')	495
AB028668.1	AB028668_3537_3615	AUACGUCAUCAGU CACGUGGGGGAAG GCGUGCCUAAACC CGGAAGCAUCCUC GUCCACGUGACUG UGACGUGUGUGGC	306	AUCAGUCACG UGGGGGAAGG CGUGC(5')	401	AUCCUCGUCC ACGUGACUGU GA(3')	496
AB028669.1	AB028669_3440_3513	CAUUUUAAGUAAG GCGGAAGCAGCUC GGCGUACACAAAA UGGCGGCGGAGCA CUUCCGGCUUGCC CAAAAUGG	307	AAGUAAGGCG GAAGCAGCUC GG(5')	402	GAGCACUUCC GGCUUGCCCA A(3')	497

AB028669.1	AB028669_3548_3619	GUCACAAGUCACGUGGGGAGGGUUGGCGUUUAACCCGGAAGCCAAUCCUCUUACGUGGCCUGUCACGUGAC	308	AGUCACGUGGGGAGGGUUGGC(5')	403	CAAUCCUCUUACGUGGCCUG(3')	498
AB037926.1	AB037926_162_232	CGACCGCGUCCCGAAGGCGGGUACCCGAGGUGAGUUUACACACCGAGGUUAAGGGCCAAUUCGGGCUUGG	309	CCCGAAGGCGGGUACCCGAGGU(5')	404	CGAGGUUAAGGGCCAAUUCGGGCU(3')	499
AB037926.1	AB037926_3454_3513	CGCGGUAUCGUAGCCGACGCGGACCCCGUUUUCGGGGCCCCCGCGGGGCUCUCGGCGCG	310	UAUCGUAGCCGACGCGGACCCCG(5')	405	GGGCCCCCGCGGGGCUCUCGGCG(3')	500
AB037926.1	AB037926_3531_3609	CGCCAUUUUGUGAUACGCGCGUCCCCUCCCGGCUUCCGUACAACGUCAGGCGGGGCGUGGCCGUAUCAGAAAAUGGCG	311	AUUUUGUGAUACGCGCGUCCCCUCCC(5')	406	GCGGGGCGUGGCCGUAUCAGAAAAUGG(3')	501
AB037926.1	AB037926_3637_3714	GCUACGUCAUAAGUCACGUGACUGGGCAGGUACUAAACCCGGAAGUAUCCUCGGUCACGUGGCCUGUCACGUAGUUG	312	AAGUCACGUGACUGGGCAGGU(5')	407	CCUCGGUCACGUGGCCUGU(3')	502
AB038621.1	AB038621_3511_3591	GGCUSUGACGUCAAAGUCACGUGGGRAGGGUGGCGUUAAACCCGGAAGUCAUCCUCGUCACGUGACCUGACAGCC	313	UGACGUCAAAGUCACGUGGGRAGGGU(5')	408	CCUCGUCACGUGACCUGACGUCACAG(3')	503
AB038622.1	AB038622_227_293	GCCCGUCCGCGGCGAGAGCGCGAGCGAAGCGAGCGAUCGAGCGUCCCGUGGGCGGGGUGCCGAAGGU	314	GAUCGAGCGUCCCGUGGGCGGGU(3')	409	CCGUCCGCGGCGAGAGCGCGAGCGA(5')	504
AB038622.1	AB038622_3510_3591	GGUUGUGACGUCAAAGUCACGUGGGGAGGGCGGCGUUAAACCCGGAAGUCAU	315	UGACGUCAAAGUCACGUGGGGAGGGCGG(5')	410	AUCCUCGUCACGUGACCUGACGUCACG(3')	505

		CCUCGUCACGUGACCUGACGUCACGGCC					
AB038623.1	AB038623_228_295	GCCCGUCCGCGGCGAGAGCGCGAGCGAAGCGAGCGAUCGAGCGUCCCGUGGGCGGGUGCCGUAGGUG	316	GAUCGAGCGUCCCGUGGGCGGGU(3')	411	CCGUCCGCGGCGAGAGCGCGAGCGA(5')	506
AB038624.1	AB038624_228_295	GCCCGUCCGCGGCGAGAGCGCGAGCGAAGCGAGCGAUCGAGCGUCCCGUGGGCGGGUGCCGUAGGUG	317	GAUCGAGCGUCCCGUGGGCGGGU(3')	412	CCGUCCGCGGCGAGAGCGCGAGCGA(5')	507
AB038624.1	AB038624_3511_3592	GGCUGUGACGUCAAAGUCACGUGGGGAGGGCGGCGUUAAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGGCC	318	UGACGUCAAAGUCACGUGGGGAGGGCGG(5')	413	AUCCUCGUCACGUGACCUGACGUCACG(3')	508
AB041957.1	AB041957_3414_3493	AGACCACGUGGUAAGUCACGUGGGGGCAGCUGCUGUAAACCCGGAAGUAGCUGACCCGCGUGACUGGUCACGUGACCUG	319	ACGUGGUAAGUCACGUGGGGGCAGCU(5')	414	CUGACCCGCGUGACUGGUCACGUGA(3')	509
AB049608.1	AB049608_3199_3277	CGCCAUUUUAUAAUACGCGCGUCCCCUCCCGGCUUCCGUACUACGUCAGGCGGGGCGUGGCCGUAUUAGAAAAUGGUG	320	AUUUUAUAAUACGCGCGUCCCCUCC(5')	415	CGGGGCGUGGCCGUAUUAGAAAAUGG(3')	510
AB050448.1	AB050448_3393_3465	UAAGUAAGGCGGAACCAGGCUGUCACCCUGUGUCAAAGGUCAAGGGACAGCCUUCCGGCUUGCACAAAAUGG	321	AAGGGACAGCCUUCCGGCUUGC(3')	416	AGUAAGGCGGAACCAGGCUGUCACCCUGU(5'))	511
AB054647.1	AB054647_3537_3615	UGCCUACGUCAUAAGUCACGUGGGGACGGCUGCUGUAAACACGGAAGUAGCU	322	CAUAAGUCACGUGGGGACGGCUGCU(5')	417	UAGCUGACCCGCGUGACUUGUCAC(3')	512

		GACCCGCGUGACU UGUCACGUGAGCA					
AB054648.1	AB054648_3439_3511	UUGUGUAAGGCGG AACAGGCUGACAC CCCGUGUCAAAGG UCAGGGGUCAGCC UCCGCUUUGCACC AAAUGGU	323	UAAGGCGGAA CAGGCUGACA CCCC(5')	418	GGUCAGCCUC CGCUUUGCA(3')	513
AB054648.1	AB054648_3538_3617	UACCUACGUCAUAA GUCACGUGGGAAG AGCUGCUGUGAAC CUGGAAGUAGCUG ACCCGCGUGGCUU GUCACGUGAGUGC	324	UACGUCAUAA GUCACGUGGG AAGAGCUG(5')	419	GCUGACCCGC GUGGCUUGUC ACGUGAGU(3')	514
AB064595.1	AB064595_116_191	UUUUCCUGGCCCG UCCGCGGCGAGAG CGCGAGCGAAGCG AGCGAUCGGGCGU CCCGAGGGCGGGU GCCGGAGGUG	325	UCGGGCGUCC CGAGGGCGGG UG(3')	420	GGCCCGUCCG CGGCGAGAGC GCGAG(5')	515
AB064595.1	AB064595_3283_3351	AAAGUGAGUGGGG CCAGACUUCGCCA UAGGGCCUUUAAC UUCCGGGUGCGUC UGGGGGCCGCCAU UUU	326	AAAGUGAGUG GGGCCAGACU UCGCC(5')	421	UCCGGGUGCG UCUGGGGGCC GCCAUUU(3')	516
AB064595.1	AB064595_3427_3500	GUGACGUUACUCU CACGUGAUGGGGG CGUGCUCUAACCC GGAAGCAUCCUCG ACCACGUGACUGU GACGUCAC	327	CUCUCACGUG AUGGGGGCGU GC(5')	422	AUCCUCGACC ACGUGACUGU G(3')	517
AB064595.1	AB064595_41_116	AGCGUCUACUACG UACACUUCCUGGG GUGUGUCCUGCCA CUGUAUAUAAACCA GAGGGGUGACGAA UGGUAGAGU	328	UCUACUACGU ACACUUCCUG GGGUGUGU(5')	423	AUAAACCAGA GGGGUGACGA AUGGUAGAGU(3')	518
AB064596.1	AB064596_3424_3497	GUGACGUCAAAGU CACGUGGUGACGG CCAUUUUAACCCG GAAGUGGCUGUUG UCACGUGACUUGA CGUCACGG	329	UGGCUGUUGU CACGUGACUU GA(3')	424	CAAAGUCACG UGGUGACGGC CAU(5')	519

AB064597.1	AB064597_3191_3253	GCUUUAGACGCCAUUUUAGGCCCUCGCGGGCACCCGUAGGCGCGUUUUAAUGACGUCACGGC	330	AGACGCCAUUUUAGGCCCUCGCGG(5')	425	GUAGGCGCGUUUUAAUGACGUCACGG(3')	520
AB064597.1	AB064597_3221_3294	CACCCGUAGGCGCGUUUUAAUGACGUCACGGCAGCCAUUUUGUCGUGACGUUUGAGACACGUGAUGGGGGCGU	331	UGUCGUGACGUUUGAGACACGUGAU(3')	426	UAGGCGCGUUUUAAUGACGUCACGGCAG(5')	521
AB064597.1	AB064597_3262_3342	GUCGUGACGUUUGAGACACGUGAUGGGGGCGUGCCUAAACCCGGAAGCAUCCCUGGUCACGUGACUCUGACGUCACGGCG	332	UGACGUUUGAGACACGUGAUGGGGGCGUGC(5')	427	AUCCCUGGUCACGUGACUCUGACGUCACG(3')	522
AB064598.1	AB064598_3179_3256	CGAAAGUGAGUGGGGCCAGACUUCGCCAUAAGGCCUUUAACUUCCGGGUGCGUGUGGGGGCCGCCAUUUUAGCUUCG	333	AGUGAGUGGGGCCAGACUUCGC(5')	428	GCGUGUGGGGGCCGCCAUUUUAGCUU(3')	523
AB064598.1	AB064598_3323_3399	CUGUGACGUCAAAGUCACGUGGGGAGGCGGCGUGUAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGG	334	UGUGACGUCAAAGUCACGUGGGAGGGCGG(5')	429	UCAUCCUCGUCACGUGACCUGACGUCACG(3')	524
AB064598.1	AB064598_3412_3485	CUGUCCGCCAUCUUGUGACUUCCUUCCGCUUUUUCAAAAAAAAAGAGGAAGUAUGACGUAGCGGCGGGGGGC	335	AAAAGAGGAAGUAUGACGUAGCGGCGG(3')	430	CGCCAUCUUGUGACUUCCUUCCGCUUUUU(5')	525
AB064599.1	AB064599_108_175	GGUAGAGUUUUUUCCGCCCGUCCGCAGCGAGGACGCGAGCGCAGCGAGCGGCCGAGCGACCCGUGG	336	AGCGAGCGGCCGAGCGACCCG(3')	431	UAGAGUUUUUUCCGCCCGUCCG(5')	526
AB064599.1	AB064599_3389_3469	GCUGUGACGUUUCAGUCACGUGGGGAGGGAACGCCUAAACCCGGAAGCGUCC	337	UUCAGUCACGUGGGGAGGGAACGC(5')	432	GUCCCUGGUCACGUGAUUGUGAC(3')	527

		CUGGUCACGUGAUUGUGACGUCACGGCC					
AB064599.1	AB064599_3483_3546	CCGCCAUUUUGUGACUUCCUUCCGCUUUUUCAAAAAAAAAGAGGAAGUGUGACGUAGCGGCGG	338	AAAAGAGGAAGUGUGACGUAGCGG(3')	433	CAUUUUGUGACUUCCUUCCGCUUUUU(5')	528
AB064600.1	AB064600_3378_3456	GACUGUGACGUCAAAGUCACGUGGGGAGGGCGGCGUGUAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGG	339	UGUGACGUCAAAGUCACGUGGGGAGGGCGG(5')	434	UCAUCCUCGUCACGUGACCUGACGUCACG(3')	529
AB064600.1	AB064600_3469_3542	CUGUCCGCCAUCUUGUGACUUCCUUCCGCUUUUUCAAAAAAAAAGAGGAAGUAUGACGUGGCGGCGGGGGGGC	340	AAAAGAGGAAGUAUGACGUGGCGG(3')	435	CCGCCAUCUUGUGACUUCCUUCCGCUUUUU(5')	530
AB064601.1	AB064601_3318_3398	GGUUGUGACGUCAAAGUCACGUGGGGAGGGCGGCGUGUAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGGCC	341	UGACGUCAAAGUCACGUGGGGAGGGCGG(5')	436	AUCCUCGUCACGUGACCUGACGUCACG(3')	531
AB064601.1	AB064601_3412_3477	CCCGCCAUCUUGUGACUUCCUUCCGCUUUUUCAAAAAAAAAGAGGAAGUGUGACGUAGCGGCGGGCGG	342	AAAAAAGAGGAAGUGUGACGUAGCGGCGG(3')	437	CGCCAUCUUGUGACUUCCUUCCGCUUUUUC(5')	532
AB064602.1	AB064602_125_192	GCCCGUCCGCGGCGAGAGCGCGAGCGAAGCGAGCGAUCGAGCGUCCCGUGGGCGGGUGCCGUAGGUG	343	GAUCGAGCGUCCCGUGGGCGGGU(3')	438	CCGUCCGCGGCGAGAGCGCGAGCGA(5')	533
AB064602.1	AB064602_3368_3446	GACUGUGACGUCAAAGUCACGUGGGGAGGAGGGCGUGUAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGG	344	UGUGACGUCAAAGUCACGUGGGGAGGAGGG(5')	439	UCAUCCUCGUCACGUGACCUGACGUCACG(3')	534

AB064603.1	AB064603_3385_3447	UCGCGUCUUAGUGACGUCACGGCAGCCAUCUUGGUCCUGACGUCACUGUCACGUGGGGAGGG	345	UUGGUCCUGACGUCACUGUCA(3')	440	CUUAGUGACGUCACGGCAGCCAU(5')	535
AB064603.1	AB064603_3422_3498	UGACGUCACUGUCACGUGGGGAGGGAACACGUGAACCCGGAAGUGUCCCUGGUCACGUGACAUGACGUCACGGCCG	346	CGUCACUGUCACGUGGGGAGGGAACAC(5')	441	GUCCCUGGUCACGUGACAUGACGUC(3')	536
AB064604.1	AB064604_3436_3514	CGCCAUUUUAAGUAAGCAUGGCGGGCGGUGAUGUCAAAUGUUAAAGGUCACAGCCGGUCAUGCUUGCACAAAAUGGCG	347	UAAGUAAGCAUGGCGGGCGGUGAU(5')	442	CACAGCCGGUCAUGCUUGCACAAA(3')	537
AB064605.1	AB064605_3440_3518	CGCCAUUUUAAGUAAGCAUGGCGGGCGGUGACGUGCAAUGUCAAAGGUCACAGCCUGUCAUGCUUGCACAAAAUGGCG	348	AAGUAAGCAUGGCGGGCGGUGA(5')	443	ACAGCCUGUCAUGCUUGCACAA(3')	538
AB064606.1	AB064606_3377_3449	CCAUCUUAAGUAGUUGAGGCGGACGGUGGCGUCGGUUCAAAGGUCACCAUCAGCCACACCUACUCAAAAUGG	349	UAAGUAGUUGAGGCGGACGGUGGC(5')	444	CACCAUCAGCCACACCUACUCAAA(3')	539
AB064607.1	AB064607_3502_3569	GCCUGUCAUGCUUGCACAAAAUGGCGGACUUCCGCUUCCGGGUCGCCGCCAUAUUUGGUCACGUGAC	350	UCAUGCUUGCACAAAAUGGCGGACUUCCG(5')	445	CGGGUCGCCGCCAUAUUUGGUCACGUGA(3')	540
AF079173.1	AF079173_3475_3551	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUAAAGGUUAAAGGUCAUCCUCGGCGGAAGCUACACACAAAAUGGU	351	AGUAGCUGACGUCAAGGAUUGAC(5')	446	CAUCCUCGGCGGAAGCUACACAA(3')	541
AF116842.1	AF116842_3475_3551	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUAAAGGUUAAAGGUCAUCC	352	AGUAGCUGACGUCAAGGAUUGAC(5')	447	CAUCCUCGGCGGAAGCUACACAA(3')	542

		UCGGCGGAAGCUA CACAAAAUGGU					
AF116842.1	AF116842_3579_3657	GCAUACGUCACAA GUCACGUGGGGGG GACCCGCUGUAAC CCGGAAGUAGGCC CCGUCACGUGACU UACCACGUGUGUA	353	ACAAGUCACG UGGGGGGGAC CCG(5')	448	GGCCCCGUCA CGUGACUUAC CAC(3')	543
AF122913.1	AF122913_3475_3551	GCCAUUUUAAGUA GCUGACGUCAAGG AUUGACGUGAAGG UUAAAGGUCAUCC UCGGCGGAAGCUA CACAAAAUGGU	354	AAGUAGCUGA CGUCAAGGAU UGACG(5')	449	UCAUCCUCGG CGGAAGCUAC ACAA(3')	544
AF122913.1	AF122913_3579_3657	GCACACGUCAUAA GUCACGUGGUGGG GACCCGCUGUAAC CCGGAAGUAGGCC CCGUCACGUGAUU UGUCACGUGUGUA	355	AUAAGUCACG UGGUGGGGAC CCG(5')	450	GGCCCCGUCA CGUGAUUUGU CAC(3')	545
AF122914.1	AF122914_3476_3552	GCCAUUUUAAGUC AGCUCUGGGGAGG CGUGACUUCCAGU UCAAAGGUCAUCC UCACCAUAACUGG CACAAAAUGGC	356	AAGUCAGCUC UGGGGAGGCG UGACUU(5')	451	GUCAUCCUCA CCAUAACUGG CACAA(3')	546
AF122915.1	AF122915_3475_3551	GCCAUUUUAAGUA GCUGACGUCAAGG AUUGACGUAAAGG UUAAAGGUCAUCC UCGGCGGAAGCUA CACAAAAUGGU	357	AGUAGCUGAC GUCAAGGAUU GAC(5')	452	CAUCCUCGGC GGAAGCUACA CAA(3')	547
AF122915.1	AF122915_3579_3657	GCAUACGUCACAA GUCACGUGGAGGG GACACGCUGUAAC CCGGAAGUAGGCC CCGUCACGUGACU UACCACGUGUGUA	358	CAAGUCACGU GGAGGGGACA CG(5')	453	GGCCCCGUCA CGUGACUUAC CAC(3')	548
AF122916.1	AF122916_3458_3537	GCGCCAUGUUAAG UGGCUGUCGCCGA GGAUUGACGUCAC AGUUCAAAGGUCA UCCUCGACGGUAA CCGCAAACAUGGC G	359	UGUUAAGUGG CUGUCGCCGA GGAUUGA(5')	454	AUCCUCGACG GUAACCGCAA ACAUG(3')	549

AF122916.1	AF122916_3565_3641	CAUGCGUCAUAAGUCACAUGACAGGGGUCCACUUAAACACGGAAGUAGGCCCCGACAUGUGACUCGUCACGUGUGU	360	UAAGUCACAUGACAGGGGUCCA(5')	455	GGCCCCGACAUGUGACUCGUC(3')	550
AF122916.1	AF122916_91_164	UGGCAGCACUUCCGAAUGGCUGAGUUUCCACGCCCGUCCGCGGAGAGGGAGCCACGGAGGUGAUCCCGAACG	361	CGGAGAGGGAGCCACGGAGGUG(3')	456	AGCACUUCCGAAUGGCUGAGUUUUCCA(5')	551
AF122917.1	AF122917_3369_3447	GCCAUUUUAAGUCAGCGCUGGGGAGGCAUGACUGUAAGUUCAAAGGUCAUCCUCACCGGAACUGACACAAAAUGGCCG	362	AAGUCAGCGCUGGGGAGGCAUGA(5')	457	AUCCUCACCGGAACUGACACAA(3')	552
AF122918.1	AF122918_3460_3540	GCCAUCUUAAGUGGCUGUCGCCGAGGAUUGACGUCACAGUUCAAAGGUCAUCCUCGGCGGUAACCGCAAAGAUGGCGGUC	363	UCUUAAGUGGCUGUCGCCGAGGAUUGAC(5')	458	CAUCCUCGGCGGUAACCGCAAAGAUG(3')	553
AF122918.1	AF122918_3566_3642	AUACGUCAUAAGUCACAUGUCUAGGGGUCCACUUAAACACGGAAGUAGGCCCCGACAUGUGACUCGUCACGUGUGU	364	AAGUCACAUGUCUAGGGGUCCACU(5')	459	UAGGCCCCGACAUGUGACUCGU(3')	554
AF122919.1	AF122919_3370_3447	CCAUUUUAAGUAAGGCGGAAGCAGCUGUCCCUGUAACAAAAUGGCGGCGACAGCCUUCCGCUUUGCACAAAAUGGAG	365	AAGUAAGGCGGAAGCAGCUGUCC(5')	460	ACAGCCUUCCGCUUUGCACAA(3')	555
AF122920.1	AF122920_3460_3540	GCCAUCUUAAGUGGCUGUCGCUGAGGAUUGACGUCACAGUUCAAAGGUCAUCCUCGGCGGUAACCGCAAAGAUGGCGGUC	366	AUCUUAAGUGGCUGUCGCUGAGGAUUGAC(5')	461	CAUCCUCGGCGGUAACCGCAAAGAUGG(3')	556

AF122920.1	AF122920_3565_3641	CAUACGUCAUAAGUCACAUGACAGGAGUCCACUUAAACACGGAAGUAGGCCCCGACAUGUGACUCGUCACGUGUGU	367	UAAGUCACAUGACAGGAGUCCACU(5')	462	UAGGCCCCGACAUGUGACUCGUC(3')	557
AF122921.1	AF122921_3459_3540	CGCCAUCUUAAGUGGCUGUCGCCGAGGAUUGGCGUCACAGUUCAAAGGUCAUCCUCGGCGGUAACCGCAAAGAUGGCGGU	368	AAGUGGCUGUCGCCGAGGAUUG(5')	463	UCCUCGGCGGUAACCGCAAA(3')	558
AF122921.1	AF122921_3565_3641	CAUACGUCAUAAGUCACAUGACAGGGGUCCACUUAAACACGGAAGUAGGCCCCGACAUGUGACUCGUCACGUGUGU	369	UAAGUCACAUGACAGGGGUCCA(5')	464	GGCCCCGACAUGUGACUCGUC(3')	559
AF129887.1	AF129887_3579_3657	GCAUACGUCACAAGUCACGUGGGGGGGACCCGCUGUAACCCGGAAGUAGGCCCCGUCACGUGACUUACCACGUGGUGU	370	ACAAGUCACGUGGGGGGGACCCG(5')	465	GGCCCCGUCACGUGACUUACCAC(3')	560
AF247137.1	AF247137_3453_3530	CCGCCAUUUUAGGCUGUUGCCGGGCGUUUGACUUCCGUGUUAAAGGUCAAACACCCAGCGACACCAAAAAAUGGCCG	371	AUUUUAGGCUGUUGCCGGGCGUUUGACU(5')	466	UCAAACACCCAGCGACACCAAAAAUGG(3')	561
AF247137.1	AF247137_3559_3636	CUACGUCAUAAGUCACGUGACAGGGAGGGGCGACAAACCCGGAAGUCAUCCUCGCCCACGUGACUUACCACGUGGUG	372	AUAAGUCACGUGACAGGGAGGGG(5')	467	CCUCGCCCACGUGACUUACCAC(3')	562
AF247138.1	AF247138_3455_3532	GCCAUUUUAAGUAGGUGACGUCCAGGACUGACGUAAAGUUCAAAGGUCAUCCUCGGCGGAACCUAUACAAAAUGGCG	373	AAGUAGGUGACGUCCAGGACU(5')	468	CCUCGGCGGAACCUAUACAA(3')	563
AF247138.1	AF247138_3561_3637	CUACGUCAUAAGUCACGUGGGGACGGCUGUACUUAAACAC	374	CAUAAGUCACGUGGGGACGGCUGU(5')	469	GCCCCGUCACGUGAUUUACCAC(3')	564

		GGAAGUAGGCCCC GUCACGUGAUUUA CCACGUGGUG					
AF261761.1	AF261761_3431_3504	GCCAUUUUAAGUA AGGCGGAAGAGCU CUAGCUAUACAAAA UGGCGGCGGAGCA CUUCCGCUUUGCC CAAAAUG	375	UAAGUAAGGC GGAAGAGCUC UAGCUA(5')	470	GCGGCGGAGC ACUUCCGCUU UGCCCAAA(3')	565
AF351132.1	AF351132_3475_3552	GCCAUUUUAAGUA GCUGACGUCAAGG AUUGACGUAGAGG UUAAAGGUCAUCC UCGGCGGAAGCUA CACAAAAUGGUG	376	AGUAGCUGAC GUCAAGGAUU GAC(5')	471	CAUCCUCGGC GGAAGCUACA CAA(3')	566
AF351132.1	AF351132_3579_3657	GCAUACGUCACAA GUCACGUGGGGGG GACCCGCUGUAAC CCGGAAGUAGGCC CCGUCACGUGACU UACCACGUGUGUA	377	ACAAGUCACG UGGGGGGGAC CCG(5')	472	GGCCCCGUCA CGUGACUUAC CAC(3')	567
AF435014.1	AF435014_3344_3426	GGCGCCAUUUUAA GUAAGCAUGGCGG GCGGCGACGUCAC AUGUCAAAGGUCA CCGCACUUCCGUG CUUGCACAAAAUG GC	378	UAAGUAAGCA UGGCGGGCGG CGAC(5')	473	CACCGCACUU CCGUGCUUGC ACAAA(3')	568
AF435014.1	AF435014_3453_3526	UGCUACGUCAUCG AGACACGUGGUGC CAGCAGCUGUAAA CCCGGAAGUCGCU GACACACGUGUCU UGUCACGU	379	AUCGAGACAC GUGGUGCCAG CAGCU(5')	474	UCGCUGACAC ACGUGUCUUG UCAC(3')	569
AJ620212.1	AJ620212_3360_3438	GCCAUUUUAAGUA AGCACCGCCUAGG GAUGACGUAUAAG UUCAAAGGUCAUC CUCAGCCGGAACU UACACAAAAUGGU	380	UCAUCCUCAG CCGGAACUUA CACAAAAUGG(3')	475	CAUUUUAAGU AAGCACCGCC UAGGGAUGAC(5')	570
AJ620212.1	AJ620212_3470_3542	ACGUCAUAUGUCA CGUGGGGAGGCCC UGCUGCGCAAACG CGGAAGUAGGCCC CGUCACGUGUCAU ACCACGU	381	AUAUGUCACG UGGGGAGGCC CUGCUG(5')	476	GUAGGCCCCG UCACGUGUCA UACCAC(3')	571

AJ620218.1	AJ620218_3381_3458	CCAUUUUAAGUAAGGCGGAAGCAGCUCCACUUUCUCACAAAAUGGCGGCGGGGCACUUCCGGCUUGCCCAAAAUGGC	382	AAGUAAGGCGGAAGCAGCUCACUUU(5')	477	GGCGGGGCACUUCCGGCUUGCCCAA(3')	572
AJ620226.1	AJ620226_3451_3523	CCAUUUUAAGUAAGGCGGAAGUUUCUCCACUAUACAAAAUGGCGGCGGAGCACUUCCGGCUUGCCCAAAAUG	383	AAGUAAGGCGGAAGUUUCUCCACU(5')	478	CGGCGGAGCACUUCCGGCUUGCCCAA(3')	573
AJ620227.1	AJ620227_3379_3451	CCAUCUUAAGUAGUUGAGGCGGACGGUGGCGUGAGUUCAAAGGUCACCAUCAGCCACACCUACUCAAAAUGG	384	UAAGUAGUUGAGGCGGACGGUGGC(5')	479	CACCAUCAGCCACACCUACUCAAA(3')	574
AJ620231.1	AJ620231_3429_3505	CGCCAUCUUAAGUAGUUGAGGCGGACGGUGGCGUGAGUUCAAAGGUCACCAUCAGCCACACCUACUCUCAAAAUGGUG	385	UAAGUAGUUGAGGCGGACGGUGG(5')	480	ACCAUCAGCCACACCUACUCAAA(3')	575
AY666122.1	AY666122_3163_3236	UUUCGGACCUUCGGCGUCGGGGGGGUCGGGGGCUUUACUAAACAGACUCCGAGAUGCCAUUGGACACUGAGGG	386	GACCUUCGGCGUCGGGGGGGUCGGGGGG(5')	481	GACUCCGAGAUGCCAUUGGACACUGAGG(3')	576
AY666122.1	AY666122_3388_3464	CCAUUUUAAGUAGGUGCCGUCCAGCACUGCUGUUCCGGGUUAAAGGGCAUCCUCGGCGGAACCUAUACAAAAUGGC	387	AUCCUCGGCGGAACCUAUA(3')	482	AGUAGGUGCCGUCCAGCA(5')	577
AY666122.1	AY666122_3494_3567	CUACGUCAUCGAUGACGUGGGGAGGCGUACUAUGAAACGCGGAAGUAGGCCCCGCUACGUCAUCAUCACGUGG	388	AUCGAUGACGUGGGGAGGCGUACUAU(5')	483	AAGUAGGCCCCGCUACGUCAUCAC(3')	578
AY823988.1	AY823988_3452_3525	CCAUUUUAAGUAAGGCGGAAGAGCUGCUCUAUAUACAAAAUGGCGGAGGAGCA	389	UGGCGGAGGAGCACUUCCGGCUUG(3')	484	AAGGCGGAAGAGCUGCUCUAUAU(5')	579

		CUUCCGGCUUGCC CAAAAUG					
AY823988.1	AY823988_3554_3629	UGCCUACGUAACA AGUCACGUGGGGA GGGUUGGCGUAUA ACCCGGAAGUCAA UCCUCCCACGUGG CCUGUCACGU	390	AACAAGUCAC GUGGGGAGGG UUGGC(5')	485	CAAUCCUCCC ACGUGGCCUG UCAC(3')	580
AY823989.1	AY823989_3551_3623	UAAGUAAGGCGGA ACCAGGCUGUCAC CCCGUGUCAAAGG UCAGGGGUCAGCC UUCCGCUUUACAC AAAAUGG	391	AGGGGUCAGC CUUCCGCUUU A(3')	486	AAGGCGGAAC CAGGCUGUCA CCCCGU(5')	581
AY823989.1	AY823989_3551_3623	UAAGUAAGGCGGA ACCAGGCUGUCAC CCCGUGUCAAAGG UCAGGGGUCAGCC UUCCGCUUUACAC AAAAUGG	392	AGGGGUCAGC CUUCCGCUUU A(3')	487	AAGGCGGAAC CAGGCUGUCA CCCCGU(5')	582
DQ361268.1	DQ361268_3413_3494	GCAGCCAUUUUAA GUCAGCUUCGGGG AGGGUCACGCAAA GUUCAAAGGUCAU CCUCACCGGAACU GGUACAAAAUGGC CG	393	UAAGUCAGCU UCGGGGAGGG UCAC(5')	488	CAUCCUCACC GGAACUGGUA CAAA(3')	583
DQ361268.1	DQ361268_3519_3593	UGCUACGUCAUAA GUGACGUAGCUGG UGUCUGCUGUAAA CACGGAAGUAGGC CCCGCCACGUCAC UUGUCACGU	394	UCAUAAGUGA CGUAGCUGGU GUCUGCU(5')	489	UAGGCCCCGC CACGUCACUU GUCACG(3')	584

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).

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).

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.

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.

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)

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.

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.

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.

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, $F(ab')_2$; 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).

In some embodiments, the effector comprises a regulatory intracellular polyeptpide. 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

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-1α, IL-1β, or a heterodimer thereof	IL-1 type 1 receptor, IL-1 type 2 receptor	3552, 3553	P01583, P01584
IL-1Ra	IL-1 type 1 receptor, IL-1 type 2 receptor	3454, 3455	P17181, P48551
IL-2	IL-2R	3558	P60568
IL-3	IL-3 receptor α + β c (CD131)	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-11Rα 1 gp130	3589	P20809
IL-12 (e.g., p35, p40, or a heterodimer thereof)	IL-12Rβ1 and IL-12Rβ2	3593, 3592	P29459, P29460
IL-13	IL-13R1α1 and IL-13R1α2	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 IL29	IL-28R1/IL-10R2	282617, 282618	Q8IZI9, Q8IU54
IL-30	IL6R/gp130	246778	Q8NEV9
IL-31	IL-31RA/OSMRβ	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-12Rβ2/gp130; IL-12Rβ2/IL-12Rβ2; gp130/gp130	10148	Q14213
IL-36	IL-36Ra	27179	Q9UHA7
IL-37	IL-18Rα and IL-18BP	27178	Q9NZH6
IL-38	IL-1R1, IL-36R	84639	Q8WWZ1
IFN-α	IFNAR	3454	P17181

IFN-β	IFNAR	3454	P17181
IFN-γ	IFNGR1/IFNGR2	3459	P15260
TGF-β	TβR-I and TβR-II	7046, 7048	P36897, P37173
TNF-α	TNFR1, TNFR2	7132, 7133	P19438, P20333

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%, 967%, 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.

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.

Exemplary cytokines and cytokine receptors are described, e.g., in Akdis et al., "Interleukins (from IL-1 to IL-38), interferons, transforming growth factor β, and TNF-α: 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, NPRC	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
Adrenocorticotropic 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)	PTH1R and PTH2R	5741	P01270
Insulin	Insulin receptor (IR)	3630	P01308
Glucagon	Glucagon receptor	2641	P01275

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.

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. Exemplary growth factors

Growth Factor		**Entrez Gene ID**	**UniProt ID**
PDGF family			
PDGF (e.g., PDGF-1, PDGF-2, or a heterodimer thereof)	PDGF receptor, e.g., PDGFRα, PDGFRβ	5156	P16234
CSF-1	CSF1R	1435	P09603
SCF	CD117	3815	P10721
VEGF family			
VEGF (e.g., isoforms VEGF 121, VEGF 165, VEGF 189, and VEGF 206)	VEGFR-1, VEGFR-2	2321	P17948
VEGF-B	VEGFR-1	2321	P17949
VEGF-C	VEGFR-2 and VEGFR -3	2324	P35916
PlGF	VEGFR-1	5281	Q07326
EGF family			
EGF	EGFR	1950	P01133
TGF-α	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

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.

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.

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)	Afibrinogenomia	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
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

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

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) Type III-C	138050	Q68CP4
ADAMTS13	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
Cystathionine 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-α-3-glucosyltransferase	Congenital disorders of N-glycosylation CDG Ih	79053	Q9BVK2
Dolichyl-P-mannose:Man-7-GlcNAc-2-PP-dolichyl-α-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
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 β-galactosidase	Krabbe's disease	2581	P54803

Ganglioside β-galactosidase	GM1 gangliosidosis, generalized	2720	P16278
Ganglioside β-galactosidase	GM2 gangliosidosis	2720	P16278
Ganglioside β-galactosidase	Sphingolipidosis Type I	2720	P16278
Ganglioside β-galactosidase	Sphingolipidosis Type II (juvenile type)	2720	P16278
Ganglioside β-galactosidase	Sphingolipidosis Type III (adult type)	2720	P16278
Glucosidase I	Congenital disorders of N-glycosylation CDG IIb	2548	P10253
Glucosylceramide β-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 lipofuscinosis Late infantile form (CLN2, Jansky-Bielschowsky disease)	1200	O14773
Mannose (Man) phosphate (P) isomerase	Congenital disorders of N-glycosylation CDG Ib	4351	P34949
Mannosyl-α-1,6-glycoprotein-β-1,2-N-acetylglucosminyltransferase	Congenital disorders of N-glycosylation CDG IIa	4247	Q10469
Metalloproteinase-2	Winchester syndrome	4313	P08253
methylmalonyl-CoA mutase	Methylmalonic acidemia (vitamin b12 non-responsive)	4594	P22033
N-Acetyl galactosamine α-4-sulfate sulfatase (arylsulfatase B)	Mucopolysaccharidosis MPS VI (Maroteaux-Lamy syndrome)	411	P15848
N-acetyl-D-glucosaminidase	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-sulfate sulfatase	Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-D	2799	P15586
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
Neuraminidase 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 α-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 glucuronyl-transferase (e.g., UGT1A1)	Crigler–Najjar syndrome	54658	P22309
α-1,2-Mannosyltransferase	Congenital disorders of N-glycosylation CDG Il (608776)	79796	Q9H6U8
α-1,2-Mannosyltransferase	Congenital disorders of N-glycosylation, type I (pre-Golgi glycosylation defects)	79796	Q9H6U8
α-1,3-Mannosyltransferase	Congenital disorders of N-glycosylation CDG Ii	440138	Q2TAA5
α-D-Mannosidase	α-Mannosidosis, type I (severe) or II (mild)	10195	Q92685
α-L-Fucosidase	Fucosidosis	4123	Q9NTJ4

α-l-Iduronidase	Mucopolysaccharidosis MPS I H/S (Hurler-Scheie syndrome)	2517	P04066
α-l-Iduronidase	Mucopolysaccharidosis MPS I-H (Hurler's syndrome)	3425	P35475
α-l-Iduronidase	Mucopolysaccharidosis MPS I-S (Scheie's syndrome)	3425	P35475
β-1,4-Galactosyltransferase	Congenital disorders of N-glycosylation CDG IId	3425	P35475
β-1,4-Mannosyltransferase	Congenital disorders of N-glycosylation CDG Ik	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

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
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	Prosaposin deficiency	5660	P07602
Protective protein/cathepsin A (PPCA)	Galactosialidosis (Goldberg's syndrome, combined neuraminidase and β-galactosidase deficiency)	5476	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	2055	Q9UBY8

	with intellectual disability		
vWF	von Willebrand disease	7450	P04275
Factor I (fibrinogen)	Afibrinogenomia	2243, 2244, 2266	P02671, P02675, P02679
erythropoietin (hEPO)			

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.

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.

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.

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.

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 polyeptpide 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

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
FGF2	NG_029067	NP_001348594
FGF1	Gene ID: 2246	NP_001341882
miR-199-3p	MIMAT0000232	

miR-590-3p	MIMAT0004801	
mi-17-92	MI0000071	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2732113/figure/F1/
miR-222	MI0000299	
miR-302-367	MIR302A And MIR367	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400607/

Transformation Factors

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, .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

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%, 967%, 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

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 Aug;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 Nov; 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.

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-organellar antigen.

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.

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

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 Oct;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.

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.

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 meningiditis). 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) 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.

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.

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.

CRISPR technology for editing the genes of eukaryotes is disclosed in US Patent Application Publications 2016/0138008A1 and US2015/0344912A1, and in US Patents 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.

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.

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.

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

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, 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.

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.

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 metallothionine (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.

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.

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 α-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.

The genetic element 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.

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.

Replication Proteins

In some embodiments, the genetic element of the anellovector, e.g., synthetic anellovector, may include sequences that encode one or more replication proteins. In some embodiments, the anellovector may replicate by a rolling-circle replication method, e.g., synthesis of the leading strand and the lagging strand is uncoupled. In such embodiments, the anellovector comprises three elements additional elements: i) a gene encoding an initiator protein, ii) a double strand origin, and iii) a single strand origin. A rolling circle replication (RCR) protein complex comprising replication proteins binds to the leading strand and destabilizes the replication origin. The RCR complex cleaves the genome to generate a free 3'OH extremity. Cellular DNA polymerase initiates viral DNA replication from the free 3'OH extremity. After the genome has been replicated, the RCR complex closes the loop covalently. This leads to the release of a positive circular single-stranded parental DNA molecule and a circular double-stranded DNA molecule composed of the negative parental strand and the newly synthesized positive strand. The single-stranded DNA molecule can be either encapsidated or involved in a second round of replication. See for example, Virology Journal 2009, 6:60 doi:10.1186/1743-422X-6-60.

The genetic element may comprise a sequence encoding a polymerase, e.g., RNA polymerase or a DNA polymerase.

Other Sequences

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).

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.

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.

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.

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

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.

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 US Patent 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 US Patent No.: 8,697,359, which are herein incorporated by reference in their entirety.

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

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 a Monodnavirus, e.g., a Shotokuvirus (e.g., a Cressdnaviricota [e.g., a redondovirus, circovirus {e.g., a porcine circovirus, e.g., PCV-1 or PCV-2; or beak-and-feather disease virus}, geminivirus {e.g., tomato golden mosaic virus}, or nanovirus {e.g., BBTV, MDV1, SCSVF, or FBNYV}]), or a Parvovirus (e.g., a dependoparavirus, e.g., a bocavirus or an AAV), e.g., as described herein, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the genetic element comprises a sequence from an Anellovirus genome, e.g., as described herein, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the sequence is from an Anellovirus genome as listed in Table 41 below.

Table 41: Examples of *Anelloviruses* and their sequences. Accessions numbers and related sequence information may be obtained at www.ncbi.nlm.nih.gov/genbank/, as referenced on December 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:TJN02
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
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
AF345525.1	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 tth6, complete genome
AJ620213.1	Torque teno virus, isolate tth10, complete genome
AJ620214.1	Torque teno virus, isolate tth11g2, complete genome

AJ620215.1	Torque teno virus, isolate tth18, complete genome
AJ620216.1	Torque teno virus, isolate tth20, complete genome
AJ620217.1	Torque teno virus, isolate tth21, complete genome
AJ620218.1	Torque teno virus, isolate tth3, complete genome
AJ620219.1	Torque teno virus, isolate tth9, complete genome
AJ620220.1	Torque teno virus, isolate tth16, complete genome
AJ620221.1	Torque teno virus, isolate tth17, complete genome
AJ620222.1	Torque teno virus, isolate tth25, complete genome
AJ620223.1	Torque teno virus, isolate tth26, complete genome
AJ620224.1	Torque teno virus, isolate tth27, complete genome
AJ620225.1	Torque teno virus, isolate tth31, complete genome
AJ620226.1	Torque teno virus, isolate tth4, complete genome
AJ620227.1	Torque teno virus, isolate tth5, complete genome
AJ620228.1	Torque teno virus, isolate tth14, complete genome
AJ620229.1	Torque teno virus, isolate tth29, complete genome
AJ620230.1	Torque teno virus, isolate tth7, complete genome
AJ620231.1	Torque teno virus, isolate tth8, complete genome
AJ620232.1	Torque teno virus, isolate tth13, complete genome
AJ620233.1	Torque teno virus, isolate tth19, complete genome
AJ620234.1	Torque teno virus, isolate tth22g4, complete genome
AJ620235.1	Torque teno virus, isolate tth23, 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
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	Anelloviridae sp. isolate ctdc048, complete genome
MH648893.1	Anelloviridae sp. isolate ctdh007, complete genome
MH648897.1	Anelloviridae sp. isolate ctcb038, complete genome
MH648900.1	Anelloviridae sp. isolate ctfc019, complete genome
MH648901.1	Anelloviridae sp. isolate ctbb022, complete genome
MH648907.1	Anelloviridae sp. isolate ctcf040, complete genome
MH648911.1	Anelloviridae sp. isolate cthi018, complete genome
MH648912.1	Anelloviridae sp. isolate ctea38, complete genome
MH648913.1	Anelloviridae sp. isolate ctbg006, complete genome
MH648916.1	Anelloviridae sp. isolate ctbg020, complete genome
MH648925.1	Anelloviridae sp. isolate ctci019, complete genome

MH648932.1	Anelloviridae sp. isolate ctid031, complete genome
MH648946.1	Anelloviridae sp. isolate ctdb017, complete genome
MH648957.1	Anelloviridae sp. isolate ctch017, complete genome
MH648958.1	Anelloviridae sp. isolate ctbh011, complete genome
MH648959.1	Anelloviridae sp. isolate ctbc020, complete genome
MH648962.1	Anelloviridae sp. isolate ctif015, complete genome
MH648966.1	Anelloviridae sp. isolate ctei055, complete genome
MH648969.1	Anelloviridae sp. isolate ctjg000, complete genome
MH648976.1	Anelloviridae sp. isolate ctcj064, complete genome
MH648977.1	Anelloviridae sp. isolate ctbj022, complete genome
MH648982.1	Anelloviridae sp. isolate ctbf014, complete genome
MH648983.1	Anelloviridae sp. isolate ctbd027, complete genome
MH648985.1	Anelloviridae sp. isolate ctch016, complete genome
MH648986.1	Anelloviridae sp. isolate ctbd020, complete genome
MH648989.1	Anelloviridae sp. isolate ctga035, complete genome
MH648990.1	Anelloviridae sp. isolate cthf001, complete genome
MH648995.1	Anelloviridae sp. isolate ctbd067, complete genome
MH648997.1	Anelloviridae sp. isolate ctce026, complete genome
MH648999.1	Anelloviridae sp. isolate ctfb058, complete genome
MH649002.1	Anelloviridae sp. isolate ctjj046, complete genome
MH649006.1	Anelloviridae sp. isolate ctcf030, complete genome
MH649008.1	Anelloviridae sp. isolate ctbg025, complete genome
MH649011.1	Anelloviridae sp. isolate ctbh052, complete genome
MH649014.1	Anelloviridae sp. isolate ctba003, complete genome
MH649017.1	Anelloviridae sp. isolate ctbb016, complete genome
MH649022.1	Anelloviridae sp. isolate ctch023, complete genome
MH649023.1	Anelloviridae sp. isolate ctbd051, complete genome
MH649028.1	Anelloviridae sp. isolate ctbf9, complete genome
MH649038.1	Anelloviridae sp. isolate ctbi030, complete genome
MH649039.1	Anelloviridae sp. isolate ctca057, complete genome
MH649040.1	Anelloviridae sp. isolate ctch033, complete genome
MH649042.1	Anelloviridae sp. isolate ctjd005, complete genome
MH649045.1	Anelloviridae sp. isolate ctdc021, complete genome
MH649051.1	Anelloviridae sp. isolate ctdg044, complete genome
MH649056.1	Anelloviridae sp. isolate ctcc062, complete genome
MH649061.1	Anelloviridae sp. isolate ctid009, complete genome

MH649062.1	Anelloviridae sp. isolate ctdc018, complete genome
MH649063.1	Anelloviridae sp. isolate ctbf012, complete genome
MH649068.1	Anelloviridae sp. isolate ctcc066, complete genome
MH649070.1	Anelloviridae sp. isolate ctda011, complete genome
MH649077.1	Anelloviridae sp. isolate ctbh034, complete genome
MH649083.1	Anelloviridae sp. isolate ctdg028, complete genome
MH649084.1	Anelloviridae sp. isolate ctii061, complete genome
MH649085.1	Anelloviridae sp. isolate cteh021, complete genome
MH649092.1	Anelloviridae sp. isolate ctbg012, complete genome
MH649101.1	Anelloviridae sp. isolate ctif053, complete genome
MH649104.1	Anelloviridae sp. isolate ctei657, complete genome
MH649106.1	Anelloviridae sp. isolate ctca015, complete genome
MH649114.1	Anelloviridae sp. isolate ctbf050, complete genome
MH649122.1	Anelloviridae sp. isolate ctdc002, complete genome
MH649125.1	Anelloviridae sp. isolate ctbb15, complete genome
MH649127.1	Anelloviridae sp. isolate ctba013, complete genome
MH649137.1	Anelloviridae sp. isolate ctbb000, complete genome
MH649141.1	Anelloviridae sp. isolate ctbc019, complete genome
MH649142.1	Anelloviridae sp. isolate ctid026, complete genome
MH649144.1	Anelloviridae sp. isolate ctfj004, complete genome
MH649152.1	Anelloviridae sp. isolate ctcj13, complete genome
MH649156.1	Anelloviridae sp. isolate ctci006, complete genome
MH649157.1	Anelloviridae sp. isolate ctbd025, complete genome
MH649158.1	Anelloviridae sp. isolate ctbf005, complete genome
MH649161.1	Anelloviridae sp. isolate ctcf045, complete genome
MH649165.1	Anelloviridae sp. isolate ctcc29, complete genome
MH649169.1	Anelloviridae sp. isolate ctib021, complete genome
MH649172.1	Anelloviridae sp. isolate ctbh857, complete genome
MH649174.1	Anelloviridae sp. isolate ctbj049, complete genome
MH649178.1	Anelloviridae sp. isolate ctfc006, complete genome
MH649179.1	Anelloviridae sp. isolate ctbe000, complete genome
MH649183.1	Anelloviridae sp. isolate ctbb031, complete genome
MH649186.1	Anelloviridae sp. isolate ctcb33, complete genome
MH649189.1	Anelloviridae sp. isolate ctcc12, complete genome
MH649196.1	Anelloviridae sp. isolate ctci060, complete genome
MH649199.1	Anelloviridae sp. isolate ctbb017, complete genome

MH649203.1	Anelloviridae sp. isolate cthc018, complete genome
MH649204.1	Anelloviridae sp. isolate ctbj003, complete genome
MH649206.1	Anelloviridae sp. isolate ctbg010, complete genome
MH649208.1	Anelloviridae sp. isolate ctid008, complete genome
MH649209.1	Anelloviridae sp. isolate ctbg056, complete genome
MH649210.1	Anelloviridae sp. isolate ctda001, complete genome
MH649212.1	Anelloviridae sp. isolate ctcf004, complete genome
MH649217.1	Anelloviridae sp. isolate ctbe029, complete genome
MH649223.1	Anelloviridae sp. isolate ctci016, complete genome
MH649224.1	Anelloviridae sp. isolate ctce11, complete genome
MH649228.1	Anelloviridae sp. isolate ctcf013, complete genome
MH649229.1	Anelloviridae sp. isolate ctcb036, complete genome
MH649241.1	Anelloviridae sp. isolate ctda027, complete genome
MH649242.1	Anelloviridae sp. isolate ctbf003, complete genome
MH649254.1	Anelloviridae sp. isolate ctjb007, complete genome
MH649255.1	Anelloviridae sp. isolate ctbb023, complete genome
MH649256.1	Anelloviridae sp. isolate ctca002, complete genome
MH649258.1	Anelloviridae sp. isolate ctcg010, complete genome
MH649263.1	Anelloviridae sp. isolate ctgh3, complete genome
MK012439.1	Anelloviridae sp. isolate cthe000, complete genome
MK012440.1	Anelloviridae sp. isolate ctjd008, complete genome
MK012448.1	Anelloviridae sp. isolate ctch012, complete genome
MK012457.1	Anelloviridae sp. isolate ctda009, complete genome
MK012458.1	Anelloviridae sp. isolate ctcd015, complete genome
MK012485.1	Anelloviridae sp. isolate ctfd011, complete genome
MK012489.1	Anelloviridae sp. isolate ctba003, complete genome
MK012492.1	Anelloviridae sp. isolate ctbb005, complete genome
MK012493.1	Anelloviridae sp. isolate ctcj014, complete genome
MK012500.1	Anelloviridae sp. isolate ctcb001, complete genome
MK012504.1	Anelloviridae sp. isolate ctcj010, complete genome
MK012516.1	Anelloviridae 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
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

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., a Monodnavirus, e.g., a Shotokuvirus (e.g., a Cressdnaviricota [e.g., a redondovirus, circovirus {e.g., a porcine circovirus, e.g., PCV-1 or PCV-2; or beak-and-feather disease virus}, geminivirus {e.g., tomato golden mosaic virus}, or nanovirus {e.g., BBTV, MDV1, SCSVF, or FBNYV}]), or a Parvovirus (e.g., a dependoparavirus, e.g., a bocavirus or an AAV). Since, in some embodiments, recombinant viruses 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 one or more genes (e.g., Rep genes and/or structural genes) of the virus under the control of regulatory sequences, e.g., within the LTR. Suitable cell lines for

replicating the anellovectors described herein include host cell lines as described herein, which can be modified, e.g., as described herein. Said genetic element can additionally contain a gene encoding a selectable marker so that the desired genetic elements can be identified.

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.

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/ 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

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.

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).

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.

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.

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.

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.

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).

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.

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.

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.

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.

In some embodiments, a first plurality of anellovectors comprising a proteinaceous exterior as described herein is administered to a subject. In some embodiments, a second plurality of anellovectors comprising a proteinaceous exterior described herein, is subsequently administered to the subject following administration of the first plurality. In some embodiments, the second plurality of anellovectors comprises the same proteinaceous exterior as the anellovectors of the first plurality. In some embodiments, the second plurality of anellovectors comprises a proteinaceous exterior with at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% amino acid sequence identity to the proteinaceous exterior of the anellovectors of the first plurality. In some embodiments, the second plurality of anellovectors comprises an ORF1 molecule with at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% amino acid sequence identity to the ORF1 molecule of the anellovectors of the first plurality. In some embodiments the second plurality of anellovectors comprises an ORF1 molecule having the same amino acid sequence as the ORF1 molecule comprised by the anellovectors of the first plurality. In some embodiments, the proteinaceous exterior of the second plurality of anellovectors comprises a polypeptide, e.g., an ORF1 molecule, having at least 70%, 75%, 80%, 85%,

90%, 95%, 96%, 97%, 98%, 99%, or 100% amino acid sequence identity to a polypeptide, e.g., an ORF1 molecule, in the proteinaceous exterior of the first plurality of anellovectors. In some embodiments, the proteinaceous exterior of the second plurality of anellovectors comprises a polypeptide, e.g., a capsid protein, having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% amino acid sequence identity to a polypeptide, e.g., a capsid protein, in the proteinaceous exterior of the first plurality of Anellovectors. In some embodiments, the second plurality of anellovectors comprises a proteinaceous exterior with at least one surface epitope in common with the anellovectors of the first plurality. In some embodiments, the second plurality of anellovectors comprises an ORF1 molecule with at least one surface epitope in common with the ORF1 of the anellovectors of the first plurality. In some embodiments, the second plurality of anellovectors comprises a proteinaceous exterior with one or more amino acid sequence difference (e.g., a conservative mutation) from the protenaceous exterior of the anellovectors of the first plurality. In some embodiments, an antibody, e.g., an antibody within the subject, that binds to the proteinaceous exterior of the first plurality of anellovectors also binds to the proteinaceous exterior of the second plurality of of anellovectors. In some embodiments, the antibody binds with about the same affinity (e.g., having a KD of about 90-110%, e.g., 95-105%) to the proteinaceous exterior of the first plurality of anellovectors as to the proteinaceous exterior of the second plurality of anellovectors.

In some embodiments, the proteinaceous exterior of the first plurality of anellovectors comprises the same tertiary structure as the proteinaceous exterior of the second plurality of anellovectors. In some embodiments, the structure, e.g., tertiary structure, of the proteinaceous exterior of the anellovectors in the first and second plurality can be determined using cryo-electron microscopy (cryo-EM), X-ray crystallography, or nuclear magnetic resonance (NMR). In some embodiments, the structure of the proteinaceous exterior of the first plurality of anellovectors is compared to structure of the proteinaceous exterior of the second plurality of anellovectors using structural alignment and measurement of the atomic coordinates of the atoms in the protein structure, e.g., a measurement of root-mean-square-deviation (RMSD). In some embodiments, the RMSD can be calculated for the backbone of the polypeptide chain of the structures being compared, the alpha carbons of the polypeptide chain of the structures being compared, or all the atoms of the structures being compared, e.g., the proteinaceous exterior of the first plurality of anellovectors and the proteinaceous exterior of the second plurality of anellovectors. In some embodiments, an RMSD of a lower value, e.g., ≤ 5 Angstroms, indicates structural similarity between the proteinaceous exterior of the first plurality of anellovectors and proteinaceous exterior of the second plurality of anellovectors. In some embodiments, an RMSD of a lower value, e.g., ≤ 3 Angstroms, indicates high structural similarity between the proteinaceous exterior of the first plurality of anellovectors and proteinaceous exterior of the second plurality of anellovectors. In some embodiments, an RMSD of 0 Angstroms indicates that two proteins comprise the same structure, e.g., that the structure

of the proteinaceous exterior of the first plurality of anellovectors is the same as the proteinaceous exterior of the second plurality of anellovectors.

III. Nucleic Acid Constructs

The genetic element described herein may be included in a nucleic acid construct (e.g., a nucleic acid genetic element construct, e.g., as described herein).

In one aspect, the invention includes a nucleic acid genetic element construct comprising a genetic element comprising (i) a sequence encoding an exterior protein (e.g., a non-pathogenic exterior protein, e.g., an Anellovirus ORF1 molecule or a splice variant or functional fragment thereof), (ii) an exterior protein binding sequence that binds the genetic element to the non-pathogenic exterior protein, and (iii) a sequence encoding an effector.

In another aspect, the invention includes a nucleic acid genetic element construct comprising a genetic element comprising (i) an exterior protein binding sequence that binds the genetic element to an exterior protein (e.g., a non-pathogenic exterior protein, e.g., an Anellovirus ORF1 molecule or a splice variant or functional fragment thereof), (ii) a non-Anellovirus sequence (e.g., a non-Anellovirus origin of replication, e.g., as described herein), and (iii) a sequence encoding an effector.

The genetic element or any of the sequences within the genetic element can be obtained using any suitable method. Various recombinant methods are known in the art, such as, for example screening libraries from cells harboring viral sequences, deriving the sequences from a nucleic acid construct known to include the same, or isolating directly from cells and tissues containing the same, using standard techniques. Alternatively or in combination, part or all of the genetic element can be produced synthetically, rather than cloned.

In some embodiments, the nucleic acid construct includes regulatory elements, nucleic acid sequences homologous to target genes, and/or various reporter constructs for causing the expression of reporter molecules within a viable cell and/or when an intracellular molecule is present within a target cell.

Reporter genes are used for identifying potentially transfected cells and for evaluating the functionality of regulatory sequences. In general, a reporter gene is a gene that is not present in or expressed by the recipient organism or tissue and that encodes a polypeptide whose expression is manifested by some easily detectable property, e.g., enzymatic activity. Expression of the reporter gene is assayed at a suitable time after the DNA has been introduced into the recipient cells. Suitable reporter genes may include genes encoding luciferase, beta-galactosidase, chloramphenicol acetyl transferase, secreted alkaline phosphatase, or the green fluorescent protein gene (e.g., Ui-Tei et al., 2000 FEBS Letters 479: 79-82). Suitable expression systems are well known and may be prepared using known

techniques or obtained commercially. In general, the construct with the minimal 5' flanking region showing the highest level of expression of reporter gene is identified as the promoter. Such promoter regions may be linked to a reporter gene and used to evaluate agents for the ability to modulate promoter-driven transcription.

In some embodiments, the nucleic acid construct is substantially non-pathogenic and/or substantially non-integrating in a host cell or is substantially non-immunogenic in a host.

In some embodiments, the nucleic acid construct is double-stranded. In some embodiments the nucleic acid construct is single-stranded. In some embodiments, the nucleic acid construct is circular (e.g., a plasmid or a minicircle, e.g., as described herein). In some embodiments the nucleic acid construct is linear.

In some embodiments, a genetic element can be produced from the nucleic acid construct, e.g., in a host cell, e.g., as described herein. In some embodiments, a genetic element can be produced from the nucleic acid construct in the presence of a Rep molecule (e.g., a non-Anellovirus Rep molecule, e.g., an AAV Rep molecule, e.g., an AAV Rep protein, or a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto). In some embodiments, a genetic element cannot be produced from the nucleic acid construct by an Anellovirus Rep protein (e.g., an ORF2 molecule as described herein).

In some embodiments, the nucleic acid construct is in an amount sufficient to modulate one or more of 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.

IV. Compositions

The anellovectors described herein may also be included in pharmaceutical compositions with a pharmaceutical excipient, e.g., as described herein. In some embodiments, the pharmaceutical 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} anellovectors. In some embodiments, the pharmaceutical composition comprises about 10^5-10^{15}, 10^5-10^{10}, or 10^{10}-10^{15} anellovectors. In some embodiments, the pharmaceutical composition comprises about 10^8 (e.g., about 10^5, 10^6, 10^7, 10^8, 10^9, or 10^{10}) genomic equivalents/mL of the anellovector. In some embodiments, the pharmaceutical composition comprises 10^5-10^{10}, 10^6-10^{10}, 10^7-10^{10}, 10^8-10^{10}, 10^9-10^{10}, 10^5-10^6, 10^5-10^7, 10^5-10^8, 10^5-10^9, 10^5-10^{11}, 10^5-10^{12}, 10^5-10^{13}, 10^5-10^{14}, 10^5-10^{15}, or 10^{10}-10^{15} genomic equivalents/mL of the anellovector, e.g., as determined according to the method of Example 18 of PCT/US19/65995. In some embodiments, the pharmaceutical composition comprises sufficient anellovectors to deliver at least 1, 2, 5, or 10, 100, 500, 1000, 2000, 5000, 8,000, 1 x 10^4, 1 x 10^5, 1 x 10^6, 1 x 10^7 or greater copies of a genetic element comprised in the anellovectors per cell to a population of the eukaryotic cells. In some

embodiments, the pharmaceutical composition comprises sufficient anellovectors to deliver at least about 1×10^4, 1×10^5, 1×10^6, $1 \times$ or 10^7, or about 1×10^4-1×10^5, 1×10^4-1×10^6, 1×10^4-1×10^7, 1×10^5-1×10^6, 1×10^5-1×10^7, or 1×10^6-1×10^7 copies of a genetic element comprised in the anellovectors per cell to a population of the eukaryotic cells.

In some embodiments, the pharmaceutical composition has one or more of the following characteristics: the pharmaceutical composition meets a pharmaceutical or good manufacturing practices (GMP) standard; the pharmaceutical composition was made according to good manufacturing practices (GMP); the pharmaceutical composition has a pathogen level below a predetermined reference value, e.g., is substantially free of pathogens; the pharmaceutical composition has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants; or the pharmaceutical composition has low immunogenicity or is substantially non-immunogenic, e.g., as described herein.

In some embodiments, the pharmaceutical composition comprises below a threshold amount of one or more contaminants. Exemplary contaminants that are desirably excluded or minimized in the pharmaceutical composition include, without limitation, host cell nucleic acids (e.g., host cell DNA and/or host cell RNA), animal-derived components (e.g., serum albumin or trypsin), replication-competent viruses, non-infectious particles, free viral capsid protein, adventitious agents, and aggregates. In embodiments, the contaminant is host cell DNA. In embodiments, the composition comprises less than about 10 ng of host cell DNA per dose. In embodiments, the level of host cell DNA in the composition is reduced by filtration and/or enzymatic degradation of host cell DNA. In embodiments, the pharmaceutical composition consists of less than 10% (e.g., less than about 10%, 5%, 4%, 3%, 2%, 1%, 0.5%, or 0.1%) contaminant by weight.

In one aspect, the invention described herein includes a pharmaceutical composition comprising:

a) an anellovector comprising a genetic element comprising (i) a sequence encoding 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 a regulatory nucleic acid; and a proteinaceous exterior that is associated with, e.g., envelops or encloses, the genetic element; and

b) a pharmaceutical excipient.

Vesicles

In some embodiments, the composition further comprises a carrier component, e.g., a microparticle, liposome, vesicle, or exosome. In some embodiments, liposomes comprise spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes may be anionic, neutral or cationic. Liposomes are generally biocompatible, nontoxic, can deliver both

hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

Vesicles can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Vesicles may comprise without limitation DOTMA, DOTAP, DOTIM, DDAB, alone or together with cholesterol to yield DOTMA and cholesterol, DOTAP and cholesterol, DOTIM and cholesterol, and DDAB and cholesterol. Methods for preparation of multilamellar vesicle lipids are known in the art (see for example U.S. Pat. No. 6,693,086, the teachings of which relating to multilamellar vesicle lipid preparation are incorporated herein by reference). Although vesicle formation can be spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review). Extruded lipids can be prepared by extruding through filters of decreasing size, as described in Templeton et al., Nature Biotech, 15:647-652, 1997, the teachings of which relating to extruded lipid preparation are incorporated herein by reference.

As described herein, additives may be added to vesicles to modify their structure and/or properties. For example, either cholesterol or sphingomyelin may be added to the mixture to help stabilize the structure and to prevent the leakage of the inner cargo. Further, vesicles can be prepared from hydrogenated egg phosphatidylcholine or egg phosphatidylcholine, cholesterol, and dicetyl phosphate. (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review). Also, vesicles may be surface modified during or after synthesis to include reactive groups complementary to the reactive groups on the recipient cells. Such reactive groups include without limitation maleimide groups. As an example, vesicles may be synthesized to include maleimide conjugated phospholipids such as without limitation DSPE-MaL-PEG2000.

A vesicle formulation may be mainly comprised of natural phospholipids and lipids such as 1,2-distearoryl-sn-glycero-3-phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines and monosialoganglioside. Formulations made up of phospholipids only are less stable in plasma. However, manipulation of the lipid membrane with cholesterol reduces rapid release of the encapsulated cargo or 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) increases stability (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

In embodiments, lipids may be used to form lipid microparticles. Lipids include, but are not limited to, DLin-KC2-DMA4, C12-200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG may be formulated (see, e.g., Novobrantseva, Molecular Therapy-Nucleic Acids (2012) 1, e4; doi:10.1038/mtna.2011.3) using a spontaneous vesicle formation procedure. The component molar ratio may be about 50/10/38.5/1.5 (DLin-KC2-DMA or C12-200/disteroylphosphatidyl choline/cholesterol/PEG-DMG). Tekmira has a portfolio of approximately 95 patent families, in the U.S. and abroad, that are directed to various aspects of lipid microparticles and lipid microparticles formulations (see, e.g., U.S. Pat. Nos. 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 and 7,838,658 and European Pat. Nos. 1766035; 1519714; 1781593 and 1664316), all of which may be used and/or adapted to the present invention.

In some embodiments, microparticles comprise one or more solidified polymer(s) that is arranged in a random manner. The microparticles may be biodegradable. Biodegradable microparticles may be synthesized, e.g., using methods known in the art including without limitation solvent evaporation, hot melt microencapsulation, solvent removal, and spray drying. Exemplary methods for synthesizing microparticles are described by Bershteyn et al., Soft Matter 4:1787-1787, 2008 and in US 2008/0014144 A1, the specific teachings of which relating to microparticle synthesis are incorporated herein by reference.

Exemplary synthetic polymers which can be used to form biodegradable microparticles include without limitation aliphatic polyesters, poly (lactic acid) (PLA), poly (glycolic acid) (PGA), co-polymers of lactic acid and glycolic acid (PLGA), polycarprolactone (PCL), polyanhydrides, poly(ortho)esters, polyurethanes, poly(butyric acid), poly(valeric acid), and poly(lactide-co-caprolactone), and natural polymers such as albumin, alginate and other polysaccharides including dextran and cellulose, collagen, chemical derivatives thereof, including substitutions, additions of chemical groups such as for example alkyl, alkylene, hydroxylations, oxidations, and other modifications routinely made by those skilled in the art), albumin and other hydrophilic proteins, zein and other prolamines and hydrophobic proteins, copolymers and mixtures thereof. In general, these materials degrade either by enzymatic hydrolysis or exposure to water, by surface or bulk erosion.

The microparticles' diameter ranges from 0.1-1000 micrometers (μm). In some embodiments, their diameter ranges in size from 1-750 μm, or from 50-500 μm, or from 100-250 μm. In some embodiments, their diameter ranges in size from 50-1000 μm, from 50-750 μm, from 50-500 μm, or from 50-250 μm. In some embodiments, their diameter ranges in size from .05-1000 μm, from 10-1000 μm, from 100-1000 μm, or from 500-1000 μm. In some embodiments, their diameter is about 0.5 μm, about 10 μm, about 50 μm, about 100 μm, about 200 μm, about 300 μm, about 350 μm, about 400 μm, about 450 μm, about 500 μm, about 550 μm, about 600 μm, about 650 μm, about 700 μm, about 750 μm, about

800 µm, about 850 µm, about 900 µm, about 950 µm, or about 1000 µm. As used in the context of microparticle diameters, the term "about" means+/-5% of the absolute value stated.

In some embodiments, a ligand is conjugated to the surface of the microparticle via a functional chemical group (carboxylic acids, aldehydes, amines, sulfhydryls and hydroxyls) present on the surface of the particle and present on the ligand to be attached. Functionality may be introduced into the microparticles by, for example, during the emulsion preparation of microparticles, incorporation of stabilizers with functional chemical groups.

Another example of introducing functional groups to the microparticle is during post-particle preparation, by direct crosslinking particles and ligands with homo- or heterobifunctional crosslinkers. This procedure may use a suitable chemistry and a class of crosslinkers (CDI, EDAC, glutaraldehydes, etc. as discussed in more detail below) or any other crosslinker that couples ligands to the particle surface via chemical modification of the particle surface after preparation. This also includes a process whereby amphiphilic molecules such as fatty acids, lipids or functional stabilizers may be passively adsorbed and adhered to the particle surface, thereby introducing functional end groups for tethering to ligands.

In some embodiments, the microparticles may be synthesized to comprise one or more targeting groups on their exterior surface to target a specific cell or tissue type (e.g., cardiomyocytes). These targeting groups include without limitation receptors, ligands, antibodies, and the like. These targeting groups bind their partner on the cells' surface. In some embodiments, the microparticles will integrate into a lipid bilayer that comprises the cell surface and the mitochondria are delivered to the cell.

The microparticles may also comprise a lipid bilayer on their outermost surface. This bilayer may be comprised of one or more lipids of the same or different type. Examples include without limitation phospholipids such as phosphocholines and phosphoinositols. Specific examples include without limitation DMPC, DOPC, DSPC, and various other lipids such as those described herein for liposomes.

In some embodiments, the carrier comprises nanoparticles, e.g., as described herein.

In some embodiments, the vesicles or microparticles described herein are functionalized with a diagnostic agent. Examples of diagnostic agents include, but are not limited to, commercially available imaging agents used in positron emissions tomography (PET), computer assisted tomography (CAT), single photon emission computerized tomography, x-ray, fluoroscopy, and magnetic resonance imaging (MRI); and contrast agents. Examples of suitable materials for use as contrast agents in MRI include gadolinium chelates, as well as iron, magnesium, manganese, copper, and chromium.

Carriers

A composition (e.g., pharmaceutical composition) described herein may comprise, be formulated with, and/or be delivered in, a carrier. In one aspect, the invention includes a composition, e.g., a pharmaceutical composition, comprising a carrier (e.g., a vesicle, a liposome, a lipid nanoparticle, an exosome, a red blood cell, an exosome (e.g., a mammalian or plant exosome), a fusosome) comprising (e.g., encapsulating) a composition described herein (e.g., an anellovector, *Anellovirus*, or genetic element described herein).

In some embodiments, the compositions and systems described herein can be formulated in liposomes or other similar vesicles. Generally, liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes may be anionic, neutral or cationic. Liposomes generally have one or more (e.g., all) of the following characteristics: biocompatibility, nontoxicity, can deliver both hydrophilic and lipophilic drug molecules, can protect their cargo from degradation by plasma enzymes, and can transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679; and Zylberberg & Matosevic. 2016. Drug Delivery, 23:9, 3319-3329, doi: 10.1080/10717544.2016.1177136).

Vesicles can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Methods for preparation of multilamellar vesicle lipids are known (see, for example, U.S. Pat. No. 6,693,086, the teachings of which relating to multilamellar vesicle lipid preparation are incorporated herein by reference). Although vesicle formation can be spontaneous when a lipid film is mixed with an aqueeous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review). Extruded lipids can be prepared by, e.g., extruding through filters of decreasing size, as described in Templeton et al., Nature Biotech, 15:647-652, 1997.

Lipid nanoparticles (LNPs) are another example of a carrier that provides a biocompatible and biodegradable delivery system for the pharmaceutical compositions described herein. See, e.g., Gordillo-Galeano et al. European Journal of Pharmaceutics and Biopharmaceutics. Volume 133, December 2018, Pages 285-308. Nanostructured lipid carriers (NLCs) are modified solid lipid nanoparticles (SLNs) that retain the characteristics of the SLN, improve drug stability and loading capacity, and prevent drug leakage. Polymer nanoparticles (PNPs) are an important component of drug delivery. These nanoparticles can effectively direct drug delivery to specific targets and improve drug stability and controlled drug release. Lipid–polymer nanoparticles (PLNs), a new type of carrier that combines liposomes and

polymers, may also be employed. These nanoparticles possess the complementary advantages of PNPs and liposomes. A PLN is composed of a core–shell structure; the polymer core provides a stable structure, and the phospholipid shell offers good biocompatibility. As such, the two components increase the drug encapsulation efficiency rate, facilitate surface modification, and prevent leakage of water-soluble drugs. For a review, see, e.g., Li et al. 2017, Nanomaterials 7, 122; doi:10.3390/nano7060122.

Exosomes can also be used as drug delivery vehicles for the compositions and systems described herein. For a review, see Ha et al. July 2016. Acta Pharmaceutica Sinica B. Volume 6, Issue 4, Pages 287-296; doi.org/10.1016/j.apsb.2016.02.001.

Ex vivo differentiated red blood cells can also be used as a carrier for a composition described herein. See, e.g., WO2015073587; WO2017123646; WO2017123644; WO2018102740; WO2016183482; WO2015153102; WO2018151829; WO2018009838; Shi et al. 2014. Proc Natl Acad Sci USA. 111(28): 10131–10136; US Patent 9,644,180; Huang et al. 2017. Nature Communications 8: 423; Shi et al. 2014. Proc Natl Acad Sci USA. 111(28): 10131–10136.

Fusosome compositions, e.g., as described in WO2018208728, can also be used as carriers to deliver a composition described herein.

Membrane Penetrating Polypeptides

In some embodiments, the composition further comprises a membrane penetrating polypeptide (MPP) to carry the components into cells or across a membrane, e.g., cell or nuclear membrane. Membrane penetrating polypeptides that are capable of facilitating transport of substances across a membrane include, but are not limited to, cell-penetrating peptides (CPPs)(see, e.g., US Pat. No.: 8,603,966), fusion peptides for plant intracellular delivery (see, e.g., Ng et al., PLoS One, 2016, 11:e0154081), protein transduction domains, Trojan peptides, and membrane translocation signals (MTS) (see, e.g., Tung et al., Advanced Drug Delivery Reviews 55:281-294 (2003)). Some MPP are rich in amino acids, such as arginine, with positively charged side chains.

Membrane penetrating polypeptides have the ability of inducing membrane penetration of a component and allow macromolecular translocation within cells of multiple tissues in vivo upon systemic administration. A membrane penetrating polypeptide may also refer to a peptide which, when brought into contact with a cell under appropriate conditions, passes from the external environment in the intracellular environment, including the cytoplasm, organelles such as mitochondria, or the nucleus of the cell, in amounts significantly greater than would be reached with passive diffusion.

Components transported across a membrane may be reversibly or irreversibly linked to the membrane penetrating polypeptide. A linker may be a chemical bond, e.g., one or more covalent bonds

or non-covalent bonds. In some embodiments, the linker is a peptide linker. Such a linker may be between 2-30 amino acids, or longer. The linker includes flexible, rigid or cleavable linkers.

Combinations

In one aspect, the anellovector or composition comprising an anellovector described herein may also include one or more heterologous moiety. In one aspect, the anellovector or composition comprising a anellovector described herein may also include one or more heterologous moiety in a fusion. In some embodiments, a heterologous moiety may be linked with the genetic element. In some embodiments, a heterologous moiety may be enclosed in the proteinaceous exterior as part of the anellovector. In some embodiments, a heterologous moiety may be administered with the anellovector.

In one aspect, the invention includes a cell or tissue comprising any one of the anellovectors and heterologous moieties described herein.

In another aspect, the invention includes a pharmaceutical composition comprising a anellovector and the heterologous moiety described herein.

In some embodiments, the heterologous moiety may be a virus (e.g., an effector (e.g., a drug, small molecule), a targeting agent (e.g., a DNA targeting agent, antibody, receptor ligand), a tag (e.g., fluorophore, light sensitive agent such as KillerRed), or an editing or targeting moiety described herein. In some embodiments, a membrane translocating polypeptide described herein is linked to one or more heterologous moieties. In one embodiment, the heterologous moiety is a small molecule (e.g., a peptidomimetic or a small organic molecule with a molecular weight of less than 2000 daltons), a peptide or polypeptide (e.g., an antibody or antigen-binding fragment thereof), a nanoparticle, an aptamer, or pharmacoagent.

Targeting Moiety

In some embodiments, the composition or anellovector described herein may further comprise a targeting moiety, e.g., a targeting moiety that specifically binds to a molecule of interest present on a target cell. The targeting moiety may modulate a specific function of the molecule of interest or cell, modulate a specific molecule (e.g., enzyme, protein or nucleic acid), e.g., a specific molecule downstream of the molecule of interest in a pathway, or specifically bind to a target to localize the anellovector or genetic element. For example, a targeting moiety may include a therapeutic that interacts with a specific molecule of interest to increase, decrease or otherwise modulate its function.

Tagging or Monitoring Moiety

In some embodiments, the composition or anellovector described herein may further comprise a tag to label or monitor the anellovector or genetic element described herein. The tagging or monitoring moiety may be removable by chemical agents or enzymatic cleavage, such as proteolysis or intein splicing. An affinity tag may be useful to purify the tagged polypeptide using an affinity technique. Some examples include, chitin binding protein (CBP), maltose binding protein (MBP), glutathione-S-transferase (GST), and poly(His) tag. A solubilization tag may be useful to aid recombinant proteins expressed in chaperone-deficient species such as *E. coli* to assist in the proper folding in proteins and keep them from precipitating. Some examples include thioredoxin (TRX) and poly(NANP). The tagging or monitoring moiety may include a light sensitive tag, e.g., fluorescence. Fluorescent tags are useful for visualization. GFP and its variants are some examples commonly used as fluorescent tags. Protein tags may allow specific enzymatic modifications (such as biotinylation by biotin ligase) or chemical modifications (such as reaction with FlAsH-EDT2 for fluorescence imaging) to occur. Often tagging or monitoring moiety are combined, in order to connect proteins to multiple other components. The tagging or monitoring moiety may also be removed by specific proteolysis or enzymatic cleavage (e.g. by TEV protease, Thrombin, Factor Xa or Enteropeptidase).

Nanoparticles

In some embodiments, the composition or anellovector described herein may further comprise a nanoparticle. Nanoparticles include inorganic materials with a size between about 1 and about 1000 nanometers, between about 1 and about 500 nanometers in size, between about 1 and about 100 nm, between about 50 nm and about 300 nm, between about 75 nm and about 200 nm, between about 100 nm and about 200 nm, and any range therebetween. Nanoparticles generally have a composite structure of nanoscale dimensions. In some embodiments, nanoparticles are typically spherical although different morphologies are possible depending on the nanoparticle composition. The portion of the nanoparticle contacting an environment external to the nanoparticle is generally identified as the surface of the nanoparticle. In nanoparticles described herein, the size limitation can be restricted to two dimensions and so that nanoparticles include composite structure having a diameter from about 1 to about 1000 nm, where the specific diameter depends on the nanoparticle composition and on the intended use of the nanoparticle according to the experimental design. For example, nanoparticles used in therapeutic applications typically have a size of about 200 nm or below.

Additional desirable properties of the nanoparticle, such as surface charges and steric stabilization, can also vary in view of the specific application of interest. Exemplary properties that can be desirable in clinical applications such as cancer treatment are described in Davis et al, Nature 2008 vol.

7, pages 771-782; Duncan, Nature 2006 vol. 6, pages 688-701; and Allen, Nature 2002 vol. 2 pages 750-763, each incorporated herein by reference in its entirety. Additional properties are identifiable by a skilled person upon reading of the present disclosure. Nanoparticle dimensions and properties can be detected by techniques known in the art. Exemplary techniques to detect particles dimensions include but are not limited to dynamic light scattering (DLS) and a variety of microscopies such at transmission electron microscopy (TEM) and atomic force microscopy (AFM). Exemplary techniques to detect particle morphology include but are not limited to TEM and AFM. Exemplary techniques to detect surface charges of the nanoparticle include but are not limited to zeta potential method. Additional techniques suitable to detect other chemical properties comprise by ^{1}H, ^{11}B, and ^{13}C and ^{19}F NMR, UV/Vis and infrared/Raman spectroscopies and fluorescence spectroscopy (when nanoparticle is used in combination with fluorescent labels) and additional techniques identifiable by a skilled person.

Small molecules

In some embodiments, the composition or anellovector described herein may further comprise a small molecule. Small molecule moieties include, but are not limited to, small peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, synthetic polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic and inorganic compounds (including heterorganic and organomettallic compounds) generally having a molecular weight less than about 5,000 grams per mole, e.g., organic or inorganic compounds having a molecular weight less than about 2,000 grams per mole, e.g., organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, e.g., organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. Small molecules 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.

Examples of suitable small molecules include those described in, "The Pharmacological Basis of Therapeutics," Goodman and Gilman, McGraw-Hill, New York, N.Y., (1996), Ninth edition, under the sections: Drugs Acting at Synaptic and Neuroeffector Junctional Sites; Drugs Acting on the Central Nervous System; Autacoids: Drug Therapy of Inflammation; Water, Salts and Ions; Drugs Affecting Renal Function and Electrolyte Metabolism; Cardiovascular Drugs; Drugs Affecting Gastrointestinal Function; Drugs Affecting Uterine Motility; Chemotherapy of Parasitic Infections; Chemotherapy of Microbial Diseases; Chemotherapy of Neoplastic Diseases; Drugs Used for Immunosuppression; Drugs Acting on Blood-Forming organs; Hormones and Hormone Antagonists; Vitamins, Dermatology; and Toxicology, all incorporated herein by reference. Some examples of small molecules include, but are not limited to, prion drugs such as tacrolimus, ubiquitin ligase or HECT ligase inhibitors such as heclin,

histone modifying drugs such as sodium butyrate, enzymatic inhibitors such as 5-aza-cytidine, anthracyclines such as doxorubicin, beta-lactams such as penicillin, anti-bacterials, chemotherapy agents, anti-virals, modulators from other organisms such as VP64, and drugs with insufficient bioavailability such as chemotherapeutics with deficient pharmacokinetics.

In some embodiments, the small molecule is an epigenetic modifying agent, for example such as those described in de Groote et al. Nuc. Acids Res. (2012):1-18. Exemplary small molecule epigenetic modifying agents are described, e.g., in Lu et al. J. Biomolecular Screening 17.5(2012):555-71, e.g., at Table 1 or 2, incorporated herein by reference. In some embodiments, an epigenetic modifying agent comprises vorinostat or romidepsin. In some embodiments, an epigenetic modifying agent comprises an inhibitor of class I, II, III, and/or IV histone deacetylase (HDAC). In some embodiments, an epigenetic modifying agent comprises an activator of SirTI. In some embodiments, an epigenetic modifying agent comprises Garcinol, Lys-CoA, C646, (+)-JQI, I-BET, BICI, MS120, DZNep, UNC0321, EPZ004777, AZ505, AMI-I, pyrazole amide 7b, benzo[d]imidazole 17b, acylated dapsone derivative (e.e.g, PRMTI), methylstat, 4,4'-dicarboxy-2,2'-bipyridine, SID 85736331, hydroxamate analog 8, tanylcypromie, bisguanidine and biguanide polyamine analogs, UNC669, Vidaza, decitabine, sodium phenyl butyrate (SDB), lipoic acid (LA), quercetin, valproic acid, hydralazine, bactrim, green tea extract (e.g., epigallocatechin gallate (EGCG)), curcumin, sulforphane and/or allicin/diallyl disulfide. In some embodiments, an epigenetic modifying agent inhibits DNA methylation, e.g., is an inhibitor of DNA methyltransferase (e.g., is 5-azacitidine and/or decitabine). In some embodiments, an epigenetic modifying agent modifies histone modification, e.g., histone acetylation, histone methylation, histone sumoylation, and/or histone phosphorylation. In some embodiments, the epigenetic modifying agent is an inhibitor of a histone deacetylase (e.g., is vorinostat and/or trichostatin A).

In some embodiments, the small molecule is a pharmaceutically active agent. In one embodiment, the small molecule is an inhibitor of a metabolic activity or component. Useful classes of pharmaceutically active agents include, but are not limited to, antibiotics, anti-inflammatory drugs, angiogenic or vasoactive agents, growth factors and chemotherapeutic (anti-neoplastic) agents (e.g., tumour suppressers). One or a combination of molecules from the categories and examples described herein or from (Orme-Johnson 2007, Methods Cell Biol. 2007;80:813-26) can be used. In one embodiment, the invention includes a composition comprising an antibiotic, anti-inflammatory drug, angiogenic or vasoactive agent, growth factor or chemotherapeutic agent.

Peptides or proteins

In some embodiments, the composition or anellovector described herein may further comprise a peptide or protein. The peptide moieties may include, but are not limited to, a peptide ligand or antibody

fragment (e.g., antibody fragment that binds a receptor such as an extracellular receptor), neuropeptide, hormone peptide, peptide drug, toxic peptide, viral or microbial peptide, synthetic peptide, and agonist or antagonist peptide.

Peptides moieties may be linear or branched. The peptide has a length from about 5 to about 200 amino acids, about 15 to about 150 amino acids, about 20 to about 125 amino acids, about 25 to about 100 amino acids, or any range therebetween.

Some examples of peptides include, but are not limited to, fluorescent tags or markers, antigens, antibodies, antibody fragments such as single domain antibodies, ligands and receptors such as glucagon-like peptide-1 (GLP-1), GLP-2 receptor 2, cholecystokinin B (CCKB) and somatostatin receptor, peptide therapeutics such as those that bind to specific cell surface receptors such as G protein-coupled receptors (GPCRs) or ion channels, synthetic or analog peptides from naturally-bioactive peptides, anti-microbial peptides, pore-forming peptides, tumor targeting or cytotoxic peptides, and degradation or self-destruction peptides such as an apoptosis-inducing peptide signal or photosensitizer peptide.

Peptides useful in the invention described herein also include small 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 small antigen binding peptides may bind a cytosolic antigen, a nuclear antigen, an intra-organellar antigen.

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.

Oligonucleotide aptamers

In some embodiments, the composition or anellovector described herein may further comprise an oligonucleotide aptamer. Aptamer moieties are oligonucleotide or peptide aptamers. Oligonucleotide aptamers are single-stranded DNA or RNA (ssDNA or ssRNA) molecules that can bind to pre-selected targets including proteins and peptides with high affinity and specificity.

Oligonucleotide aptamers are nucleic acid species that may be engineered through repeated rounds of *in vitro* selection or equivalently, SELEX (systematic evolution of ligands by exponential enrichment) to bind to various molecular targets such as small molecules, proteins, nucleic acids, and even cells, tissues and organisms. Aptamers provide discriminate molecular recognition, and can be produced by chemical synthesis. In addition, aptamers may possess desirable storage properties, and elicit little or no immunogenicity in therapeutic applications.

Both DNA and RNA aptamers can show robust binding affinities for various targets. For example, DNA and RNA aptamers have been selected for t lysozyme, thrombin, human

immunodeficiency virus trans-acting responsive element (HIV TAR),(see en.wikipedia.org/wiki/Aptamer - cite_note-10), hemin, interferon γ, vascular endothelial growth factor (VEGF), prostate specific antigen (PSA), dopamine, and the non-classical oncogene, heat shock factor 1 (HSF1).

Peptide aptamers

In some embodiments, the composition or anellovector described herein may further comprise a peptide aptamer. Peptide aptamers have one (or more) short variable peptide domains, including peptides having low molecular weight, 12–14 kDa. Peptide aptamers may be designed to specifically bind to and interfere with protein-protein interactions inside cells.

Peptide aptamers are artificial proteins selected or engineered to bind specific target molecules. These proteins include of one or more peptide loops of variable sequence. They are typically isolated from combinatorial libraries and often subsequently improved by directed mutation or rounds of variable region mutagenesis and selection. *In vivo*, peptide aptamers can bind cellular protein targets and exert biological effects, including interference with the normal protein interactions of their targeted molecules with other proteins. In particular, a variable peptide aptamer loop attached to a transcription factor binding domain is screened against the target protein attached to a transcription factor activating domain. In vivo binding of the peptide aptamer to its target via this selection strategy is detected as expression of a downstream yeast marker gene. Such experiments identify particular proteins bound by the aptamers, and protein interactions that the aptamers disrupt, to cause the phenotype. In addition, peptide aptamers derivatized with appropriate functional moieties can cause specific post-translational modification of their target proteins, or change the subcellular localization of the targets.

Peptide aptamers can also recognize targets *in vitro*. They have found use in lieu of antibodies in biosensors and used to detect active isoforms of proteins from populations containing both inactive and active protein forms. Derivatives known as tadpoles, in which peptide aptamer "heads" are covalently linked to unique sequence double-stranded DNA "tails", allow quantification of scarce target molecules in mixtures by PCR (using, for example, the quantitative real-time polymerase chain reaction) of their DNA tails.

Peptide aptamer selection can be made using different systems, but the most used is currently the yeast two-hybrid system. Peptide aptamers can also be selected from combinatorial peptide libraries constructed by phage display and other surface display technologies such as mRNA display, ribosome display, bacterial display and yeast display. These experimental procedures are also known as biopannings. Among peptides obtained from biopannings, mimotopes can be considered as a kind of peptide aptamers. All the peptides panned from combinatorial peptide libraries have been stored in a special database with the name MimoDB.

VI. Methods of Use

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. In some embodiments, the anellovectors may be administered in a single dose, e.g., a first plurality. In some embodiments, anellovectors may be administered in at least two doses, e.g., a first plurality, followed by a second plurality. In some embodiments, the anellovectors may be administered in multiple doses, e.g., a first plurality, a second plurality, a third plurality, optionally a fourth plurality, optionally a fifth plurality, and/or optionally further pluralities.

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 infusable solutions or suspensions or suppositories or aerosols.

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.

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 a 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.

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. 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.

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., an 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.

VII. Administration/Delivery

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).

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.

In some embodiments, the subject to which administration of the pharmaceutical compositions is contemplated is a human. In some embodiments, the subject is a neonate, e.g., between 0 and 4 weeks of age. In some embodiments, the subject is an infant, e.g., between 4 weeks of age and 1 year of age. In some embodiments, the subject is a a child, e.g., between 1 year of age and 12 years of age. In some embodiments, the subject is less than 18 years of age. In some embodiments, the subject is an adolescent, e.g., between 12 years of age and 18 years of age. In some embodiments, the subject is above the age of 18. In some embodiments, the subject is a young adult, e.g., between 18 years of age and 25 years of age. In some embodiments, the subject is an adult, e.g., between 25 years of age to 50 years of age. In some embodiments, the subject is an older adult, e.g., an adult at least 50 years of age or older.

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.

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).

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).

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.

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.

In one aspect, the invention features a method of delivering an effector to a subject, e.g., a human subject, who has previously been administered an anellovector, e.g., a first plurality of anellovectors, the method comprising administration of a second plurality of anellovectors. In another aspect, the invention features a method of delivering an effector to a subject, e.g., a human subject, the method comprising administering a first plurality of anellovectors to the subject and subsequently administering to the subject a second plurality of anellovectors. In some emodiments, the methods described herein, further comprise administration of a third, fourth, fifth, and/or further plurality of anellovectors. In some embodiments, the first and second plurality are administred via the same route of administration, e.g., intravenous administration. In some embodiments, the first and second plurality are administered via different routes of administration. In some embodiments, the first plurality of anellovectors is administered to the subject as part of a first pharmaceutical composition. In some embodiments, the second plurality of anellovectors is administered to the subject as part of a second pharmaceutical composition.

In some embodiments, the first and the second plurality comprise about the same dosage of anellovectors, e.g., wherein the first plurality and the second plurality of anellovectors comprise about the same quantity and/or concentration of anellovectors. In some embodiments, the second plurality comprises 90-110%, e.g., 95-105% of the number of anellovectors in the first plurality. In some embodiments, the first plurality comprises a greater dosage of anellovectors than the second plurality, e.g., wherein the first plurality comprises a greater quantity and/or concentration of anellovectors relative to the second plurality. In some embodiments, the first plurality comprises a lower dosage of anellovectors than the second plurality, e.g., wherein the first plurality comprises a greater quantity and/or concentration of anellovectors relative to the second plurality. In some embodiments, the subject receives repeated doses of anellovectors, wherein the repeated doses are administered over the course of at least 1,

2, 3, 4, or 5 years. In some embodiments, the repeated dose is administered about every 1, 2, 3, or 4 weeks, or about every 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 months.

In some embodiments, the genetic element comprised in the anellovectors of the first plurality administered to the subject are detectable in the subject at least 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 days after administration thereof, e.g., by a high-resolution melting (HRM) assay, e.g., as described in Example 1. In some embodiments, the genetic element comprised in the anellovectors of the second plurality administered to the subject are detectable in the subject at least 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 days after administration thereof, e.g., by a high-resolution melting (HRM) assay, e.g., as described in Example 1.

In some embodiments, the first and/or second plurality of anellovectors administered to the subject comprises an effector. In some embodiments, the first and/or second plurality comprises an exogenous effector. In some embodiments, the first and/or second plurality comprises an endogenous effector. In some embodiments, the effector of the second plurality of anellovectors is the same effector as the effector of the first plurality of anellovectors. In some embodiments, the effector of the second plurality of anellovectors is different from the effector of the first plurality of anellovectors. In some embodiments, the second plurality of anellovectors delivers about the same number of copies of the effector to the subject as the number of effectors delivered by the first plurality of anellovectors. In some embodiments, the second plurality of anellovectors delivers the effector to the subject at a level of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of copies of the effector delivered to the subject by the first plurality of anellovectors (e.g., wherein the effector delivered by the first plurality may be the same or different form the effector delivered by the second plurality), In some embodiments, the second plurality of anellovectors delivers delivers more copies (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 500, or 1000-fold as many copies) of the effector to the subject than the first plurality of anellovectors. In some embodiments, the second plurality of anellovectors has a biological effect on the subject (e.g., knockdown of a target gene, or upregulation of a biomarker) that is no less than the biological effect of administration of the first plurality of anellovectors.

In some embodiments, identifying or selecting a subject on the basis of having received a plurality of anellovectors comprises performing an assay on a sample from the subject. In some embodiments, identifying or selecting a subject on the basis of having received a plurality of anellovectors comprises obtaining information from a third party (e.g., a laboratory), wherein the third party performed an assay on a sample from the subject. In some embodiments, identifying or selecting a

subject on the basis of having received a plurality of anellovectors comprises reviewing the subject's medical history.

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.

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, Kunjin 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, Rosavirus A, Ross river virus, Rotavirus A, Rotavirus B, Rotavirus C, Rubella virus, Sagiyama 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).

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

Example 7: Effective Transduction of Specific Cell Lines by Different Anellovectors Encoding Human Growth Hormone

Example 8: Purification of Ring 2 Anellovectors for rapid assessment of vector transduction

Example 1: Expression of a panel of full-length Anellovirus ORF1 proteins in mammalian cells

In this example, ORF1 proteins from a panel of anellovirus genomes were expressed in Expi-293 cells. ORF1 sequences for 8 different anelloviruses were identified; 3 Alphatorqueviruses (Ring1, Ring5, and Ring20), 3 Betatorqueviruses (Ring2, Ring9, and Ring10), and 2 Gammatorqueviruses (Ring3 and Ring4). Each nucleotide sequence was codon optimized for expression in human cells using IDT's codon optimization too. The codon optimized sequences were ordered as gene fragments from IDT, subcloned, then cloned into expression plasmids with a hEF1a promoter and with an N-terminal 3xFlag tag.

Each plasmid harboring the hEF1a-driven 3xFlag-ORF1 genes was transfected into Expi-293 cells. Briefly, 2.5µg of plasmid DNA was mixed with 2.5µL of PEI in 100µL of serum-free media. After a 20 minute incubation for complexation, PEI-DNA mixes were added dropwise to $1x10^6$ Expi-293 cells. Cells were then incubated at 37˚C at 8% CO2, shaking at 225 rpm for 2 days.

Transfected cell lysates were run on a Western blot. Briefly, $5x10^5$ cells in 100µL of media were collected and mixed with 25µL of 4x LDS sample buffer and 12.5µL of 20% BME. Samples were boiled at 95˚C for 5min before running. 20µL of each sample was run on a NuPAGE 4-12% Bis-Tris gel (Invitrogen) in 1x MES SDS Running buffer at 190V. Proteins were then transferred to a nitrocellulose membrane via wet transfer at 90V for 1hr. The blot was blocked for 1hr in 20mM Tris, 0.5 M NaCl, 0.1% Brij58 pH 7.5. A 1:2000 dilution of Mouse anti-Flag antibody was added to the blot and incubated overnight at room temperature. The blot was washed and soaked in a 1:5000 dilution of AP-rabbit anti-mouse secondary antibody for 2 hours. Then the blot was washed and soaked in blot developer solution until bands appeared.

Expression was observed for N-terminally 3xFlag-tagged anellovirus ORF1 proteins (FIG. 1). Each ran at the expected size for 3xFlag-tagged ORF1: Alphatorque Ring1 ORF1 at 91 kda, Betatorque Ring2 ORF1 at 79 kda, and Gammatorques Ring4 at 82 kda. Expression was also observed for a number of ORF1 proteins from other Anellovirus strains (data not shown).

Example 2: Replication of AAV ITR-flanked DNA by AAV Rep in the absence of AAV capsid

In this example, an ITR-flanked reporter gene construct was replicated off of a plasmid by AAV-Rep expression plasmids that did not produce AAV Capsid proteins. An expression vector with the full AAV2 Rep gene, producing Rep78, Rep68, Rep52, and Rep40, under control of the native AAV P5 promoter was constructed. Additionally, an expression vector with the full AAV2 Rep gene under control

of an inducible TRE-tight promoter was constructed. As a positive control for replication, an AAV2 RepCap expression plasmid was used (Cell BioLabs #VPK-422). As a replication target, a plasmid harboring an hrGFP reporter, driven by a CMV promoter and flanked by AAV ITRs, was used (Cell BioLabs #AAV-400). Each condition included the AAV pHelper plasmid (Cell BioLabs #340202), and plasmids expressing the Ring2 ORF1 and ORF2 proteins. Plasmids were transfected into Expi293 cells using PEI. Four days post-transfection, cell pellets were collected.

Total DNA from each sample was then run on a Southern blot. Briefly, total DNA was isolated from the cell pellets, digested with restriction endonucleases, run on an agarose gel, and transferred to a nylon-membrane. Three untransfected DNA controls were included on the Southern blot; pITR-hrGFP plasmid, ITR-hrGFP genome DNA produced by extracting the ITR-hrGFP DNA from the plasmid via restriction enzyemes, and pHelper plasmid DNA. The blot was probed for hrGFP and pHelper DNA sequence using biotinylated DNA fragments, and detected with streptavidin-linked IRDye800 on a LiCor Odyssey imager (FIG. 2). To determine relative replication efficiencies, the densities of the ITR-hrGFP genome bands and the pHelper bands on the Southern were quantified using ImageJ. The amount of replicated ITR-hrGFP was normalized to the amount of pHelper plasmid transfection input, then analyzed relative to pRepCap replication levels.

Southern blot analysis demonstrated that the CAP-free AAV Rep constructs successfully replicated ITR-hrGFP genomes from the plasmid (FIG. 2). After quantifying the band intensities and normalizing for transfection input, the P5-driven Rep construct replicated the 60% of the ITR-hrGFP genomes of RepCap, while the TRE-tight-driven Rep performed nearly identically to RepCap. These results demonstrated that ITR-containing DNA constructs can be efficiently replicated with Cap-free AAV-Rep expression vectors. Furthermore, the TRE-tight-promoter Rep construct replicated the DNA to the same levels as the standard pRepCap plasmid, without producing the AAV Cap proteins.

Example 3: Production of AnelloVectors through cross-packing with AAV variant transgene reporter constructs.

In this example, anellovectors were shown to be produced through co-expression of Anello ORF proteins (ORF1, ORF2), in conjunction with traditional AAV production components (AAV rep expressing plasmids and pHelper plasmid) and a transgene plasmid encompassing the reporter nanoluciferase (nLuc) along with Anellovirus non-coding sequences flanked between AAV2 ITRs. The transgenes were of a size similar to the corresponding Anellovirus genome (plus or minus 0.3kb). In other variations, non-coding Anellovirus sequences were included because, in some experiments, vector DNA was found to package more efficiently when comprising Anellovirus sequences. These anellovectors were produced as Anellovirus protein exteriors encapsulating a reporter construct containing AAV2 ITRs.

In this example, replication and amplification of the transgene occurred through AAV Rep-mediated activities, while the components required for encapsulation of the replicated transgene occurred through trans-expression of the Anellovirus ORF1 and ORF2 proteins.

Briefly, the above listed plasmids were co-transfected, using PEI-Pro, into Expi-293F cells at a plasmid to plasmid ratio of 1:1 and DNA to PEI molar ratio of 1:1. At 4 days post transfection (dpt), cells were harvested and pelleted away from the conditioned media (CM) by centrifugation. Cells were then lysed by either chemical or mechanical means, treated with a DNase in the presence of a protease inhibitor, and then treated with a detergent for lipid removal. Anellovector particles were then isolated away from cell debris and host protein through two ultracentrifugation steps. The first spin consisted of a 2-step CsCl density gradient in which material between densities of 1.25g/ml and 1.4g/ml was extracted. After an overnight dialysis, this material was then applied onto a linear CsCl gradient. Fractions were then extracted in 1ml aliquots, refractive indexes were taken, and the material was desalted for quantification using quantitative real-time PCR (qPCR) to detect DNase protected transgene specific genomes. Fractions within the density range of 1.27-1.35 were pooled together and then dialyzed overnight using a 50kDa MWCO in buffer containing 0.001% PS-80. Material was then concentrated using a centrifugal membrane concentrator with a MWCO of 100kDa. Final material was then quantified using quantitative real-time PCR (qPCR) to detect Anelloviral nucleic acids.

FIG. 3A shows the vector genome copy number obtained by qPCR of an amplicon in the nanoluciferase transgene in the linear gradient fractions. A clear peak in vector copies was observed at a fraction density of 1.31 g/mL. In contrast, as shown in FIG. 3B, if the ORF1 anellovirus gene was omitted from the transfection, no such peak was observed. These data indicate that the vector signal was dependent on ORF1 being expressed. Together, these data are consistent with an Anellovector being produced.

Example 4: Delivery of reporter constructs via Anellovector transduction in mammalian and non-human primate cells of different origins

In this example, anellovectors were produced through co-expression of Anellovirus ORF proteins (ORF1, ORF2) in conjunction with traditional AAV production components (AAV rep expressing plasmids and pHelper plasmid) and a transgene plasmid encompassing a reporter along with Anellovirus non-coding sequences flanked between AAV2-ITRs. In these cases, anellovectors were made with transgenes expressing a luciferase reporter (nLuc) or fluorescent reporters (mCherry, GFP). In this example, sucessful transduction of human (Vero) and non-human primate (Vero) cell lines was demonstrated using R2-anellovectors encompassing ITR-flanked transgenes expressing nLuc, mCherry or GFP.

Vectors were purified over linear density gradients then dialized using 50kDa MWCO membranes to reduce transgene protein carry-over. Transductions were performed through incubation of vector material on Vero and IGR-OV1 cells for 3 hours at 37°C – conditions which permit binding and internalization of the virus in the cells. Day 0 (D0) samples were harvested immediately following this incubation (for nLuc transductions) and remaining samples were incubated for 2 days prior to analysis. For R2-nLuc vectors, luciferase assays were performed which measure the amount of the nLuc protein through a luminescent based readout. As shown in FIGS. 4-5, transduction with anellovectors resulted in a 1.5-log increase from D0 to D2, whereas transductions with material not expressing Anellovirus ORF1 and ORF2 proteins decreased from D0 to D2. 3-log increases were observed in IGR-OV1 cells (FIG. 5). In both cell lines, identical MOIs were used (0.4). These results were further highlighted by transduction of Vero and IGR-OV1 cells with anellovectors carrying additional reporters (i.e., GFP and mCherry) at an MOI of 0.2. Microscopy showed successful transduction of both Vero and IGR-OV1 cells by these anellovectors and expression of the respective fluorescent reporter. Control cells transduced with material not expressing Anellovirus ORF1 and ORF2 proteins did not show substantial fluorescence by either reporter.

Example 5: Generation of Anello-AAV vectors and successful transduction in MOLT4 cells

In this example, whether Anellovirus capsid protein (ORF1) could package non-cognate replicating ssDNA in cyto was tested. Several AAV components (plasmids encoding AAV Rep, reporter transgene, and a pHelper plasmid component) that can generate ssDNA encoding a red fluorescent "mKate" reporter gene packaged by ORF1 protein were used. The following transfections were carried out in 293F cells using PEI:

(1) the main components of AAV particle generation minus the AAV Capsid plasmid (mKate plasmid, AAV Rep, and pHelper plasmid),

(2) the main components of the AAV system plus ORF1 and ORF2 of Ring2, or

(3) the main components of the AAV system with Ring2 ORF2 only.

After four days, cells were lysed, then processed over CsCl step gradients (FIG. 6). Fractions within the density range of 1.2-1.4g/ml were collected and dialyzed then used to infect MOLT4 cells (human T Lymphoblast cell line) at an MOI of 1 vector per cell. Positive transduction events were measured 3 days post infection (dpi) through quantification of mKate expressing cells using flow cytometry. Condition 1, which only contained the AAV replication machinery and the mKate transgene,

failed to give a positive population of cells expressing mKate, while condition 2, containing ORF1 and 2 alongside the AAV replication machinery, resulted in 35% of the cells expressing mKate.

To further confirm whether this was a true transduction event, condition 3 was introduced, in which the capsid protein of Anelloviruses (ORF1) was left out. This resulted in no detectable transduction events, suggesting that in the setting of condition 2, we were able to transduce MOLT4 cells and that this transduction was ORF1-dependent. Further work extended these transductions to additional cell types and a Ring 4.0 Anello-AAV vector. Interestingly, when transductions were performed, there appeared to be a higher transduction efficiency of Raji cells for Ring2 vectors and 293T cells for Ring4.

Example 6: Engineered Ring2 Anellovirus DNA replicates through AAV Rep protein

Ring2 Anellovirus genomes have been shown, e.g., as described herein, to be capable of naturally replicating in MOLT-4 cells, but have thus far replicated poorly in HEK293 cells. To drive more robust genome replication in the tractable HEK293 cell line, versions of Ring2 were engineered to harbor known *cis* elements for AAV replication. In wild-type AAV, AAV Rep proteins bind to DNA sequences (*cis* elements) within the AAV ITR and drive DNA replication. The minimal sequences required for this activity were identified herein as a "Rep binding motif" (RBM) and a "terminal resolution site" (TRS). In this example, 62bp of AAV ITR sequence containing these sites was incorporated into the 3' non-coding region (NCR) of the Ring2 genome (FIG. 7A).

To test whether AAV Rep proteins drive replication of the Ring2+RBM/TRS DNA, plasmids harboring the engineered Anellovirus genome comprising the AAV ITR elements (RBM and TRS) were co-transfected into Expi-293 cells with or without *trans*-expressed AAV Rep. Total DNA was harvested four days post-transfection, digested to linearize the plasmid and to degrade non-replicated DNA with DpnI, and then run on Southern blots probing for Ring2 genomes (FIG. 7B). For wild-type Ring2 genomes without AAV-RBM/TRS, linearized input plasmid DNA was observed (lanes 1 and 3), but was degraded in the presence of DpnI (lanes 2 and 4), indicating that the DNA did not replicate in the cells. However, Ring2 with RBM/TRS in the 3' NCR did successfully replicate in the presence of AAV Rep, as indicated by a DpnI-resistant band (lane 8, green arrow). Without Rep, the linearized plasmid (lane 5) was digested by DpnI (lane 6), confirming that replication was Rep-dependent.

These data demonstrated successful engineering of a system for replication of Anellovirus DNA in Expi-HEK293 cells. Without wishing to be bound by theory, it is contemplated that *in vitro* circularization can be used to remove the plasmid backbone from Ring2-3'NCR-RBM/TRS, and that the resulting construct can be replicated with AAV-Rep and/or packaged using *trans*-expressed Ring2 ORF1 protein.

Example 7: Effective Transduction of Specific Cell Lines by Different Anellovectors Encoding Human Growth Hormone

The above examples have demonstrated the production of anellovectors by taking advantage of the AAV replication machinery in Expi293 cells, including anellovectors encoding fluorescent and luminescent payloads that are able to transduce cell lines *in vitro*. In this example, anellovectors encoding human growth hormone (hGH), a biologically active payload, were prepared that can be suitable for *in vivo* experiments. Briefly, Expi293 cells were transfected with plasmids required to produce the viral vectors (payload, AAV Rep, and pHelper) and either AAV2 capsid (positive control), RING2 capsid, RING9 capsid, or no capsid (negative control). Four days after transfection, cells were harvested and lysed by two rounds of freeze-thaws in 0.5% Triton X-100-containing buffer. Lysates were then treated with benzonase, followed by partial vector purification using cesium chloride step gradient. Step gradient material was dialyzed overnight to remove cesium chloride and then incubated with either human ovarian cancer cell line IGR-OV1 or monkey kidney cell line Vero for 3 hours. After this treatment, cells were washed with PBS three times to remove any contaminating DNA or protein, including carryover hGH from the vector production step. Fresh medium was added to transduced cells and incubated in at 37°C and 5% CO_2. Culture medium was harvested after 30 minutes (day 0 time point), 48 hours (day 2 time point), and 72 hours (day 3 time point), to quantify by ELISA the amount of hGH secreted by transduced cells.

As shown in FIGS. 8A-8B, there was an increase in the amount of secreted hGH in the culture medium of IGR-OV1 cells (FIG. 8A) and Vero cells (FIG. 8B) transduced with RING2 or RING9 vectors. AAV2 carrying hGH (positive control) also showed secretion of hGH on days 2 and 3, albeit at lower levels. Samples treated with the negative control did not demonstrate a similar increase in the amount of secreted hGH. These data demonstrated successful production of two transduction-competent anellovectors with different capsids, each encoding a biologically active payload.

Example 8: Purification of Ring 2 Anellovectors for rapid assessment of vector transduction

Assessing viral transduction without partially purifying vectors has historically been difficult due to high cell death caused by crude lysates. In this example, a quick method is described that allows the direct analysis of lysates, which bypasses the current 2-day process of vector purification, and allows decisions to be made faster concerning improvements in vector production or design. Lysates from 293F cells transfected with Ring2-ITR-nanoLuciferase (nLuc) vectors produced in either the presence (+ AAV Rep) or the absence (- AAV Rep) of all necessary components. Samples were clarified then diluted 1:1 in a buffer to adjust to pH 9 and lower the conductivity to 15mS/cm. Adjusted lysates were then loaded onto MustangQ columns and unbound material was collected. Bound material was eluted using a buffer

containing high salt with a neutral pH. Samples were then assessed for vector recovery by qPCR and transduction assays. Transduction assays were performed by adding 100ul (approx. 1/20) of total eluted samples onto IGR cells and measuring nLuc activity at Day 0 and Day 2. Transduction was measured by an increase in luminescence from D0 to D2.

As shown in FIG. 9, only samples in which all necessary plasmids were co-transfected showed positive transduction signals. Furthermore, crude cell lysates resulted in high cell death after 24 h. These results demonstrated a quick procedure (30 minutes of hands-on time) by which we can concentrate and partially purify anellovectors from crude cell lysates to measure transduction efficiencies. This approach can be used as a screening method to improve the througput of production and design optimization.

What is claimed is:

1. A viral particle comprising a circular DNA comprising (i) an AAV origin of replication, (ii) a promoter operably linked to a sequence encoding a therapeutic RNA or polypeptide, and (iii) a sequence that binds an Anellovirus ORF1 molecule, the circular DNA being encapsidated by a capsid comprising an Anellovirus ORF1 molecule.

2. A vector comprising:

a) a proteinaceous exterior comprising an Anellovirus ORF1 molecule; and

b) a genetic element comprising a non-Anellovirus origin of replication;

optionally wherein the genetic element further comprises: (i) a nucleic acid sequence encoding an exogenous effector, and/or (ii) a promoter element operatively linked to the nucleic acid sequence encoding the exogenous effector.

3. A genetic element comprising:

a protein binding sequence that specifically binds an Anellovirus ORF1 molecule (e.g., a 5' UTR); and

an AAV origin of replication, e.g., comprised in a first AAV inverted terminal repeat (ITR);

optionally, a nucleic acid sequence encoding an exogenous effector (e.g., a therapeutic exogenous effector); and

optionally, a promoter element operatively linked to the nucleic acid sequence encoding the exogenous effector.

4. A system comprising:

a) a first nucleic acid, wherein the first nucleic acid is a genetic element or a genetic element construct, the first nucleic acid comprising:

an AAV origin of replication, e.g., comprised in a first AAV inverted terminal repeat (ITR);

optionally, a nucleic acid sequence encoding an exogenous effector (e.g., a therapeutic exogenous effector); and

optionally, a promoter element operatively linked to the nucleic acid sequence encoding the exogenous effector;

b) a second nucleic acid encoding an Anellovirus ORF1 molecule.

5. A method of delivering an exogenous effector to a target cell (e.g., a vertebrate cell, e.g., a mammalian cell, e.g., a human cell), the method comprising introducing into the cell a vector of claim 2.

6. A method of treating or preventing a disease or disorder in a subject in need thereof, the method comprising introducing into the subject a vector of claim 2.

7. A method of making a therapeutic composition, comprising:

(a) providing one or a plurality of host cells comprising exogenous DNA comprising

(i) an AAV origin of replication,

(ii) a promoter operably linked to a sequence encoding a therapeutic effector (e.g., a therapeutoic RNA or polypeptide),

(iii) a sequence encoding an Anellovirus ORF1 molecule,

(iv) optionally a sequence encoding an Anellovirus ORF2 molecule,

(v) optionally a sequence encoding an AAV REP2 sequence

(vi) optionally a sequence encoding one or a plurality of helper proteins, e.g., an Adenovirus helper protein, e.g., an E2A molecule, an Adenovirus E4 molecule, and/or an Adenovirus VARNA molecule;

(b) culturing the one or plurality of host cells under conditions suitable for formation of vectors (e.g., anellovectors, e.g., viral particles) comprising a proteinaceous exterior (e.g., capsid) comprising a sufficient number of the ORF1 molecules to enclose (e.g., encapsidate) the genetic element;

(c) purifying the vectors produced in step (b) from the cell culture,

thereby making a therapeutic composition.

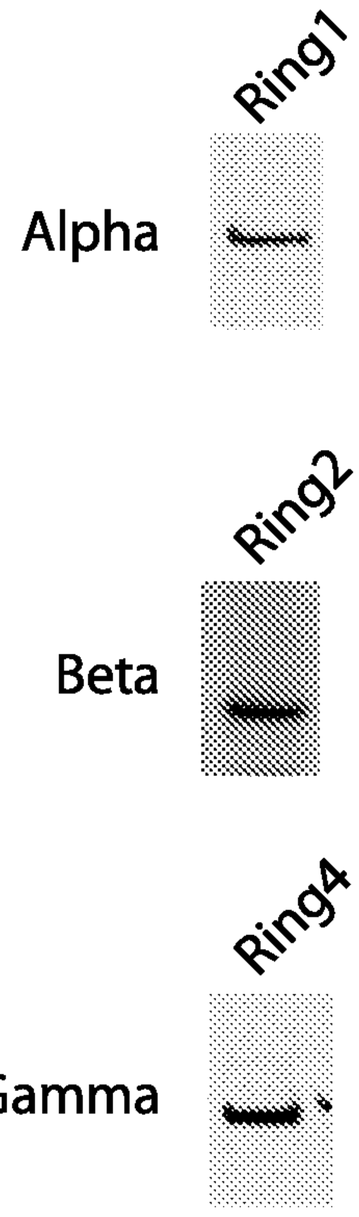
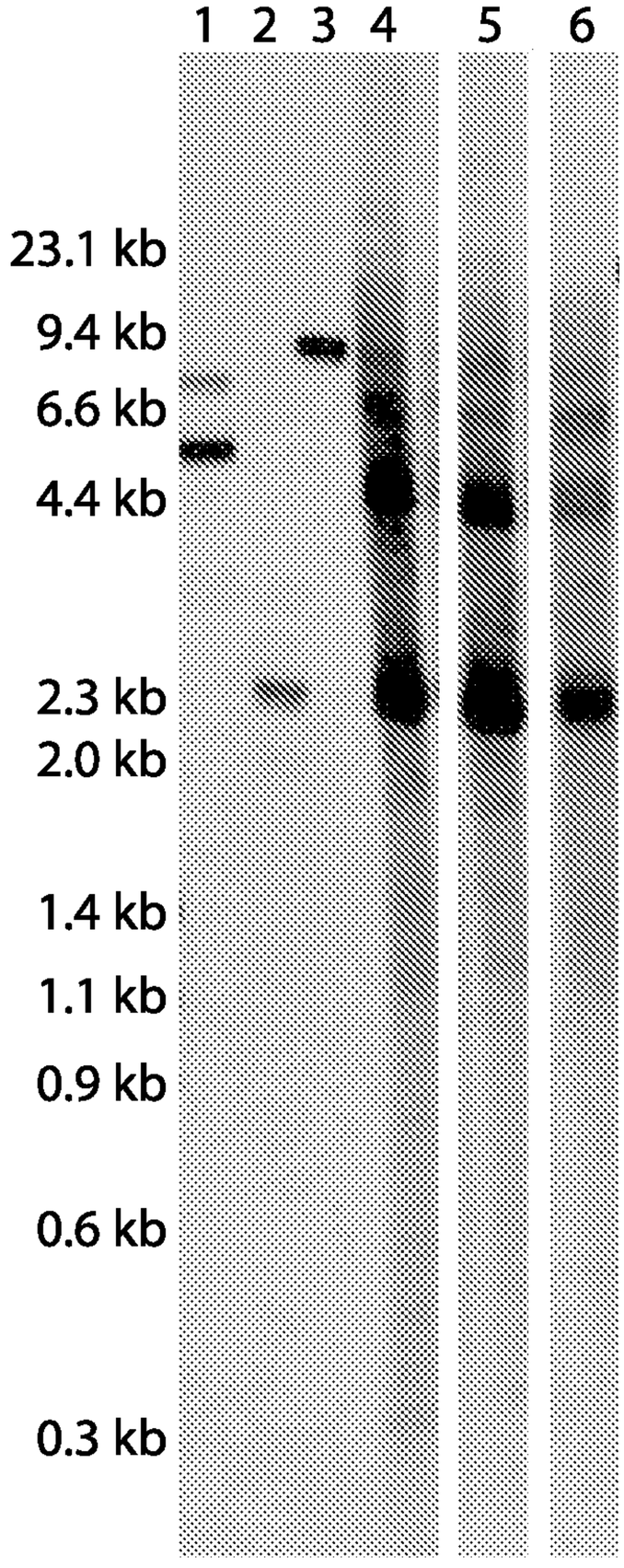


FIG. 1



Lane	Rep
1	pITR-hrGFP Plasmid control
2	ITR-hrGFP genome control
3	pHelper Plasmid control
4	P5_Rep
5	TRE-t_Rep
6	RepCap

FIG. 2

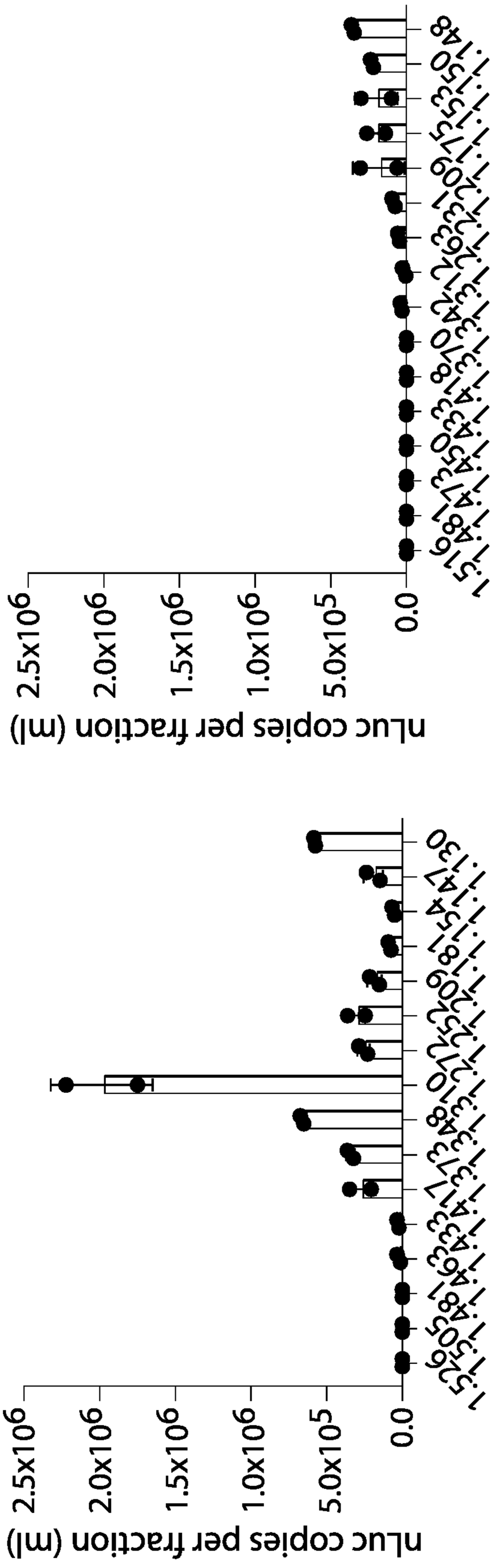

nluc copies per fraction (ml)
2.5x10^6
2.0x10^6
1.5x10^6
1.0x10^6
5.0x10^5
0.0
1:5:1:48
1:5:1:50
1:5:1:52
1:5:1:54
1:5:1:56
1:5:1:58
1:5:1:60
1:5:1:62
1:5:1:64
1:5:1:66
1:5:1:68
1:5:1:70
1:5:1:72
1:5:1:74
1:5:1:76
1:5:1:78
1:5:1:80
1:5:1:82
1:5:1:84
1:5:1:86
1:5:1:88
1:5:1:90
1:5:1:92
1:5:1:94
1:5:1:96
1:5:1:98
1:5:1:100
1:5:1:102
1:5:1:104
1:5:1:106
1:5:1:108
1:5:1:110
1:5:1:112
1:5:1:114
1:5:1:116
1:5:1:118
1:5:1:120
1:5:1:122
1:5:1:124
1:5:1:126
1:5:1:128
1:5:1:130
nluc copies per fraction (ml)
2.5x10^6
2.0x10^6
1.5x10^6
1.0x10^6
5.0x10^5
0.0
1:5:1:48
1:5:1:50
1:5:1:52
1:5:1:54
1:5:1:56
1:5:1:58
1:5:1:60
1:5:1:62
1:5:1:64
1:5:1:66
1:5:1:68
1:5:1:70
1:5:1:72
1:5:1:74
1:5:1:76
1:5:1:78
1:5:1:80
1:5:1:82
1:5:1:84
1:5:1:86
1:5:1:88
1:5:1:90
1:5:1:92
1:5:1:94
1:5:1:96
1:5:1:98
1:5:1:100
1:5:1:102
1:5:1:104
1:5:1:106
1:5:1:108
1:5:1:110
1:5:1:112
1:5:1:114
1:5:1:116
1:5:1:118
1:5:1:120
1:5:1:122
1:5:1:124
1:5:1:126
1:5:1:128
1:5:1:130

FIG. 3B

FIG. 3A

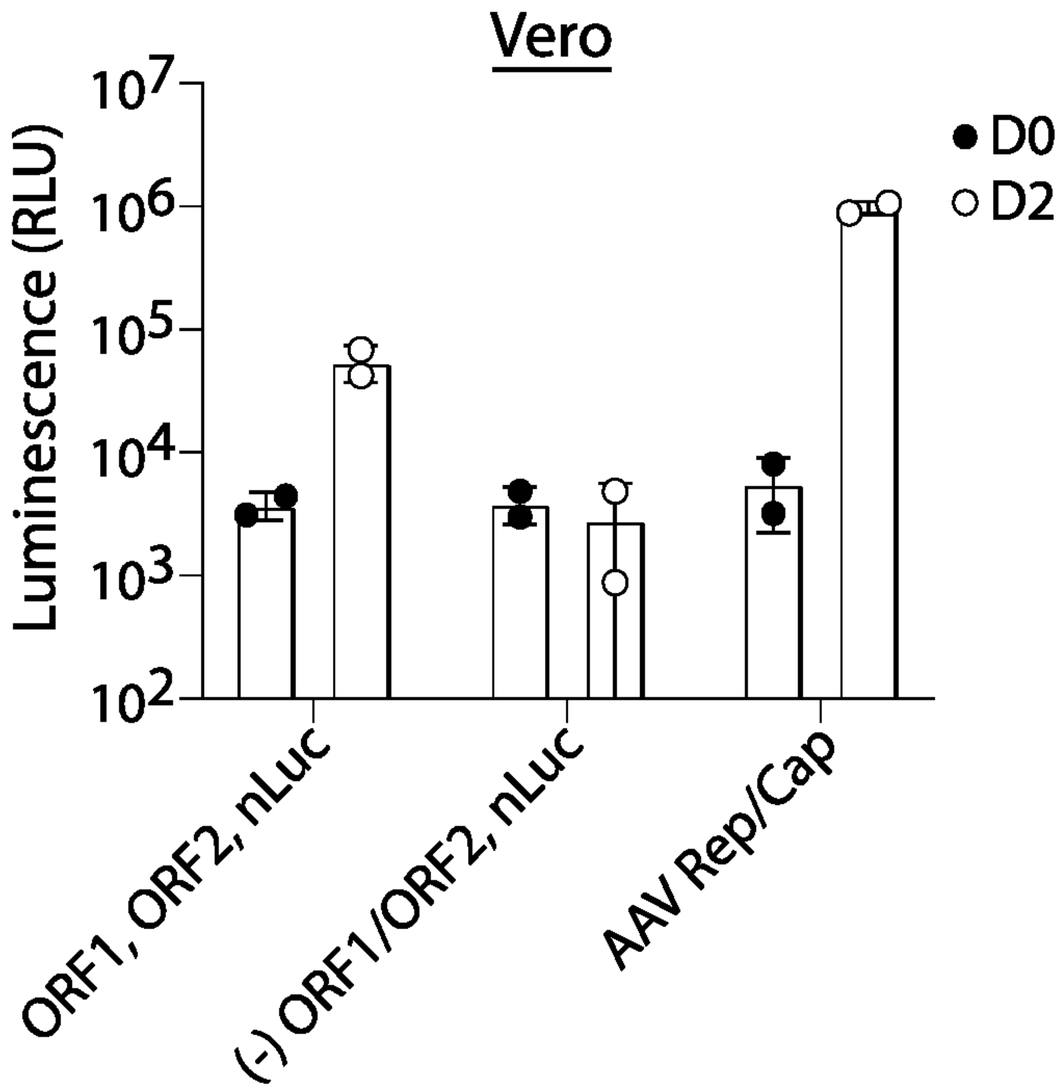

Vero
Luminescence (RLU)
10^7
10^6
10^5
10^4
10^3
10^2
D0
D2
ORF1, ORF2, nLuc
(-) ORF1/ORF2, nLuc
AAV Rep/Cap

FIG. 4

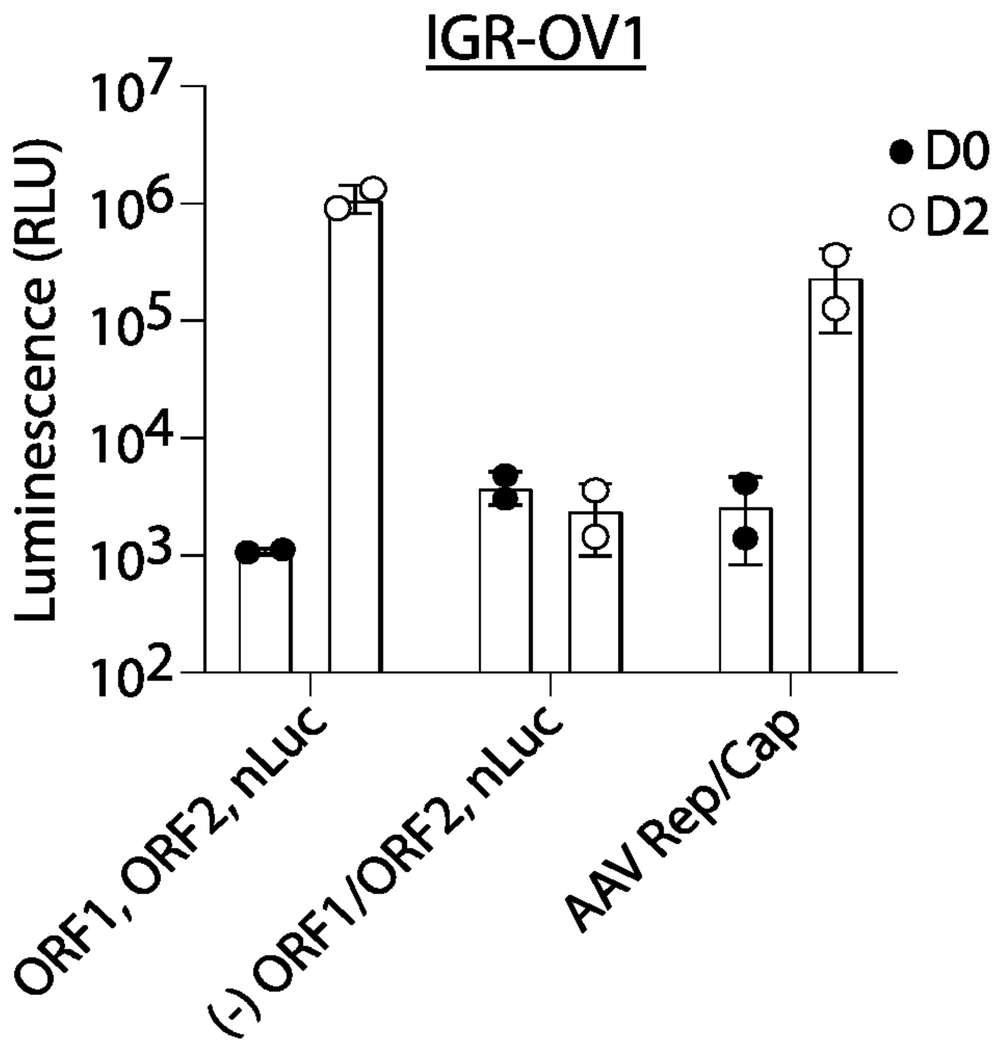

IGR-OV1
Luminescence (RLU)
10^7
10^6
10^5
10^4
10^3
10^2
D0
D2
ORF1, ORF2, nLuc
(-) ORF1/ORF2, nLuc
AAV Rep/Cap

FIG. 5

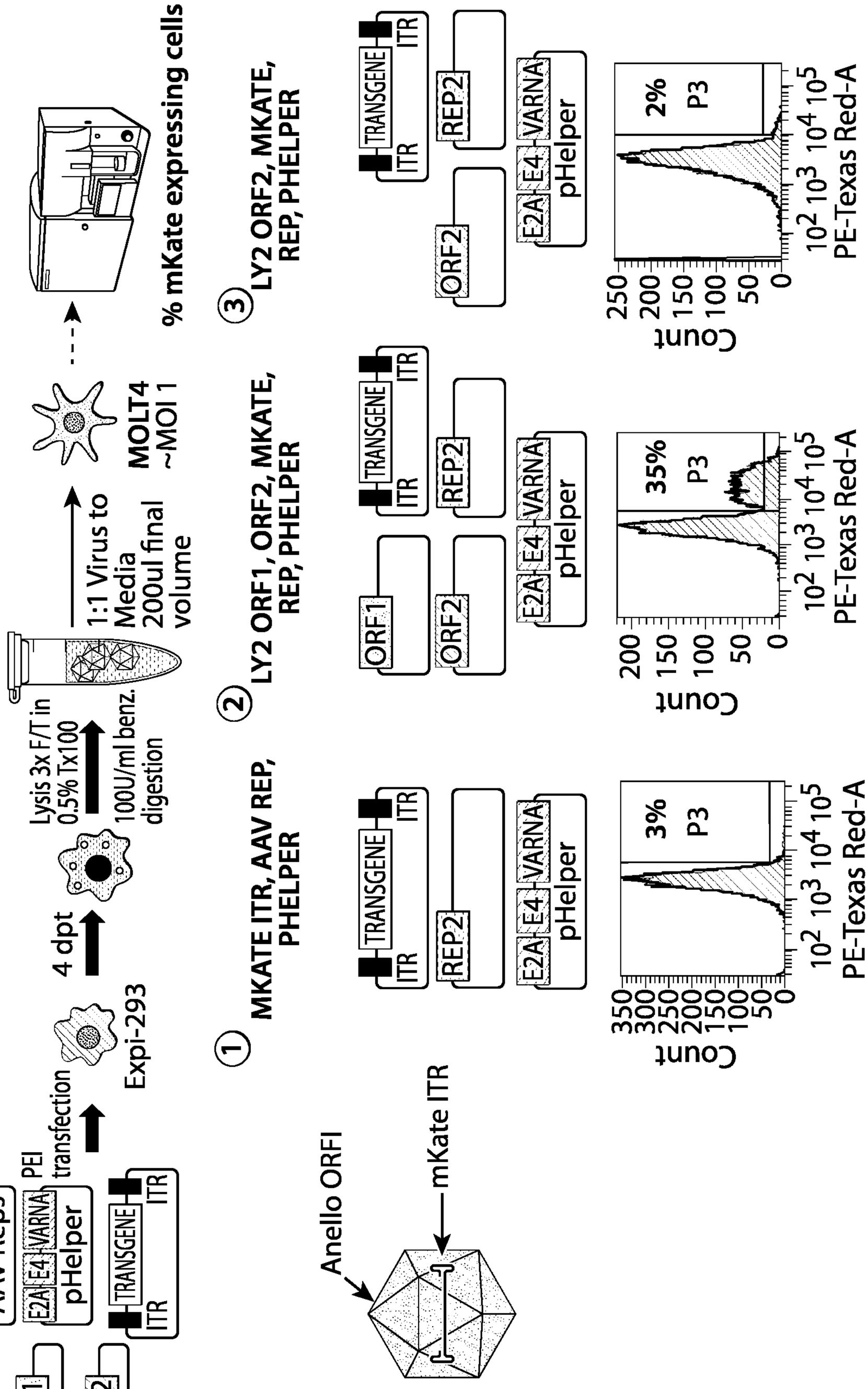

Anello ORFI
mKate ITR
Expi-293
PEI transfection
4 dpt
Lysis 3x F/T in 0.5% Tx100
100U/ml benz. digestion
1:1 Virus to Media 200ul final volume
MOLT4 ~MOI 1
% mKate expressing cells
2%
35%
3%
P3
Count
PE-Texas Red-A
TRANSGENE
ITR
REP2
AAV Reps
ORF1
ORF2
E2A
E4
VARNA
pHelper
① MKATE ITR, AAV REP, PHELPER
② LY2 ORF1, ORF2, MKATE, REP, PHELPER
③ LY2 ORF2, MKATE, REP, PHELPER

FIG. 6

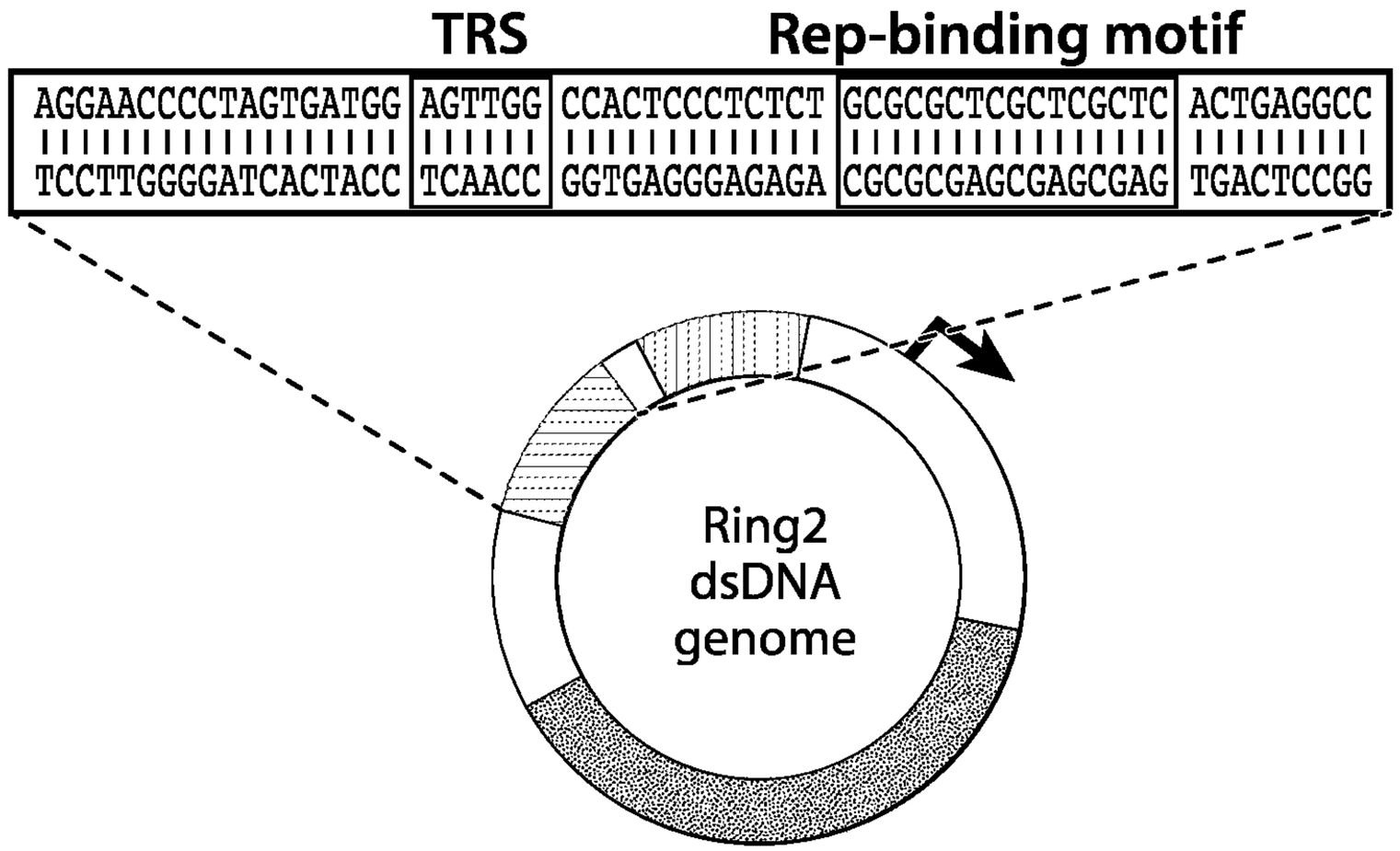

TRS
Rep-binding motif
AGGAACCCCTAGTGATGG AGTTGG CCACTCCCTCTCT GCGCGCTCGCTCGCTC ACTGAGGCC
TCCTTGGGGATCACTACC TCAACC GGTGAGGGAGAGA CGCGCGAGCGAGCGAG TGACTCCGG
Ring2
dsDNA
genome

FIG. 7A

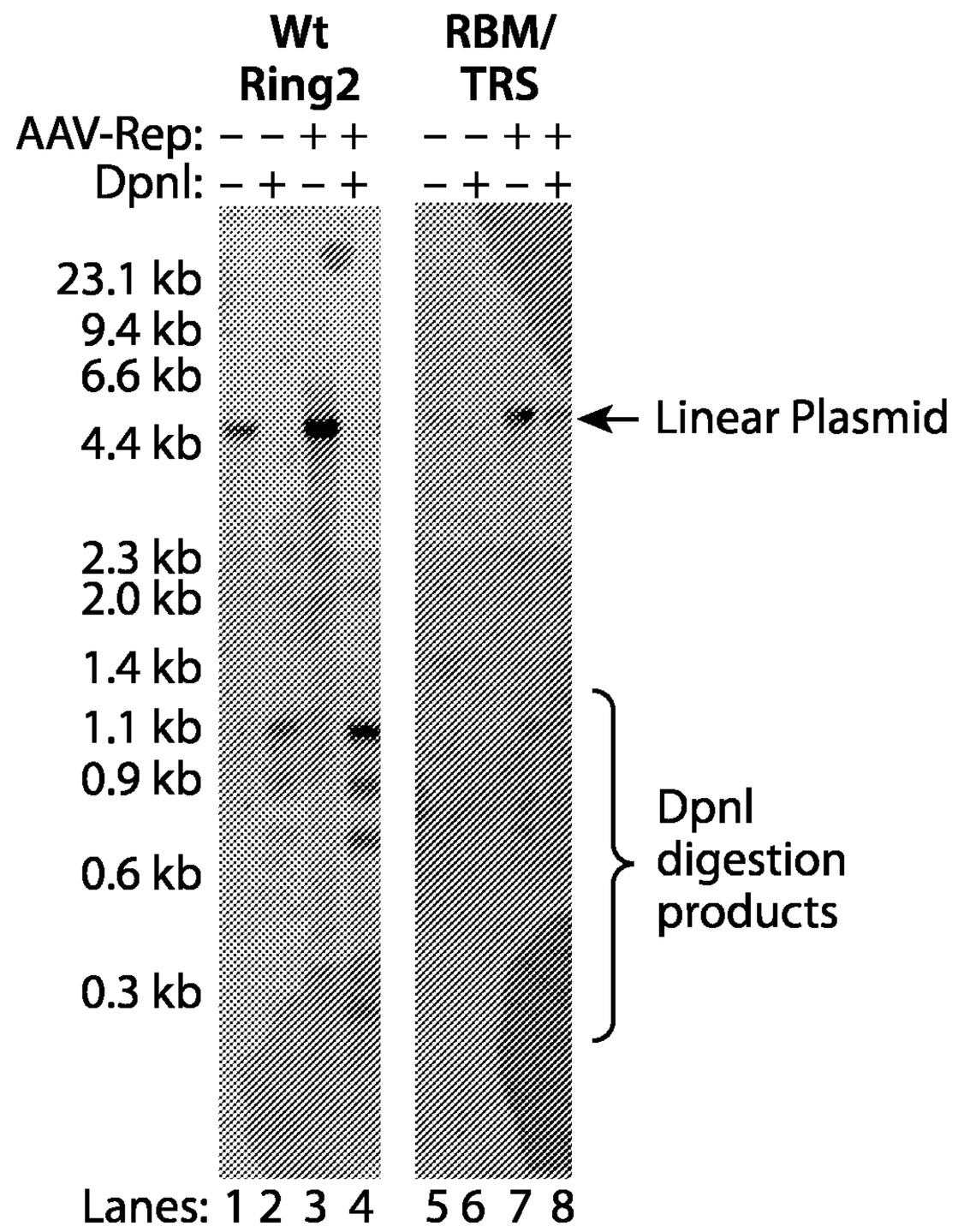

Wt
Ring2
RBM/
TRS
AAV-Rep: – – + + – – + +
DpnI: – + – + – + – +
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
1.4 kb
1.1 kb
0.9 kb
0.6 kb
0.3 kb
Linear Plasmid
DpnI
digestion
products
Lanes: 1 2 3 4 5 6 7 8

FIG. 7B

IGR-OV1
150
100
50
0
hGH (pg/ml)
Day 0
Day 2
Day 3
No capsid
AAV2
RING2
RING9

Vero
150
100
50
0
hGH (pg/ml)
No capsid
AAV2
RING2
RING9

FIG. 8A

FIG. 8B

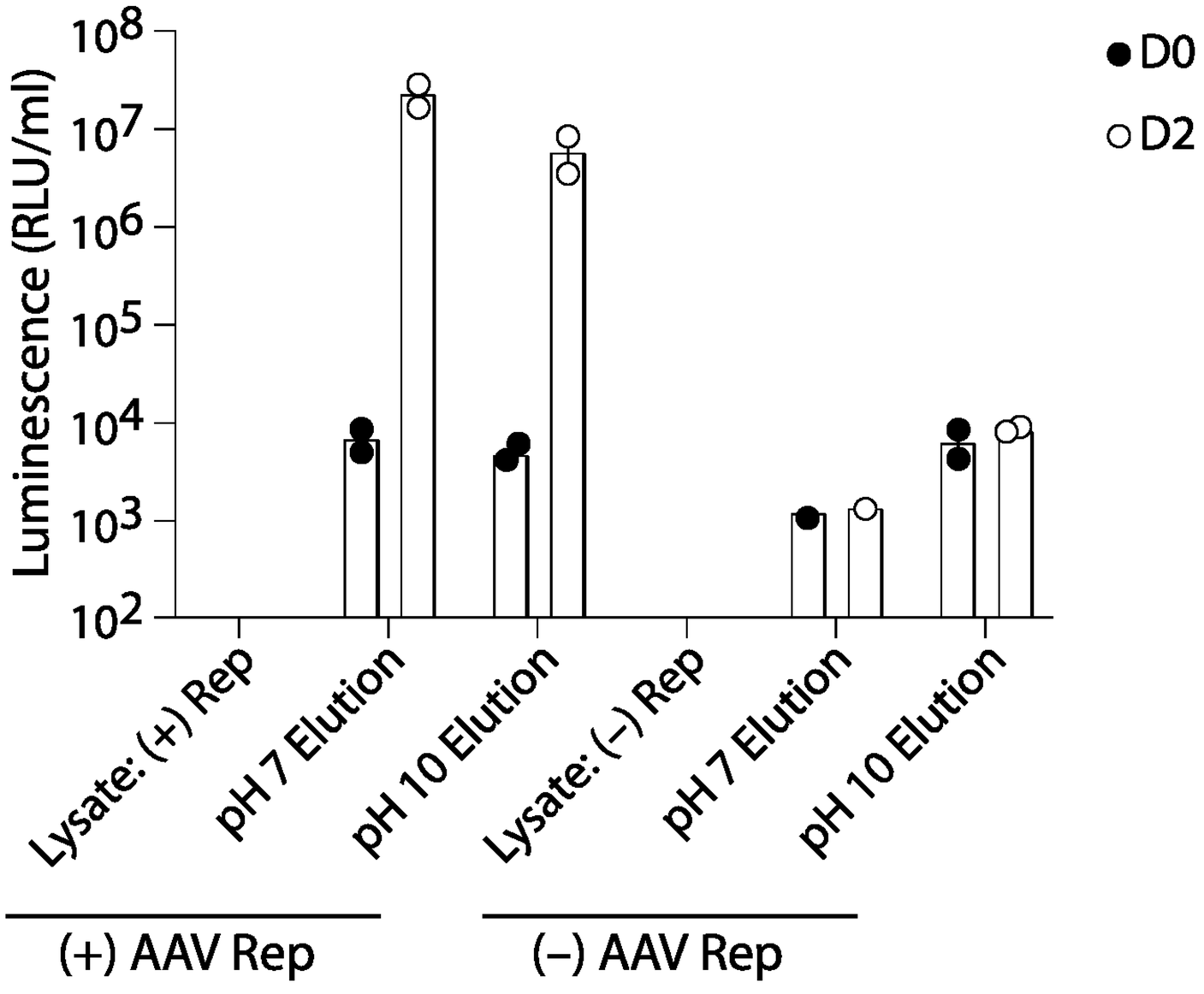

Luminescence (RLU/ml)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
D0
D2
Lysate: (+) Rep
pH 7 Elution
pH 10 Elution
Lysate: (–) Rep
pH 7 Elution
pH 10 Elution
(+) AAV Rep
(–) AAV Rep

FIG. 9

FIG. 10A

FIG. 10B

FIG. 10C

FIG. 10D

FIG. 10E

FIG. 10F

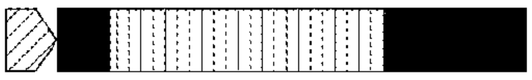


FIG. 10G

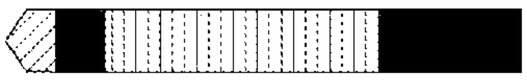


FIG. 10H

FIG. 10I

FIG. 10J

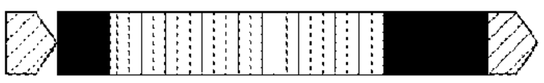


FIG. 10K

FIG. 10L

SEQUENCE LISTING

<110> FLAGSHIP PIONEERING INNOVATIONS V, INC.

<120> HYBRID AAV-ANELLOVECTORS

<130> V2057-7015WO

<140>
<141>

<150> 63/147,102
<151> 2021-02-08

<160> 1067

<170> PatentIn version 3.5

<210> 1

<400> 1
000

<210> 2

<400> 2
000

<210> 3

<400> 3
000

<210> 4

<400> 4
000

<210> 5

<400> 5
000

<210> 6

<400> 6
000

<210> 7

<400> 7
000

<210> 8

<400> 8
000

<210> 9

<400> 9
000

<210> 10

<400> 10
000

<210> 11

<400> 11
000

<210> 12

<400> 12
000

<210> 13

<400> 13
000

<210> 14

<400> 14
000

<210> 15

<400> 15
000

<210> 16
<211> 3753
<212> DNA
<213> Alphatorquevirus sp.

<400> 16
tgctacgtca ctaacccacg tgtcctctac aggccaatcg cagtctatgt cgtgcacttc 60

ctgggcatgg tctacataat tatataaatg cttgcacttc cgaatggctg agtttttgct 120

gcccgtccgc ggagaggagc cacggcaggg gatccgaacg tcctgagggc gggtgccgga 180

ggtgagttta cacaccgaag tcaaggggca attcgggctc aggactggcc gggctttggg 240

caaggctctt aaaaatgcac ttttctcgaa taagcagaaa gaaaaggaaa gtgctactgc 300

tttgcgtgcc agcagctaag aaaaaaccaa ctgctatgag cttctggaaa cctccggtac 360

acaatgtcac ggggatccaa cgcatgtggt atgagtcctt tcaccgtggc cacgcttctt 420

tttgtggttg tgggaatcct atacttcaca ttactgcact tgctgaaaca tatggccatc 480

caacaggccc gagaccttct gggccaccgg gagtagaccc caacccccac atccgtagag 540

ccaggcctgc cccggccgct ccggagccct cacaggttga ttcgagacca gccctgacat 600

ggcatgggga tggtggaagc gacggaggcg ctggtggttc cggaagcggt ggacccgtgg 660

cagacttcgc agacgatggc ctcgatcagc tcgtcgccgc cctagacgac gaagagtaag 720

gaggcgcaga cggtggagga gggggagacg aaaaacaagg acttacagac gcaggagacg 780

ctttagacgc aggggacgaa aagcaaaact tataataaaa ctgtggcaac ctgcagtaat 840

taaaagatgc agaataaagg gatacatacc actgattata agtgggaacg gtacctttgc 900

cacaaacttt accagtcaca taaatgacag aataatgaaa ggccccttcg ggggaggaca 960

cagcactatg aggttcagcc tctacatttt gtttgaggag cacctcagac acatgaactt 1020

ctggaccaga agcaacgata acctagagct aaccagatac ttgggggctt cagtaaaaat 1080

atacaggcac ccagaccaag actttatagt aatatacaac agaagaaccc ctctaggagg 1140

caacatctac acagcaccct ctctacaccc aggcaatgcc attttagcaa aacacaaaat 1200

attagtacca agtttacaga caagaccaaa gggtagaaaa gcaattagac taagaatagc 1260

acccccaca ctctttacag acaagtggta ctttcaaaag gacatagccg acctcaccct 1320

tttcaacatc atggcagttg aggctgactt gcggtttccg ttctgctcac cacaaactga 1380

caacacttgc atcagcttcc aggtccttag ttccgtttac aacaactacc tcagtattaa 1440

taccttttaat aatgacaact cagactcaaa gttaaaagaa tttttaaata aagcatttcc 1500

aacaacaggc acaaaaggaa caagtttaaa tgcactaaat acatttagaa cagaaggatg 1560

cataagtcac ccacaactaa aaaaaccaaa cccacaaata aacaaaccat tagagtcaca 1620

atactttgca cctttagatg ccctctgggg agaccccata tactataatg atctaaatga 1680

aaacaaaagt ttgaacgata tcattgagaa aatactaata aaaaacatga ttacatacca 1740

tgcaaaacta agagaatttc caaattcata ccaaggaaac aaggcctttt gccacctaac 1800

aggcatatac agcccaccat acctaaacca aggcagaata tctccagaaa tatttggact 1860

gtacacagaa ataatttaca acccttacac agacaaagga actggaaaca aagtatggat 1920

ggacccacta actaaagaga acaacatata taaagaagga cagagcaaat gcctactgac 1980

tgacatgccc ctatggactt tacttttttgg atatacagac tggtgtaaaa aggacactaa 2040

taactgggac ttaccactaa actacagact agtactaata tgcccttata cctttccaaa 2100

attgtacaat gaaaaagtaa aagactatgg gtacatcccg tactcctaca aattcggagc 2160

gggtcagatg ccagacggca gcaactacat acccttttcag tttagagcaa agtggtaccc 2220

cacagtacta caccagcaac aggtaatgga ggacataagc aggagcgggc cctttgcacc 2280

taaggtagaa aaaccaagca ctcagctggt aatgaagtac tgttttaact ttaactgggg 2340

cggtaaccct atcattgaac agattgttaa agaccccagc ttccagccca cctatgaaat 2400

acccggtacc ggtaacatcc ctagaagaat acaagtcatc gacccgcggg tcctgggacc 2460

gcactactcg ttccggtcat gggacatgcg cagacacaca tttagcagag caagtattaa 2520

gagagtgtca gaacaacaag aaacttctga ccttgtattc tcaggcccaa aaaagcctcg 2580

ggtcgacatc ccaaaacaag aaacccaaga agaaagctca cattcactcc aaagagaatc 2640

gagaccgtgg gagaccgagg aagaaagcga gacagaagcc ctctcgcaag agagccaaga 2700

ggtccccttc caacagcagt tgcagcagca gtaccaagag cagctcaagc tcagacaggg 2760

aatcaaagtc ctcttcgagc agctcataag gacccaacaa ggggtccatg taaacccatg 2820

cctacggtag gtcccaggca gtggctgttt ccagagagaa agccagcccc agctcctagc 2880

agtggagact gggccatgga gtttctcgca gcaaaaatat ttgataggcc agttagaagc 2940

aaccttaaag atacccctta ctacccatat gttaaaaacc aatacaatgt ctactttgac 3000

cttaaatttg aataaacagc agcttcaaac ttgcaaggcc gtgggagttt cactggtcgg 3060

tgtctacctc taaaggtcac taagcactcc gagcgtaagc gaggagtgcg accctccccc 3120

ctggaacaac ttcttcggag tccggcgcta cgccttcggc tgcgccggac acctcagacc 3180

ccccctccac ccgaaacgct tgcgcgtttc ggaccttcgg cgtcgggggg gtcgggagct 3240

ttattaaacg gactccgaag tgctcttgga cactgagggg gtgaacagca acgaaagtga 3300

gtggggccag acttcgccat aaggccttta tcttcttgcc atttgtcagt gtccggggtc 3360

gccataggct tcgggctcgt ttttaggcct tccggactac aaaaatcgcc attttggtga 3420

cgtcacggcc gccatcttaa gtagttgagg cggacggtgg cgtgagttca aaggtcacca 3480

tcagccacac ctactcaaaa tggtggacaa tttcttccgg gtcaaaggtt acagccgcca 3540

tgttaaaaca cgtgacgtat gacgtcacgg ccgccatttt gtgacacaag atggccgact 3600

tccttcctct ttttcaaaaa aaagcggaag tgccgccgcg gcggcggggg gcggcgcgct 3660

gcgcgcgccg cccagtaggg ggagccatgc gccccccccc gcgcatgcgc ggggcccccc 3720

cccgcggggg gctccgcccc ccggcccccc ccg 3753

<210> 17
<211> 127
<212> PRT
<213> Alphatorquevirus sp.

<400> 17
Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg
1 5 10 15

```
Met Trp Tyr Glu Ser Phe His Arg Gly His Ala Ser Phe Cys Gly Cys
            20                  25                  30

Gly Asn Pro Ile Leu His Ile Thr Ala Leu Ala Glu Thr Tyr Gly His
        35                  40                  45

Pro Thr Gly Pro Arg Pro Ser Gly Pro Pro Gly Val Asp Pro Asn Pro
    50                  55                  60

His Ile Arg Arg Ala Arg Pro Ala Pro Ala Ala Pro Glu Pro Ser Gln
65                  70                  75                  80

Val Asp Ser Arg Pro Ala Leu Thr Trp His Gly Asp Gly Gly Ser Asp
                85                  90                  95

Gly Gly Ala Gly Gly Ser Gly Ser Gly Gly Pro Val Ala Asp Phe Ala
            100                 105                 110

Asp Asp Gly Leu Asp Gln Leu Val Ala Ala Leu Asp Asp Glu Glu
        115                 120                 125
```

<210> 18
<211> 268
<212> PRT
<213> Alphatorquevirus sp.

<400> 18

```
Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg
1               5                   10                  15

Met Trp Tyr Glu Ser Phe His Arg Gly His Ala Ser Phe Cys Gly Cys
            20                  25                  30

Gly Asn Pro Ile Leu His Ile Thr Ala Leu Ala Glu Thr Tyr Gly His
        35                  40                  45

Pro Thr Gly Pro Arg Pro Ser Gly Pro Pro Gly Val Asp Pro Asn Pro
    50                  55                  60
```

```
His Ile Arg Arg Ala Arg Pro Ala Pro Ala Ala Pro Glu Pro Ser Gln
65                  70                  75                  80

Val Asp Ser Arg Pro Ala Leu Thr Trp His Gly Asp Gly Gly Ser Asp
                85                  90                  95

Gly Gly Ala Gly Gly Ser Gly Ser Gly Gly Pro Val Ala Asp Phe Ala
            100                 105                 110

Asp Asp Gly Leu Asp Gln Leu Val Ala Ala Leu Asp Asp Glu Glu Leu
        115                 120                 125

Leu Lys Thr Pro Ala Ser Ser Pro Pro Met Lys Tyr Pro Val Pro Val
    130                 135                 140

Thr Ser Leu Glu Glu Tyr Lys Ser Ser Thr Arg Gly Ser Trp Asp Arg
145                 150                 155                 160

Thr Thr Arg Ser Gly His Gly Thr Cys Ala Asp Thr His Leu Ala Glu
                165                 170                 175

Gln Val Leu Arg Glu Cys Gln Asn Asn Lys Lys Leu Leu Thr Leu Tyr
            180                 185                 190

Ser Gln Ala Gln Lys Ser Leu Gly Ser Thr Ser Gln Asn Lys Lys Pro
        195                 200                 205

Lys Lys Lys Ala His Ile His Ser Lys Glu Asn Arg Asp Arg Gly Arg
    210                 215                 220

Pro Arg Lys Lys Ala Arg Gln Lys Pro Ser Arg Lys Arg Ala Lys Arg
225                 230                 235                 240

Ser Pro Ser Asn Ser Ser Cys Ser Ser Ser Thr Lys Ser Ser Ser Ser
                245                 250                 255

Ser Asp Arg Glu Ser Lys Ser Ser Ser Ser Ser Ser Ser
```

260 265

<210> 19
<211> 276
<212> PRT
<213> Alphatorquevirus sp.

<400> 19
Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg
1 5 10 15

Met Trp Tyr Glu Ser Phe His Arg Gly His Ala Ser Phe Cys Gly Cys
20 25 30

Gly Asn Pro Ile Leu His Ile Thr Ala Leu Ala Glu Thr Tyr Gly His
35 40 45

Pro Thr Gly Pro Arg Pro Ser Gly Pro Pro Gly Val Asp Pro Asn Pro
50 55 60

His Ile Arg Arg Ala Arg Pro Ala Pro Ala Ala Pro Glu Pro Ser Gln
65 70 75 80

Val Asp Ser Arg Pro Ala Leu Thr Trp His Gly Asp Gly Gly Ser Asp
85 90 95

Gly Gly Ala Gly Gly Ser Gly Ser Gly Gly Pro Val Ala Asp Phe Ala
100 105 110

Asp Asp Gly Leu Asp Gln Leu Val Ala Ala Leu Asp Asp Glu Glu Pro
115 120 125

Lys Lys Ala Ser Gly Arg His Pro Lys Thr Arg Asn Pro Arg Arg Lys
130 135 140

Leu Thr Phe Thr Pro Lys Arg Ile Glu Thr Val Gly Asp Arg Gly Arg
145 150 155 160

Lys Arg Asp Arg Ser Pro Leu Ala Arg Glu Pro Arg Gly Pro Leu Pro

```
                165                 170                 175

Thr Ala Val Ala Ala Ala Val Pro Arg Ala Ala Gln Ala Gln Thr Gly
            180                 185                 190

Asn Gln Ser Pro Leu Arg Ala Ala His Lys Asp Pro Thr Arg Gly Pro
        195                 200                 205

Cys Lys Pro Met Pro Thr Val Gly Pro Arg Gln Trp Leu Phe Pro Glu
    210                 215                 220

Arg Lys Pro Ala Pro Ala Pro Ser Ser Gly Asp Trp Ala Met Glu Phe
225                 230                 235                 240

Leu Ala Ala Lys Ile Phe Asp Arg Pro Val Arg Ser Asn Leu Lys Asp
                245                 250                 255

Thr Pro Tyr Tyr Pro Tyr Val Lys Asn Gln Tyr Asn Val Tyr Phe Asp
            260                 265                 270

Leu Lys Phe Glu
        275


<210> 20
<211> 167
<212> PRT
<213> Alphatorquevirus sp.

<400> 20
Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg
1               5                   10                  15

Met Trp Pro Lys Lys Ala Ser Gly Arg His Pro Lys Thr Arg Asn Pro
            20                  25                  30

Arg Arg Lys Leu Thr Phe Thr Pro Lys Arg Ile Glu Thr Val Gly Asp
        35                  40                  45

Arg Gly Arg Lys Arg Asp Arg Ser Pro Leu Ala Arg Glu Pro Arg Gly
```

```
    50                  55                  60

Pro Leu Pro Thr Ala Val Ala Ala Ala Val Pro Arg Ala Ala Gln Ala
65                  70                  75                  80

Gln Thr Gly Asn Gln Ser Pro Leu Arg Ala Ala His Lys Asp Pro Thr
                85                  90                  95

Arg Gly Pro Cys Lys Pro Met Pro Thr Val Gly Pro Arg Gln Trp Leu
            100                 105                 110

Phe Pro Glu Arg Lys Pro Ala Pro Ala Pro Ser Ser Gly Asp Trp Ala
        115                 120                 125

Met Glu Phe Leu Ala Ala Lys Ile Phe Asp Arg Pro Val Arg Ser Asn
    130                 135                 140

Leu Lys Asp Thr Pro Tyr Tyr Pro Tyr Val Lys Asn Gln Tyr Asn Val
145                 150                 155                 160

Tyr Phe Asp Leu Lys Phe Glu
                165
```

<210> 21
<211> 743
<212> PRT
<213> Alphatorquevirus sp.

<400> 21

```
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1               5                   10                  15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
            20                  25                  30

Arg Arg Pro Arg Arg Arg Arg Val Arg Arg Arg Arg Arg Trp Arg Arg
        35                  40                  45

Gly Arg Arg Lys Thr Arg Thr Tyr Arg Arg Arg Arg Arg Phe Arg Arg
```

```
    50                  55                  60

Arg Gly Arg Lys Ala Lys Leu Ile Ile Lys Leu Trp Gln Pro Ala Val
65                  70                  75                  80

Ile Lys Arg Cys Arg Ile Lys Gly Tyr Ile Pro Leu Ile Ile Ser Gly
                85                  90                  95

Asn Gly Thr Phe Ala Thr Asn Phe Thr Ser His Ile Asn Asp Arg Ile
            100                 105                 110

Met Lys Gly Pro Phe Gly Gly Gly His Ser Thr Met Arg Phe Ser Leu
        115                 120                 125

Tyr Ile Leu Phe Glu Glu His Leu Arg His Met Asn Phe Trp Thr Arg
    130                 135                 140

Ser Asn Asp Asn Leu Glu Leu Thr Arg Tyr Leu Gly Ala Ser Val Lys
145                 150                 155                 160

Ile Tyr Arg His Pro Asp Gln Asp Phe Ile Val Ile Tyr Asn Arg Arg
                165                 170                 175

Thr Pro Leu Gly Gly Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly
            180                 185                 190

Asn Ala Ile Leu Ala Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr
        195                 200                 205

Arg Pro Lys Gly Arg Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr
    210                 215                 220

Leu Phe Thr Asp Lys Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr
225                 230                 235                 240

Leu Phe Asn Ile Met Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys
                245                 250                 255
```

```
Ser Pro Gln Thr Asp Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser
            260                 265                 270

Val Tyr Asn Asn Tyr Leu Ser Ile Asn Thr Phe Asn Asn Asp Asn Ser
        275                 280                 285

Asp Ser Lys Leu Lys Glu Phe Leu Asn Lys Ala Phe Pro Thr Thr Gly
    290                 295                 300

Thr Lys Gly Thr Ser Leu Asn Ala Leu Asn Thr Phe Arg Thr Glu Gly
305                 310                 315                 320

Cys Ile Ser His Pro Gln Leu Lys Lys Pro Asn Pro Gln Ile Asn Lys
                325                 330                 335

Pro Leu Glu Ser Gln Tyr Phe Ala Pro Leu Asp Ala Leu Trp Gly Asp
            340                 345                 350

Pro Ile Tyr Tyr Asn Asp Leu Asn Glu Asn Lys Ser Leu Asn Asp Ile
        355                 360                 365

Ile Glu Lys Ile Leu Ile Lys Asn Met Ile Thr Tyr His Ala Lys Leu
    370                 375                 380

Arg Glu Phe Pro Asn Ser Tyr Gln Gly Asn Lys Ala Phe Cys His Leu
385                 390                 395                 400

Thr Gly Ile Tyr Ser Pro Pro Tyr Leu Asn Gln Gly Arg Ile Ser Pro
                405                 410                 415

Glu Ile Phe Gly Leu Tyr Thr Glu Ile Ile Tyr Asn Pro Tyr Thr Asp
            420                 425                 430

Lys Gly Thr Gly Asn Lys Val Trp Met Asp Pro Leu Thr Lys Glu Asn
        435                 440                 445

Asn Ile Tyr Lys Glu Gly Gln Ser Lys Cys Leu Leu Thr Asp Met Pro
```

```
    450                 455                 460

Leu Trp Thr Leu Leu Phe Gly Tyr Thr Asp Trp Cys Lys Lys Asp Thr
465                 470                 475                 480

Asn Asn Trp Asp Leu Pro Leu Asn Tyr Arg Leu Val Leu Ile Cys Pro
                485                 490                 495

Tyr Thr Phe Pro Lys Leu Tyr Asn Glu Lys Val Lys Asp Tyr Gly Tyr
            500                 505                 510

Ile Pro Tyr Ser Tyr Lys Phe Gly Ala Gly Gln Met Pro Asp Gly Ser
        515                 520                 525

Asn Tyr Ile Pro Phe Gln Phe Arg Ala Lys Trp Tyr Pro Thr Val Leu
    530                 535                 540

His Gln Gln Gln Val Met Glu Asp Ile Ser Arg Ser Gly Pro Phe Ala
545                 550                 555                 560

Pro Lys Val Glu Lys Pro Ser Thr Gln Leu Val Met Lys Tyr Cys Phe
                565                 570                 575

Asn Phe Asn Trp Gly Gly Asn Pro Ile Ile Glu Gln Ile Val Lys Asp
            580                 585                 590

Pro Ser Phe Gln Pro Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro
        595                 600                 605

Arg Arg Ile Gln Val Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser
    610                 615                 620

Phe Arg Ser Trp Asp Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile
625                 630                 635                 640

Lys Arg Val Ser Glu Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly
                645                 650                 655
```

```
Pro Lys Lys Pro Arg Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu
            660                 665                 670

Ser Ser His Ser Leu Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu
        675                 680                 685

Glu Ser Glu Thr Glu Ala Leu Ser Gln Glu Ser Gln Glu Val Pro Phe
    690                 695                 700

Gln Gln Gln Leu Gln Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln
705                 710                 715                 720

Gly Ile Lys Val Leu Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val
                725                 730                 735

His Val Asn Pro Cys Leu Arg
            740
```

<210> 22
<211> 194
<212> PRT
<213> Alphatorquevirus sp.

<400> 22

```
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1               5                   10                  15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
            20                  25                  30

Arg Arg Pro Arg Arg Arg Arg Ile Val Lys Asp Pro Ser Phe Gln Pro
        35                  40                  45

Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro Arg Arg Ile Gln Val
    50                  55                  60

Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser Phe Arg Ser Trp Asp
65                  70                  75                  80
```

```
Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile Lys Arg Val Ser Glu
                85                  90                  95

Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly Pro Lys Lys Pro Arg
            100                 105                 110

Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu Ser Ser His Ser Leu
        115                 120                 125

Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu Glu Ser Glu Thr Glu
    130                 135                 140

Ala Leu Ser Gln Glu Ser Gln Glu Val Pro Phe Gln Gln Gln Leu Gln
145                 150                 155                 160

Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln Gly Ile Lys Val Leu
                165                 170                 175

Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val His Val Asn Pro Cys
            180                 185                 190

Leu Arg
```

<210> 23
<211> 113
<212> PRT
<213> Alphatorquevirus sp.

<400> 23

```
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1               5                   10                  15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
            20                  25                  30

Arg Arg Pro Arg Arg Arg Arg Ala Gln Lys Ser Leu Gly Ser Thr Ser
        35                  40                  45
```

Gln Asn Lys Lys Pro Lys Lys Lys Ala His Ile His Ser Lys Glu Asn
    50   55   60

Arg Asp Arg Gly Arg Pro Arg Lys Lys Ala Arg Gln Lys Pro Ser Arg
65   70   75   80

Lys Arg Ala Lys Arg Ser Pro Ser Asn Ser Ser Cys Ser Ser Ser Thr
                85   90   95

Lys Ser Ser Ser Ser Ser Asp Arg Glu Ser Lys Ser Ser Ser Ser Ser
            100   105   110

Ser

<210> 24

<400> 24
000

<210> 25

<400> 25
000

<210> 26

<400> 26
000

<210> 27

<400> 27
000

<210> 28

<400> 28
000

<210> 29

<400> 29
000

<210> 30

<400> 30
000

<210> 31

<400> 31
000

<210> 32

<400> 32
000

<210> 33

<400> 33
000

<210> 34

<400> 34
000

<210> 35

<400> 35
000

<210> 36

<400> 36
000

<210> 37

<400> 37
000

<210> 38

<400> 38
000

<210> 39

<400> 39
000

<210> 40

<400> 40
000

<210> 41

<400> 41
000

<210> 42

<400> 42
000

<210> 43

<400> 43
000

<210> 44

<400> 44
000

<210> 45

<400> 45

000

<210> 46

<400> 46
000

<210> 47

<400> 47
000

<210> 48

<400> 48
000

<210> 49

<400> 49
000

<210> 50

<400> 50
000

<210> 51

<400> 51
000

<210> 52

<400> 52
000

<210> 53

<400> 53
000

<210> 54
<211> 2979
<212> DNA
<213> Betatorquevirus sp.

<400> 54
taataaatat tcaacaggaa aaccacctaa tttaaattgc cgaccacaaa ccgtcactta 60

gttccccttt ttgcaacaac ttctgctttt ttccaactgc cggaaaacca cataatttgc 120

atggctaacc acaaactgat atgctaatta acttccacaa aacaacttcc ccttttaaaa 180

ccacacctac aaattaatta ttaaacacag tcacatcctg ggaggtacta ccacactata 240

ataccaagtg cacttccgaa tggctgagtt tatgccgcta gacggagaac gcatcagtta 300

ctgactgcgg actgaacttg ggcgggtgcc gaaggtgagt gaaaccaccg aagtcaaggg 360

gcaattcggg ctagttcagt ctagcggaac gggcaagaaa cttaaaatta ttttattttt 420

cagatgagcg actgctttaa accaacatgc tacaacaaca aaacaaagca aactcactgg 480

attaataacc tgcatttaac ccacgacctg atctgcttct gcccaacacc aactagacac 540

ttattactag ctttagcaga acaacaagaa acaattgaag tgtctaaaca agaaaaagaa 600

aaaataacaa gatgccttat tactacagaa gaagacggta caactacaga cgtcctagat 660

ggtatggacg aggttggatt agacgccctt ttcgcagaag atttcgaaga aaaagaaggg 720

taagacctac ttatactact attcctctaa agcaatggca accgccatat aaaagaacat 780

gctatataaa aggacaagac tgtttaatat actatagcaa cttaagactg ggaatgaata 840

gtacaatgta tgaaaaaagt attgtacctg tacattggcc gggaggggggt tctttttctg 900

taagcatgtt aactttagat gccttgtatg atatacataa actttgtaga aactggtgga 960

catccacaaa ccaagactta ccactagtaa gatataaagg atgcaaaata acattttatc 1020

aaagcacatt tacagactac atagtaagaa tacatacaga actaccagct aacagtaaca 1080

aactaacata cccaaacaca catccactaa tgatgatgat gtctaagtac aaacacatta 1140

tacctagtag acaaacaaga agaaaaaaga aaccatacac aaaaatattt gtaaaaccac 1200

ctccgcaatt tgaaaacaaa tggtactttg ctacagacct ctacaaaatt ccattactac 1260

aaatacactg cacagcatgc aacttacaaa acccatttgt aaaaccagac aaattatcaa 1320

acaatgttac attatggtca ctaaacacca taagcataca aaatagaaac atgtcagtgg 1380

atcaaggaca atcatggcca tttaaaatac taggaacaca aagcttttat ttttactttt 1440

acaccggagc aaacctacca ggtgacacaa cacaaatacc agtagcagac ctattaccac 1500

taacaaaccc aagaataaac agaccaggac aatcactaaa tgaggcaaaa attacagacc 1560

atattacttt cacagaatac aaaaacaaat ttacaaatta ttggggtaac ccatttaata 1620

aacacattca agaacaccta gatatgatac tatactcact aaaaagtcca gaagcaataa 1680

aaaacgaatg gacaacagaa aacatgaaat ggaaccaatt aaacaatgca ggaacaatgg 1740

cattaacacc atttaacgag ccaatattca cacaaataca atataaccca gatagagaca 1800

caggagaaga cactcaatta tacctactct ctaacgctac aggaacagga tgggacccac 1860

caggaattcc agaattaata ctagaaggat ttccactatg gttaatatat tggggatttg 1920

cagactttca aaaaaaccta aaaaaagtaa caaacataga cacaaattac atgttagtag 1980

caaaaacaaa atttacacaa aaacctggca cattctactt agtaatacta aatgacacct 2040

ttgtagaagg caatagccca tatgaaaaac aacctttacc tgaagacaac attaaatggt 2100

acccacaagt acaataccaa ttagaagcac aaaacaaact actacaaact gggccattta 2160

caccaaacat acaaggacaa ctatcagaca atatatcaat gttttataaa ttttacttta 2220

aatggggagg aagcccacca aaagcaatta atgttgaaaa tcctgcccac cagattcaat 2280

atcccatacc ccgtaacgag catgaaacaa cttcgttaca gagtccaggg gaagccccag 2340

aatccatctt atactccttc gactatagac acgggaacta cacaacaaca gctttgtcac 2400

gaattagcca agactgggca cttaaagaca ctgtttctaa aattacagag ccagatcgac 2460

agcaactgct caaacaagcc ctcgaatgcc tgcaaatctc ggaagaaacg caggagaaaa 2520

aagaaaaaga agtacagcag ctcatcagca acctcagaca gcagcagcag ctgtacagag 2580

agcgaataat atcattatta aaggaccaat aacttttaac tgtgtaaaaa aggtgaaatt 2640

gtttgatgat aaaccaaaaa accgtagatt tacacctgag gaatttgaaa ctgagttaca 2700

aatagcaaaa tggttaaaga gacccccaag atcctttgta aatgatcctc ccttttaccc 2760

atggttacca cctgaacctg ttgtaaactt taagcttaat tttactgaat aaaggccagc 2820

attaattcac ttaaggagtc tgtttattta agttaaacct taataaacgg tcaccgcctc 2880

cctaatacgc aggcgcagaa agggggctcc gcccccttta acccccaggg ggctccgccc 2940

cctgaaaccc ccaaggggggc tacgcccccct tacaccccc 2979

<210> 55
<211> 99
<212> PRT
<213> Betatorquevirus sp.

<400> 55
Met Ser Asp Cys Phe Lys Pro Thr Cys Tyr Asn Asn Lys Thr Lys Gln
1 5 10 15

Thr His Trp Ile Asn Asn Leu His Leu Thr His Asp Leu Ile Cys Phe
20 25 30

Cys Pro Thr Pro Thr Arg His Leu Leu Leu Ala Leu Ala Glu Gln Gln
35 40 45

Glu Thr Ile Glu Val Ser Lys Gln Glu Lys Glu Lys Ile Thr Arg Cys
50 55 60

Leu Ile Thr Thr Glu Glu Asp Gly Thr Thr Thr Asp Val Leu Asp Gly
65 70 75 80

Met Asp Glu Val Gly Leu Asp Ala Leu Phe Ala Glu Asp Phe Glu Glu
85 90 95

Lys Glu Gly

<210> 56
<211> 203
<212> PRT
<213> Betatorquevirus sp.

<400> 56
Met Ser Asp Cys Phe Lys Pro Thr Cys Tyr Asn Asn Lys Thr Lys Gln
1 5 10 15

```
Thr His Trp Ile Asn Asn Leu His Leu Thr His Asp Leu Ile Cys Phe
            20                  25                  30

Cys Pro Thr Pro Thr Arg His Leu Leu Leu Ala Leu Ala Glu Gln Gln
        35                  40                  45

Glu Thr Ile Glu Val Ser Lys Gln Glu Lys Glu Lys Ile Thr Arg Cys
    50                  55                  60

Leu Ile Thr Thr Glu Glu Asp Gly Thr Thr Thr Asp Val Leu Asp Gly
65                  70                  75                  80

Met Asp Glu Val Gly Leu Asp Ala Leu Phe Ala Glu Asp Phe Glu Glu
                85                  90                  95

Lys Glu Gly Phe Asn Ile Pro Tyr Pro Val Thr Ser Met Lys Gln Leu
            100                 105                 110

Arg Tyr Arg Val Gln Gly Lys Pro Gln Asn Pro Ser Tyr Thr Pro Ser
        115                 120                 125

Thr Ile Asp Thr Gly Thr Thr Gln Gln Gln Leu Cys His Glu Leu Ala
    130                 135                 140

Lys Thr Gly His Leu Lys Thr Leu Phe Leu Lys Leu Gln Ser Gln Ile
145                 150                 155                 160

Asp Ser Asn Cys Ser Asn Lys Pro Ser Asn Ala Cys Lys Ser Arg Lys
                165                 170                 175

Lys Arg Arg Arg Lys Lys Lys Lys Lys Tyr Ser Ser Ser Ser Ala Thr
            180                 185                 190

Ser Asp Ser Ser Ser Ser Cys Thr Glu Ser Glu
        195                 200
```

<210> 57
<211> 219
<212> PRT
<213> Betatorquevirus sp.

<400> 57
Met Ser Asp Cys Phe Lys Pro Thr Cys Tyr Asn Asn Lys Thr Lys Gln
1   5   10   15

Thr His Trp Ile Asn Asn Leu His Leu Thr His Asp Leu Ile Cys Phe
            20   25   30

Cys Pro Thr Pro Thr Arg His Leu Leu Leu Ala Leu Ala Glu Gln Gln
        35   40   45

Glu Thr Ile Glu Val Ser Lys Gln Glu Lys Glu Lys Ile Thr Arg Cys
    50   55   60

Leu Ile Thr Thr Glu Glu Asp Gly Thr Thr Thr Asp Val Leu Asp Gly
65   70   75   80

Met Asp Glu Val Gly Leu Asp Ala Leu Phe Ala Glu Asp Phe Glu Glu
                85   90   95

Lys Glu Gly Ala Arg Ser Thr Ala Thr Ala Gln Thr Ser Pro Arg Met
            100   105   110

Pro Ala Asn Leu Gly Arg Asn Ala Gly Glu Lys Arg Lys Arg Ser Thr
        115   120   125

Ala Ala His Gln Gln Pro Gln Thr Ala Ala Ala Ala Val Gln Arg Ala
    130   135   140

Asn Asn Ile Ile Ile Lys Gly Pro Ile Thr Phe Asn Cys Val Lys Lys
145   150   155   160

Val Lys Leu Phe Asp Asp Lys Pro Lys Asn Arg Arg Phe Thr Pro Glu
                165   170   175

```
Glu Phe Glu Thr Glu Leu Gln Ile Ala Lys Trp Leu Lys Arg Pro Pro
            180                 185                 190

Arg Ser Phe Val Asn Asp Pro Pro Phe Tyr Pro Trp Leu Pro Pro Glu
        195                 200                 205

Pro Val Val Asn Phe Lys Leu Asn Phe Thr Glu
    210                 215
```

<210> 58
<211> 666
<212> PRT
<213> Betatorquevirus sp.

<400> 58

```
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1               5                   10                  15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
            20                  25                  30

Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys Gln
        35                  40                  45

Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp Cys
    50                  55                  60

Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met Tyr
65                  70                  75                  80

Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe Ser
                85                  90                  95

Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu Cys
            100                 105                 110

Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr
        115                 120                 125
```

```
Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile
    130                 135                 140

Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr
145                 150                 155                 160

Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His Ile
                165                 170                 175

Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile
            180                 185                 190

Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr
        195                 200                 205

Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys Asn
    210                 215                 220

Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val Thr
225                 230                 235                 240

Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser Val
                245                 250                 255

Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe
            260                 265                 270

Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln
        275                 280                 285

Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg
    290                 295                 300

Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr Phe
305                 310                 315                 320

Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn
                325                 330                 335
```

```
Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser
            340                 345                 350

Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp Asn
        355                 360                 365

Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu Pro
    370                 375                 380

Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp
385                 390                 395                 400

Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro
                405                 410                 415

Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile
            420                 425                 430

Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr Asn
        435                 440                 445

Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln Lys
    450                 455                 460

Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu Gly
465                 470                 475                 480

Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp
                485                 490                 495

Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln
            500                 505                 510

Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile
        515                 520                 525
```

```
Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys
    530                 535                 540

Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro
545                 550                 555                 560

Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro
                565                 570                 575

Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr
            580                 585                 590

Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr Val
        595                 600                 605

Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu
    610                 615                 620

Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu
625                 630                 635                 640

Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg
                645                 650                 655

Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
            660                 665


<210> 59
<211> 148
<212> PRT
<213> Betatorquevirus sp.

<400> 59
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1               5                   10                  15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
            20                  25                  30
```

```
Arg Lys Arg Arg Ile Gln Tyr Pro Ile Pro Arg Asn Glu His Glu Thr
        35                  40                  45

Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro Glu Ser Ile Leu Tyr Ser
    50                  55                  60

Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr Thr Ala Leu Ser Arg Ile
65                  70                  75                  80

Ser Gln Asp Trp Ala Leu Lys Asp Thr Val Ser Lys Ile Thr Glu Pro
                85                  90                  95

Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu Glu Cys Leu Gln Ile Ser
            100                 105                 110

Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu Val Gln Gln Leu Ile Ser
        115                 120                 125

Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg Glu Arg Ile Ile Ser Leu
    130                 135                 140

Leu Lys Asp Gln
145
```

<210> 60
<211> 82
<212> PRT
<213> Betatorquevirus sp.

<400> 60

```
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1               5                   10                  15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
            20                  25                  30

Arg Lys Arg Arg Ser Gln Ile Asp Ser Asn Cys Ser Asn Lys Pro Ser
        35                  40                  45
```

Asn Ala Cys Lys Ser Arg Lys Lys Arg Arg Arg Lys Lys Lys Lys Lys
50 55 60

Tyr Ser Ser Ser Ser Ala Thr Ser Asp Ser Ser Ser Ser Cys Thr Glu
65 70 75 80

Ser Glu

<210> 61

<400> 61
000

<210> 62

<400> 62
000

<210> 63

<400> 63
000

<210> 64

<400> 64
000

<210> 65

<400> 65
000

<210> 66

<400> 66
000

<210> 67

<400> 67
000

<210> 68

<400> 68
000

<210> 69

<400> 69
000

<210> 70

<400> 70
000

<210> 71

<400> 71
000

<210> 72

<400> 72
000

<210> 73

<400> 73
000

<210> 74

<400> 74
000

<210> 75

<400> 75
000

<210> 76

<400> 76
000

<210> 77

<400> 77
000

<210> 78

<400> 78
000

<210> 79

<400> 79
000

<210> 80

<400> 80
000

<210> 81

<400> 81
000

<210> 82

<400> 82
000

<210> 83

<400> 83
000

<210> 84

<400> 84
000

<210> 85

<400> 85
000

<210> 86

<400> 86
000

<210> 87

<400> 87
000

<210> 88

<400> 88
000

<210> 89

<400> 89
000

<210> 90

<400> 90
000

<210> 91

<400> 91
000

<210> 92

<400> 92
000

<210> 93

<400> 93
000

<210> 94

<400> 94
000

<210> 95

<400> 95
000

<210> 96

<400> 96
000

<210> 97

<400> 97
000

<210> 98

<400> 98
000

<210> 99

<400> 99
000

<210> 100

<400> 100
000

<210> 101

<400> 101
000

<210> 102

<400> 102
000

<210> 103

<400> 103
000

<210> 104

<400> 104
000

<210> 105
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 105
cgggtgccgk aggtgagttt acacaccgma gtcaaggggc aattcgggct crggactggc 60

cgggcyhtgg g 71

<210> 106
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 106
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctwtgg g 71

<210> 107
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 107
cgggtgccgt aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 108
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 108
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggccctgg g 71

<210> 109
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 109
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctttgg g 71

<210> 110
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 110
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 111
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 111
cgggtgccgg aggtgagttt acacaccgaa gtcaaggggc aattcgggct caggactggc 60

cgggctttgg g 71

<210> 112
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 112
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggcyhtgg g 71

<210> 113
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 113
cgggtgccgt aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 114
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 114
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggcccggg 70

<210> 115
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 115
cgggtgccgg aggtgagttt acacaccgaa gtcaaggggc aattcgggct caggactggc 60

cgggctttgg g 71

<210> 116
<211> 69
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 116
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggaggccg 60

ggccatggg 69

<210> 117
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 117
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggccccgg g 71

<210> 118
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 118
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 119
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 119
cgggtgccga aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 120
<211> 117
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<220>
<221> misc_feature
<222> (10)..(10)
<223> May or may not be present

<220>
<221> misc_feature
<222> (12)..(12)
<223> May or may not be present

<220>
<221> misc_feature
<222> (30)..(32)
<223> May or may not be present

<220>
<221> misc_feature
<222> (34)..(34)
<223> May or may not be present

<220>
<221> misc_feature
<222> (43)..(46)
<223> May or may not be present

<220>
<221> misc_feature
<222> (52)..(54)
<223> May or may not be present

<220>
<221> misc_feature
<222> (70)..(71)
<223> May or may not be present

<220>
<221> misc_feature
<222> (89)..(90)
<223> May or may not be present

<220>
<221> misc_feature
<222> (103)..(103)
<223> May or may not be present

<400> 120
cggcggsggs gcsscgcgct dcgcgcgcsg cccrsyrggg grdssmmwgc skcsccccccc 60

cscgcgcatg cgcrcgggkc ccccccccyv sgggggggctc cgcccccccg gcccccc 117

<210> 121
<211> 169
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<220>
<221> modified_base
<222> (20)..(20)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (22)..(22)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (40)..(42)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (53)..(56)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (62)..(62)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (64)..(64)

<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (97)..(98)
<223> a, c, t, g, unknown or other

<400> 121
gccgccgcgg cggcggsggn gnsgcgcgct dcgcgcgcsn nncrccrggg ggnnnncwgc 60

sncnccccccc cccgcgcatg cgcgggkccc ccccccnncg gggggctccg ccccccggcc 120

ccccccgtg ctaaacccac cgcgcatgcg cgaccacgcc cccgccgcc 169

<210> 122
<211> 79
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (20)..(20)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (22)..(22)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (40)..(42)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (53)..(56)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (62)..(62)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (64)..(64)
<223> a, c, t, g, unknown or other

<400> 122
gccgccgcgg cggcggsggn gnsgcgcgct dcgcgcgcsn nncrccrggg ggnnnncwgc 60

sncnccccccc cccgcgcat 79

<210> 123
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (18)..(19)
<223> a, c, t, g, unknown or other

<400> 123
gcgcgggkcc ccccccnnc ggggggctcc g 31

<210> 124
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 124
ccccccggcc ccccccgtg ctaaacccac cgcgcatgcg cgaccacgcc cccgccgcc 59

<210> 125
<211> 156
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 125
gcggcggggg ggcggccgcg ttcgcgcgcc gcccaccagg gggtgctgcg cgccccccc 60

cgcgcatgcg cggggccccc ccccgggggg gctccgcccc cccggccccc ccccgtgcta 120

aacccaccgc gcatgcgcga ccacgccccc gccgcc 156

<210> 126
<211> 7
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 126
gcggcgg 7

<210> 127
<211> 7
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 127
gggggcg 7

<210> 128
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 128
gccgcg 6

<210> 129

<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 129
ttcgcgcgcc gcccaccagg gggtg 25

<210> 130
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 130
ctgcg 5

<210> 131
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 131
cgccccccc cgcgcat 17

<210> 132
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 132
gcgcggggcc ccccccc 17

<210> 133
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 133
gggggggctc cgccccccg gccccccccc gtgctaaacc caccgcgcat gcgcgaccac 60

gcccccgccg cc 72

<210> 134
<211> 115
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 134
cggcggcggc ggcgcgcgcg ctgcgcgcgc gcgccggggg ggcgccagcg ccccccccccc 60

cgcgcatgca cgggtccccc cccccacggg gggctccgcc ccccggcccc ccccc 115

<210> 135
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 135
cggcggcggc ggcg 14

<210> 136
<211> 17
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 136
cgcgcgctgc gcgcgcg 17

<210> 137
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 137
cgccgggggg gcgccagcg 19

<210> 138
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 138
ccccccccc cgcgcat 17

<210> 139
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 139
gcacgggtcc ccccccccac ggggggctcc g 31

<210> 140
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 140
ccccccggcc ccccccc 17

<210> 141
<211> 121
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 141
ccgtcggcgg gggggccgcg cgctgcgcgc gcggcccccg ggggaggcac agcctccccc 60

ccccgcgcgc atgcgcgcgg gtcccccccc ctccgggggg ctccgccccc cggccccccc 120

c 121

<210> 142
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 142
ccgtcggcgg gggggccgcg cgctgcgcgc gcggccc 37

<210> 143
<211> 84
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 143
ccgggggagg cacagcctcc ccccccccgcg cgcatgcgcg cgggtccccc cccctccggg 60

gggctccgcc ccccggcccc cccc 84

<210> 144
<211> 104
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 144
cggcggcggc gcgcgcgcta cgcgcgcgcg ccgggggggct gccgcccccc ccccgcgcat 60

gcgcggggcc cccccccgcg gggggctccg ccccccggcc cccc 104

<210> 145
<211> 11
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 145
cggcggcggc g 11

<210> 146
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 146
cgcgcgctac gcgcgcg 17

<210> 147
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 147
cgccgggggg 10

<210> 148
<211> 7
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 148
ctgccgc 7

<210> 149
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 149
ccccccccg cgcat 15

<210> 150
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 150
gcgcggggcc ccccccc 17

<210> 151
<211> 13
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 151
gcgggggct ccg 13

<210> 152
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 152
ccccccggcc cccc 14

<210> 153
<211> 122
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 153
gccgccgcgg cggcgggggg cggcgcgctg cgcgcgccgc ccagtagggg gagccatgcg 60

cccccccccg cgcatgcgcg gggcccccc ccgcgggggg ctccgccccc cggcccccc 120

cg 122

<210> 154
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 154

gccgccgcgg cggcggggg 19

<210> 155
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 155
gcggcgcgct gcgcgcgccg cccagtaggg ggagccatgc g 41

<210> 156
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 156
ccccccccg cgcat 15

<210> 157
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 157
gcgcggggcc ccccccc 17

<210> 158
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 158
gcgggggct ccg 13

<210> 159
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 159
ccccccggcc ccccccg 17

<210> 160
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 160
cgcgctgcgc gcgccgccca gtagggggag ccatgc 36

<210> 161
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 161
ccgccatctt aagtagttga ggcggacggt ggcgtgagtt caaaggtcac catcagccac 60

acctactcaa aatggtgg 78

<210> 162
<211> 172
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 162
cttaagtagt tgaggcggac ggtggcgtga gttcaaaggt caccatcagc cacacctact 60

caaaatggtg gacaatttct tccgggtcaa aggttacagc cgccatgtta aaacacgtga 120

cgtatgacgt cacggccgcc attttgtgac acaagatggc cgacttcctt cc 172

<210> 163
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 163
cgcgctgcgc gcgccgccca gtaggggggag ccatgc 36

<210> 164
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 164
gcgctdcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 165
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 165
gcgcttcgcg cgccgcccac taggggggcgt tgcgcg 36

<210> 166
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 166
gcgctgcgcg cgccgcccag taggggcgc aatgcg 36

<210> 167
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 167
gcgctgcgcg cgcggccccc gggggaggca ttgcct 36

<210> 168
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 168
gcgctgcgcg cgcgcgccgg gggggcgcca gcgccc 36

<210> 169
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 169
gcgcttcgcg cgcgcgccgg ggggctccgc cccccc 36

<210> 170
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 170
gcgcttcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 171
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 171
gcgctacgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 172
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 172
gcgctacgcg cgcgcgccgg ggggctctgc cccccc 36

<210> 173

<400> 173
000

<210> 174

<400> 174
000

<210> 175

<400> 175
000

<210> 176

<400> 176
000

<210> 177

<400> 177
000

<210> 178

<400> 178
000

<210> 179

<400> 179
000

<210> 180

<400> 180
000

<210> 181

<400> 181
000

<210> 182

<400> 182

000

<210> 183

<400> 183
000

<210> 184

<400> 184
000

<210> 185
<211> 743
<212> PRT
<213> Alphatorquevirus sp.

<400> 185
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1   5   10   15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
            20   25   30

Arg Arg Pro Arg Arg Arg Arg Val Arg Arg Arg Arg Arg Trp Arg Arg
        35   40   45

Gly Arg Arg Lys Thr Arg Thr Tyr Arg Arg Arg Arg Arg Phe Arg Arg
    50   55   60

Arg Gly Arg Lys Ala Lys Leu Ile Ile Lys Leu Trp Gln Pro Ala Val
65   70   75   80

Ile Lys Arg Cys Arg Ile Lys Gly Tyr Ile Pro Leu Ile Ile Ser Gly
                85   90   95

Asn Gly Thr Phe Ala Thr Asn Phe Thr Ser His Ile Asn Asp Arg Ile
            100   105   110

Met Lys Gly Pro Phe Gly Gly Gly His Ser Thr Met Arg Phe Ser Leu

```
        115                 120                 125

Tyr Ile Leu Phe Glu Glu His Leu Arg His Met Asn Phe Trp Thr Arg
    130                 135                 140

Ser Asn Asp Asn Leu Glu Leu Thr Arg Tyr Leu Gly Ala Ser Val Lys
145                 150                 155                 160

Ile Tyr Arg His Pro Asp Gln Asp Phe Ile Val Ile Tyr Asn Arg Arg
                165                 170                 175

Thr Pro Leu Gly Gly Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly
            180                 185                 190

Asn Ala Ile Leu Ala Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr
        195                 200                 205

Arg Pro Lys Gly Arg Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr
    210                 215                 220

Leu Phe Thr Asp Lys Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr
225                 230                 235                 240

Leu Phe Asn Ile Met Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys
                245                 250                 255

Ser Pro Gln Thr Asp Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser
            260                 265                 270

Val Tyr Asn Asn Tyr Leu Ser Ile Asn Thr Phe Asn Asn Asp Asn Ser
        275                 280                 285

Asp Ser Lys Leu Lys Glu Phe Leu Asn Lys Ala Phe Pro Thr Thr Gly
    290                 295                 300

Thr Lys Gly Thr Ser Leu Asn Ala Leu Asn Thr Phe Arg Thr Glu Gly
305                 310                 315                 320
```

```
Cys Ile Ser His Pro Gln Leu Lys Lys Pro Asn Pro Gln Ile Asn Lys
               325                 330                 335

Pro Leu Glu Ser Gln Tyr Phe Ala Pro Leu Asp Ala Leu Trp Gly Asp
            340                 345                 350

Pro Ile Tyr Tyr Asn Asp Leu Asn Glu Asn Lys Ser Leu Asn Asp Ile
        355                 360                 365

Ile Glu Lys Ile Leu Ile Lys Asn Met Ile Thr Tyr His Ala Lys Leu
    370                 375                 380

Arg Glu Phe Pro Asn Ser Tyr Gln Gly Asn Lys Ala Phe Cys His Leu
385                 390                 395                 400

Thr Gly Ile Tyr Ser Pro Pro Tyr Leu Asn Gln Gly Arg Ile Ser Pro
               405                 410                 415

Glu Ile Phe Gly Leu Tyr Thr Glu Ile Ile Tyr Asn Pro Tyr Thr Asp
            420                 425                 430

Lys Gly Thr Gly Asn Lys Val Trp Met Asp Pro Leu Thr Lys Glu Asn
        435                 440                 445

Asn Ile Tyr Lys Glu Gly Gln Ser Lys Cys Leu Leu Thr Asp Met Pro
    450                 455                 460

Leu Trp Thr Leu Leu Phe Gly Tyr Thr Asp Trp Cys Lys Lys Asp Thr
465                 470                 475                 480

Asn Asn Trp Asp Leu Pro Leu Asn Tyr Arg Leu Val Leu Ile Cys Pro
               485                 490                 495

Tyr Thr Phe Pro Lys Leu Tyr Asn Glu Lys Val Lys Asp Tyr Gly Tyr
            500                 505                 510

Ile Pro Tyr Ser Tyr Lys Phe Gly Ala Gly Gln Met Pro Asp Gly Ser
```

```
        515                 520                 525

Asn Tyr Ile Pro Phe Gln Phe Arg Ala Lys Trp Tyr Pro Thr Val Leu
    530                 535                 540

His Gln Gln Gln Val Met Glu Asp Ile Ser Arg Ser Gly Pro Phe Ala
545                 550                 555                 560

Pro Lys Val Glu Lys Pro Ser Thr Gln Leu Val Met Lys Tyr Cys Phe
                565                 570                 575

Asn Phe Asn Trp Gly Gly Asn Pro Ile Ile Glu Gln Ile Val Lys Asp
            580                 585                 590

Pro Ser Phe Gln Pro Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro
        595                 600                 605

Arg Arg Ile Gln Val Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser
    610                 615                 620

Phe Arg Ser Trp Asp Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile
625                 630                 635                 640

Lys Arg Val Ser Glu Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly
                645                 650                 655

Pro Lys Lys Pro Arg Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu
            660                 665                 670

Ser Ser His Ser Leu Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu
        675                 680                 685

Glu Ser Glu Thr Glu Ala Leu Ser Gln Glu Ser Gln Glu Val Pro Phe
    690                 695                 700

Gln Gln Gln Leu Gln Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln
705                 710                 715                 720
```

Gly Ile Lys Val Leu Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val
725 730 735

His Val Asn Pro Cys Leu Arg
740

<210> 186
<211> 68
<212> PRT
<213> Alphatorquevirus sp.

<400> 186
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1 5 10 15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
20 25 30

Arg Arg Pro Arg Arg Arg Arg Val Arg Arg Arg Arg Arg Trp Arg Arg
35 40 45

Gly Arg Arg Lys Thr Arg Thr Tyr Arg Arg Arg Arg Arg Phe Arg Arg
50 55 60

Arg Gly Arg Lys
65

<210> 187
<211> 212
<212> PRT
<213> Alphatorquevirus sp.

<400> 187
Ala Lys Leu Ile Ile Lys Leu Trp Gln Pro Ala Val Ile Lys Arg Cys
1 5 10 15

Arg Ile Lys Gly Tyr Ile Pro Leu Ile Ile Ser Gly Asn Gly Thr Phe
20 25 30

Ala Thr Asn Phe Thr Ser His Ile Asn Asp Arg Ile Met Lys Gly Pro

```
        35                  40                  45

Phe Gly Gly Gly His Ser Thr Met Arg Phe Ser Leu Tyr Ile Leu Phe
    50                  55                  60

Glu Glu His Leu Arg His Met Asn Phe Trp Thr Arg Ser Asn Asp Asn
65                  70                  75                  80

Leu Glu Leu Thr Arg Tyr Leu Gly Ala Ser Val Lys Ile Tyr Arg His
                85                  90                  95

Pro Asp Gln Asp Phe Ile Val Ile Tyr Asn Arg Arg Thr Pro Leu Gly
            100                 105                 110

Gly Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly Asn Ala Ile Leu
        115                 120                 125

Ala Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr Arg Pro Lys Gly
    130                 135                 140

Arg Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr Leu Phe Thr Asp
145                 150                 155                 160

Lys Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr Leu Phe Asn Ile
                165                 170                 175

Met Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys Ser Pro Gln Thr
            180                 185                 190

Asp Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser Val Tyr Asn Asn
        195                 200                 205

Tyr Leu Ser Ile
    210

<210> 188
<211> 133
<212> PRT
```

<213> Alphatorquevirus sp.

<400> 188
Asn Thr Phe Asn Asn Asp Asn Ser Asp Ser Lys Leu Lys Glu Phe Leu
1   5   10   15

Asn Lys Ala Phe Pro Thr Thr Gly Thr Lys Gly Thr Ser Leu Asn Ala
            20   25   30

Leu Asn Thr Phe Arg Thr Glu Gly Cys Ile Ser His Pro Gln Leu Lys
        35   40   45

Lys Pro Asn Pro Gln Ile Asn Lys Pro Leu Glu Ser Gln Tyr Phe Ala
    50   55   60

Pro Leu Asp Ala Leu Trp Gly Asp Pro Ile Tyr Tyr Asn Asp Leu Asn
65   70   75   80

Glu Asn Lys Ser Leu Asn Asp Ile Ile Glu Lys Ile Leu Ile Lys Asn
                85   90   95

Met Ile Thr Tyr His Ala Lys Leu Arg Glu Phe Pro Asn Ser Tyr Gln
            100   105   110

Gly Asn Lys Ala Phe Cys His Leu Thr Gly Ile Tyr Ser Pro Pro Tyr
        115   120   125

Leu Asn Gln Gly Arg
    130

<210> 189
<211> 166
<212> PRT
<213> Alphatorquevirus sp.

<400> 189
Ile Ser Pro Glu Ile Phe Gly Leu Tyr Thr Glu Ile Ile Tyr Asn Pro
1   5   10   15

Tyr Thr Asp Lys Gly Thr Gly Asn Lys Val Trp Met Asp Pro Leu Thr

```
            20                  25                  30

Lys Glu Asn Asn Ile Tyr Lys Glu Gly Gln Ser Lys Cys Leu Leu Thr
        35                  40                  45

Asp Met Pro Leu Trp Thr Leu Leu Phe Gly Tyr Thr Asp Trp Cys Lys
    50                  55                  60

Lys Asp Thr Asn Asn Trp Asp Leu Pro Leu Asn Tyr Arg Leu Val Leu
65                  70                  75                  80

Ile Cys Pro Tyr Thr Phe Pro Lys Leu Tyr Asn Glu Lys Val Lys Asp
                85                  90                  95

Tyr Gly Tyr Ile Pro Tyr Ser Tyr Lys Phe Gly Ala Gly Gln Met Pro
            100                 105                 110

Asp Gly Ser Asn Tyr Ile Pro Phe Gln Phe Arg Ala Lys Trp Tyr Pro
        115                 120                 125

Thr Val Leu His Gln Gln Gln Val Met Glu Asp Ile Ser Arg Ser Gly
    130                 135                 140

Pro Phe Ala Pro Lys Val Glu Lys Pro Ser Thr Gln Leu Val Met Lys
145                 150                 155                 160

Tyr Cys Phe Asn Phe Asn
                165


<210> 190
<211> 164
<212> PRT
<213> Alphatorquevirus sp.

<400> 190
Trp Gly Gly Asn Pro Ile Ile Glu Gln Ile Val Lys Asp Pro Ser Phe
1               5                   10                  15

Gln Pro Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro Arg Arg Ile
```

```
            20                  25                  30

Gln Val Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser Phe Arg Ser
        35                  40                  45

Trp Asp Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile Lys Arg Val
    50                  55                  60

Ser Glu Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly Pro Lys Lys
65                  70                  75                  80

Pro Arg Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu Ser Ser His
                85                  90                  95

Ser Leu Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu Glu Ser Glu
            100                 105                 110

Thr Glu Ala Leu Ser Gln Glu Ser Gln Glu Val Pro Phe Gln Gln Gln
        115                 120                 125

Leu Gln Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln Gly Ile Lys
    130                 135                 140

Val Leu Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val His Val Asn
145                 150                 155                 160

Pro Cys Leu Arg
```

<210> 191

<400> 191
000

<210> 192

<400> 192
000

<210> 193

<400> 193
000

<210> 194

<400> 194
000

<210> 195

<400> 195
000

<210> 196

<400> 196
000

<210> 197

<400> 197
000

<210> 198

<400> 198
000

<210> 199

<400> 199
000

<210> 200

<400> 200
000

<210> 201

<400> 201
000

<210> 202

<400> 202
000

<210> 203

<400> 203
000

<210> 204

<400> 204
000

<210> 205

<400> 205
000

<210> 206

<400> 206
000

<210> 207

<400> 207
000

<210> 208

<400> 208
000

<210> 209

<400> 209

000

<210> 210

<400> 210
000

<210> 211

<400> 211
000

<210> 212

<400> 212
000

<210> 213

<400> 213
000

<210> 214

<400> 214
000

<210> 215
<211> 666
<212> PRT
<213> Betatorquevirus sp.

<400> 215
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1 5 10 15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
            20 25 30

Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys Gln
        35 40 45

```
Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp Cys
    50                  55                  60

Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met Tyr
65                  70                  75                  80

Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe Ser
                85                  90                  95

Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu Cys
            100                 105                 110

Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr
        115                 120                 125

Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile
    130                 135                 140

Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr
145                 150                 155                 160

Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His Ile
                165                 170                 175

Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile
            180                 185                 190

Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr
        195                 200                 205

Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys Asn
    210                 215                 220

Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val Thr
225                 230                 235                 240

Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser Val
```

```
                245                 250                 255

Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe
            260                 265                 270

Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln
        275                 280                 285

Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg
    290                 295                 300

Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr Phe
305                 310                 315                 320

Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn
                325                 330                 335

Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser
            340                 345                 350

Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp Asn
        355                 360                 365

Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu Pro
    370                 375                 380

Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp
385                 390                 395                 400

Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro
                405                 410                 415

Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile
            420                 425                 430

Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr Asn
        435                 440                 445
```

```
Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln Lys
    450                 455                 460

Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu Gly
465                 470                 475                 480

Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp
                485                 490                 495

Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln
            500                 505                 510

Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile
        515                 520                 525

Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys
    530                 535                 540

Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro
545                 550                 555                 560

Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro
                565                 570                 575

Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr
            580                 585                 590

Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr Val
        595                 600                 605

Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu
    610                 615                 620

Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu
625                 630                 635                 640

Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg
```

645 650 655

Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
660 665

<210> 216
<211> 38
<212> PRT
<213> Betatorquevirus sp.

<400> 216
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1 5 10 15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
20 25 30

Arg Lys Arg Arg Val Arg
35

<210> 217
<211> 208
<212> PRT
<213> Betatorquevirus sp.

<400> 217
Pro Thr Tyr Thr Thr Ile Pro Leu Lys Gln Trp Gln Pro Pro Tyr Lys
1 5 10 15

Arg Thr Cys Tyr Ile Lys Gly Gln Asp Cys Leu Ile Tyr Tyr Ser Asn
20 25 30

Leu Arg Leu Gly Met Asn Ser Thr Met Tyr Glu Lys Ser Ile Val Pro
35 40 45

Val His Trp Pro Gly Gly Gly Ser Phe Ser Val Ser Met Leu Thr Leu
50 55 60

Asp Ala Leu Tyr Asp Ile His Lys Leu Cys Arg Asn Trp Trp Thr Ser
65 70 75 80

```
Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr Lys Gly Cys Lys Ile Thr
                85                  90                  95

Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile Val Arg Ile His Thr Glu
            100                 105                 110

Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr Pro Asn Thr His Pro Leu
        115                 120                 125

Met Met Met Met Ser Lys Tyr Lys His Ile Ile Pro Ser Arg Gln Thr
    130                 135                 140

Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile Phe Val Lys Pro Pro Pro
145                 150                 155                 160

Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr Asp Leu Tyr Lys Ile Pro
                165                 170                 175

Leu Leu Gln Ile His Cys Thr Ala Cys Asn Leu Gln Asn Pro Phe Val
            180                 185                 190

Lys Pro Asp Lys Leu Ser Asn Asn Val Thr Leu Trp Ser Leu Asn Thr
        195                 200                 205
```

<210> 218
<211> 128
<212> PRT
<213> Betatorquevirus sp.

<400> 218

```
Ile Ser Ile Gln Asn Arg Asn Met Ser Val Asp Gln Gly Gln Ser Trp
1               5                   10                  15

Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe Tyr Phe Tyr Phe Tyr Thr
            20                  25                  30

Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln Ile Pro Val Ala Asp Leu
        35                  40                  45
```

Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg Pro Gly Gln Ser Leu Asn
50 55 60

Glu Ala Lys Ile Thr Asp His Ile Thr Phe Thr Glu Tyr Lys Asn Lys
65 70 75 80

Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn Lys His Ile Gln Glu His
85 90 95

Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser Pro Glu Ala Ile Lys Asn
100 105 110

Glu Trp Thr Thr Glu Asn Met Lys Trp Asn Gln Leu Asn Asn Ala Gly
115 120 125

<210> 219
<211> 163
<212> PRT
<213> Betatorquevirus sp.

<400> 219
Thr Met Ala Leu Thr Pro Phe Asn Glu Pro Ile Phe Thr Gln Ile Gln
1 5 10 15

Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp Thr Gln Leu Tyr Leu Leu
20 25 30

Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro Pro Gly Ile Pro Glu Leu
35 40 45

Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile Tyr Trp Gly Phe Ala Asp
50 55 60

Phe Gln Lys Asn Leu Lys Lys Val Thr Asn Ile Asp Thr Asn Tyr Met
65 70 75 80

Leu Val Ala Lys Thr Lys Phe Thr Gln Lys Pro Gly Thr Phe Tyr Leu
85 90 95

```
Val Ile Leu Asn Asp Thr Phe Val Glu Gly Asn Ser Pro Tyr Glu Lys
            100                 105                 110

Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp Tyr Pro Gln Val Gln Tyr
        115                 120                 125

Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln Thr Gly Pro Phe Thr Pro
    130                 135                 140

Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile Ser Met Phe Tyr Lys Phe
145                 150                 155                 160

Tyr Phe Lys
```

<210> 220
<211> 129
<212> PRT
<213> Betatorquevirus sp.

<400> 220

```
Trp Gly Gly Ser Pro Pro Lys Ala Ile Asn Val Glu Asn Pro Ala His
1               5                   10                  15

Gln Ile Gln Tyr Pro Ile Pro Arg Asn Glu His Glu Thr Thr Ser Leu
            20                  25                  30

Gln Ser Pro Gly Glu Ala Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr
        35                  40                  45

Arg His Gly Asn Tyr Thr Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp
    50                  55                  60

Trp Ala Leu Lys Asp Thr Val Ser Lys Ile Thr Glu Pro Asp Arg Gln
65                  70                  75                  80

Gln Leu Leu Lys Gln Ala Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr
                85                  90                  95
```

Gln Glu Lys Lys Glu Lys Glu Val Gln Gln Leu Ile Ser Asn Leu Arg
100 105 110

Gln Gln Gln Gln Leu Tyr Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp
115 120 125

Gln

<210> 221

<400> 221
000

<210> 222

<400> 222
000

<210> 223

<400> 223
000

<210> 224

<400> 224
000

<210> 225

<400> 225
000

<210> 226

<400> 226
000

<210> 227

<211> 220
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<220>
<221> MOD_RES
<222> (29)..(31)
<223> Any amino acid

<220>
<221> SITE
<222> (29)..(31)
<223> This region may encompass 0-3 residues

<220>
<221> MOD_RES
<222> (100)..(100)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (125)..(129)
<223> Any amino acid

<220>
<221> SITE
<222> (125)..(129)
<223> This region may encompass 1-5 residues

<220>
<221> MOD_RES
<222> (181)..(181)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (211)..(211)
<223> Any amino acid

<400> 227
Leu Val Leu Thr Gln Trp Gln Pro Asn Thr Val Arg Arg Cys Tyr Ile
1 5 10 15

Arg Gly Tyr Leu Pro Leu Ile Ile Cys Gly Glu Asn Xaa Xaa Xaa Thr

```
            20                  25                  30

Thr Ser Arg Asn Tyr Ala Thr His Ser Asp Asp Thr Ile Gln Lys Gly
        35                  40                  45

Pro Phe Gly Gly Gly Met Ser Thr Thr Thr Phe Ser Leu Arg Val Leu
    50                  55                  60

Tyr Asp Glu Tyr Gln Arg Phe Met Asn Arg Trp Thr Tyr Ser Asn Glu
65                  70                  75                  80

Asp Leu Asp Leu Ala Arg Tyr Leu Gly Cys Lys Phe Thr Phe Tyr Arg
                85                  90                  95

His Pro Asp Xaa Asp Phe Ile Val Gln Tyr Asn Thr Asn Pro Pro Phe
            100                 105                 110

Lys Asp Thr Lys Leu Thr Ala Pro Ser Ile His Pro Xaa Xaa Xaa Xaa
        115                 120                 125

Xaa Gly Met Leu Met Leu Ser Lys Arg Lys Ile Leu Ile Pro Ser Leu
    130                 135                 140

Lys Thr Arg Pro Lys Gly Lys His Tyr Val Lys Val Arg Ile Gly Pro
145                 150                 155                 160

Pro Lys Leu Phe Glu Asp Lys Trp Tyr Thr Gln Ser Asp Leu Cys Asp
                165                 170                 175

Val Pro Leu Val Xaa Leu Tyr Ala Thr Ala Ala Asp Leu Gln His Pro
            180                 185                 190

Phe Gly Ser Pro Gln Thr Asp Asn Pro Cys Val Thr Phe Gln Val Leu
        195                 200                 205

Gly Ser Xaa Tyr Asn Lys His Leu Ser Ile Ser Pro
    210                 215                 220
```

<210> 228
<211> 172
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<220>
<221> MOD_RES
<222> (38)..(38)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (44)..(46)
<223> Any amino acid

<220>
<221> SITE
<222> (44)..(46)
<223> This region may encompass 0-3 residues

<220>
<221> MOD_RES
<222> (77)..(77)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (79)..(79)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (98)..(101)
<223> Any amino acid

<220>
<221> SITE
<222> (98)..(101)
<223> This region may encompass 0-4 residues

<400> 228
Ser Asn Phe Glu Phe Pro Gly Ala Tyr Thr Asp Ile Thr Tyr Asn Pro
1 5 10 15

```
Leu Thr Asp Lys Gly Val Gly Asn Met Val Trp Ile Gln Tyr Leu Thr
            20                  25                  30

Lys Pro Asp Thr Ile Xaa Asp Lys Thr Gln Ser Xaa Xaa Xaa Lys Cys
        35                  40                  45

Leu Ile Glu Asp Leu Pro Leu Trp Ala Ala Leu Tyr Gly Tyr Val Asp
    50                  55                  60

Phe Cys Glu Lys Glu Thr Gly Asp Ser Ala Ile Ile Xaa Asn Xaa Gly
65                  70                  75                  80

Arg Val Leu Ile Arg Cys Pro Tyr Thr Lys Pro Pro Leu Tyr Asp Lys
                85                  90                  95

Thr Xaa Xaa Xaa Xaa Asn Lys Gly Phe Val Pro Tyr Ser Thr Asn Phe
            100                 105                 110

Gly Asn Gly Lys Met Pro Gly Gly Ser Gly Tyr Val Pro Ile Tyr Trp
        115                 120                 125

Arg Ala Arg Trp Tyr Pro Thr Leu Phe His Gln Lys Glu Val Leu Glu
    130                 135                 140

Asp Ile Val Gln Ser Gly Pro Phe Ala Tyr Lys Asp Glu Lys Pro Ser
145                 150                 155                 160

Thr Gln Leu Val Met Lys Tyr Cys Phe Asn Phe Asn
                165                 170
```

<210> 229
<211> 258
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<220>
<221> MOD_RES
<222> (20)..(22)
<223> Any amino acid

<220>
<221> SITE
<222> (20)..(22)
<223> This region may encompass 0-3 residues

<220>
<221> MOD_RES
<222> (25)..(25)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (78)..(78)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (89)..(89)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (91)..(91)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (95)..(98)
<223> Any amino acid

<220>
<221> SITE
<222> (95)..(98)
<223> This region may encompass 1-4 residues

<220>
<221> MOD_RES
<222> (107)..(120)
<223> Any amino acid

<220>
<221> SITE
<222> (107)..(120)
<223> This region may encompass 2-14 residues

<220>
<221> MOD_RES
<222> (129)..(129)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (139)..(168)
<223> Any amino acid

<220>
<221> SITE
<222> (139)..(168)
<223> This region may encompass 0-30 residues

<220>
<221> MOD_RES
<222> (201)..(204)
<223> Any amino acid

<220>
<221> SITE
<222> (201)..(204)
<223> This region may encompass 0-4 residues

<220>
<221> MOD_RES
<222> (219)..(258)
<223> Any amino acid

<220>
<221> SITE
<222> (219)..(258)
<223> This region may encompass 0-40 residues

<400> 229
Trp Gly Gly Asn Pro Ile Ser Gln Gln Val Val Arg Asn Pro Cys Lys
1   5   10   15

Asp Ser Gly Xaa Xaa Xaa Ser Gly Xaa Gly Arg Gln Pro Arg Ser Val
            20   25   30

Gln Val Val Asp Pro Lys Tyr Met Gly Pro Glu Tyr Thr Phe His Ser
        35   40   45

Trp Asp Trp Arg Arg Gly Leu Phe Gly Glu Lys Ala Ile Lys Arg Met

```
    50                  55                  60

Ser Glu Gln Pro Thr Asp Asp Glu Ile Phe Thr Gly Gly Xaa Pro Lys
65                  70                  75                  80

Arg Pro Arg Arg Asp Pro Pro Thr Xaa Gln Xaa Pro Glu Glu Xaa Xaa
                85                  90                  95

Xaa Xaa Gln Lys Glu Ser Ser Ser Phe Arg Xaa Xaa Xaa Xaa Xaa Xaa
            100                 105                 110

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Trp Glu Ser Ser Ser Gln Glu
        115                 120                 125

Xaa Glu Ser Glu Ser Gln Glu Glu Glu Glu Xaa Xaa Xaa Xaa Xaa Xaa
    130                 135                 140

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
145                 150                 155                 160

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Glu Gln Thr Val Gln Gln Gln Leu
                165                 170                 175

Arg Gln Gln Leu Arg Glu Gln Arg Arg Leu Arg Val Gln Leu Gln Leu
            180                 185                 190

Leu Phe Gln Gln Leu Leu Lys Thr Xaa Xaa Xaa Xaa Gln Ala Gly Leu
        195                 200                 205

His Ile Asn Pro Leu Leu Leu Ser Gln Ala Xaa Xaa Xaa Xaa Xaa Xaa
    210                 215                 220

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
225                 230                 235                 240

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
                245                 250                 255
```

Xaa Xaa

<210> 230
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<220>
<221> MOD_RES
<222> (136)..(136)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (138)..(141)
<223> Any amino acid

<220>
<221> SITE
<222> (138)..(141)
<223> This region may encompass 1-4 residues

<220>
<221> MOD_RES
<222> (179)..(179)
<223> Any amino acid

<400> 230
Leu Lys Gln Trp Gln Pro Ser Thr Ile Arg Lys Cys Lys Ile Lys Gly
1 5 10 15

Tyr Leu Pro Leu Phe Gln Cys Gly Lys Gly Arg Ile Ser Asn Asn Tyr
20 25 30

Thr Gln Tyr Lys Glu Ser Ile Val Pro His His Glu Pro Gly Gly Gly
35 40 45

Gly Trp Ser Ile Gln Gln Phe Thr Leu Gly Ala Leu Tyr Glu Glu His
50 55 60

```
Leu Lys Leu Arg Asn Trp Trp Thr Lys Ser Asn Asp Gly Leu Pro Leu
65                  70                  75                  80

Val Arg Tyr Leu Gly Cys Thr Ile Lys Leu Tyr Arg Ser Glu Asp Thr
                85                  90                  95

Asp Tyr Ile Val Thr Tyr Gln Arg Cys Tyr Pro Met Thr Ala Thr Lys
            100                 105                 110

Leu Thr Tyr Leu Ser Thr Gln Pro Ser Arg Met Leu Met Asn Lys His
        115                 120                 125

Lys Ile Ile Val Pro Ser Lys Xaa Thr Xaa Xaa Xaa Xaa Asn Lys Lys
    130                 135                 140

Lys Lys Pro Tyr Lys Lys Ile Phe Ile Lys Pro Pro Ser Gln Met Gln
145                 150                 155                 160

Asn Lys Trp Tyr Phe Gln Gln Asp Ile Ala Asn Thr Pro Leu Leu Gln
                165                 170                 175

Leu Thr Xaa Thr Ala Cys Ser Leu Asp Arg Met Tyr Leu Ser Ser Asp
            180                 185                 190

Ser Ile Ser Asn Asn Ile Thr Phe Thr Ser Leu Asn Thr Asn Phe Phe
        195                 200                 205

Gln Asn Pro Asn Phe Gln
    210
```

<210> 231
<211> 187
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<220>
<221> MOD_RES
<222> (1)..(10)
<223> Any amino acid

<220>
<221> SITE
<222> (1)..(10)
<223> This region may encompass 4-10 residues

<220>
<221> MOD_RES
<222> (38)..(45)
<223> Any amino acid

<220>
<221> SITE
<222> (38)..(45)
<223> This region may encompass 1-8 residues

<220>
<221> MOD_RES
<222> (94)..(94)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (100)..(102)
<223> Any amino acid

<220>
<221> SITE
<222> (100)..(102)
<223> This region may encompass 1-3 residues

<220>
<221> MOD_RES
<222> (112)..(112)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (114)..(115)
<223> Any amino acid

<220>
<221> SITE
<222> (114)..(115)

<223> This region may encompass 0-2 residues

<220>
<221> MOD_RES
<222> (124)..(139)
<223> Any amino acid

<220>
<221> SITE
<222> (124)..(139)
<223> This region may encompass 3-16 residues

<220>
<221> MOD_RES
<222> (154)..(154)
<223> Any amino acid

<400> 231
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Pro Leu Tyr Phe Glu
1   5   10   15

Cys Arg Tyr Asn Pro Phe Lys Asp Lys Gly Thr Gly Asn Lys Val Tyr
            20   25   30

Leu Val Ser Asn Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Gly Trp
        35   40   45

Asp Pro Pro Thr Asp Pro Asp Leu Ile Ile Glu Gly Phe Pro Leu Trp
    50   55   60

Leu Leu Leu Trp Gly Trp Leu Asp Trp Gln Lys Lys Leu Gly Lys Ile
65   70   75   80

Gln Asn Ile Asp Thr Asp Tyr Ile Leu Val Ile Gln Ser Xaa Tyr Tyr
                85   90   95

Ile Pro Pro Xaa Xaa Xaa Lys Leu Pro Tyr Tyr Val Pro Leu Asp Xaa
            100   105   110

Asp Xaa Xaa Phe Leu His Gly Arg Ser Pro Tyr Xaa Xaa Xaa Xaa Xaa
        115   120   125

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Ser Asp Lys Gln
    130   135   140

His Trp His Pro Lys Val Arg Phe Gln Xaa Glu Thr Ile Asn Asn Ile
145   150   155   160

Ala Leu Thr Gly Pro Gly Thr Pro Lys Leu Pro Asn Gln Lys Ser Ile
                165   170   175

Gln Ala His Met Lys Tyr Lys Phe Tyr Phe Lys
            180   185

<210> 232
<211> 163
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<220>
<221> MOD_RES
<222> (34)..(34)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (65)..(65)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (77)..(78)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (86)..(87)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (96)..(96)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (102)..(106)
<223> Any amino acid

<220>
<221> SITE
<222> (102)..(106)
<223> This region may encompass 0-5 residues

<220>
<221> MOD_RES
<222> (125)..(125)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (135)..(135)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (138)..(163)
<223> Any amino acid

<220>
<221> SITE
<222> (138)..(163)
<223> This region may encompass 0-26 residues

<400> 232
Trp Gly Gly Cys Pro Ala Pro Met Glu Thr Ile Thr Asp Pro Cys Lys
1   5   10   15

Gln Pro Lys Tyr Pro Ile Pro Asn Asn Leu Leu Gln Thr Thr Ser Leu
            20   25   30

Gln Xaa Pro Thr Thr Pro Ile Glu Thr Tyr Leu Tyr Lys Phe Asp Glu
        35   40   45

Arg Arg Gly Leu Leu Thr Lys Lys Ala Ala Lys Arg Ile Lys Lys Asp
    50   55   60

Xaa Thr Thr Glu Thr Thr Leu Phe Thr Asp Thr Gly Xaa Xaa Thr Ser
65   70   75   80

Thr Thr Leu Pro Thr Xaa Xaa Gln Thr Glu Thr Thr Gln Glu Glu Xaa
85 90 95

Thr Ser Glu Glu Glu Xaa Xaa Xaa Xaa Xaa Glu Thr Leu Leu Gln Gln
100 105 110

Leu Gln Gln Leu Arg Arg Lys Gln Lys Gln Leu Arg Xaa Arg Ile Leu
115 120 125

Gln Leu Leu Gln Leu Leu Xaa Leu Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa
130 135 140

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
145 150 155 160

Xaa Xaa Xaa

<210> 233
<211> 203
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<220>
<221> MOD_RES
<222> (79)..(79)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (104)..(104)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (116)..(116)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (120)..(121)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (125)..(125)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (170)..(170)
<223> Any amino acid

<400> 233
Thr Ile Pro Leu Lys Gln Trp Gln Pro Glu Ser Ile Arg Lys Cys Lys
1   5   10   15

Ile Lys Gly Tyr Gly Thr Leu Val Leu Gly Ala Glu Gly Arg Gln Phe
            20   25   30

Tyr Cys Tyr Thr Asn Glu Lys Asp Glu Tyr Thr Pro Pro Lys Ala Pro
        35   40   45

Gly Gly Gly Gly Phe Gly Val Glu Leu Phe Ser Leu Glu Tyr Leu Tyr
    50   55   60

Glu Gln Trp Lys Ala Arg Asn Asn Ile Trp Thr Lys Ser Asn Xaa Tyr
65   70   75   80

Lys Asp Leu Cys Arg Tyr Thr Gly Cys Lys Ile Thr Phe Tyr Arg His
                85   90   95

Pro Thr Thr Asp Phe Ile Val Xaa Tyr Ser Arg Gln Pro Pro Phe Glu
            100   105   110

Ile Asp Lys Xaa Thr Tyr Met Xaa Xaa His Pro Gln Xaa Leu Leu Leu
        115   120   125

Arg Lys His Lys Lys Ile Ile Leu Ser Lys Ala Thr Asn Pro Lys Gly

```
    130                 135                 140

Lys Leu Lys Lys Lys Ile Lys Ile Lys Pro Pro Lys Gln Met Leu Asn
145                 150                 155                 160

Lys Trp Phe Phe Gln Lys Gln Phe Ala Xaa Tyr Gly Leu Val Gln Leu
                165                 170                 175

Gln Ala Ala Ala Cys Asx Leu Arg Tyr Pro Arg Leu Gly Cys Cys Asn
            180                 185                 190

Glu Asn Arg Leu Ile Thr Leu Tyr Tyr Leu Asn
        195                 200
```

<210> 234
<211> 162
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<220>
<221> MOD_RES
<222> (12)..(12)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (20)..(20)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (23)..(23)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (30)..(30)
<223> Any amino acid

<220>

<221> MOD_RES
<222> (58)..(58)
<223> I or L

<220>
<221> MOD_RES
<222> (84)..(84)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (90)..(90)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (95)..(95)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (105)..(105)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (111)..(111)
<223> I or L

<220>
<221> MOD_RES
<222> (113)..(113)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (154)..(154)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (156)..(156)
<223> Any amino acid

<400> 234
Leu Pro Ile Val Val Ala Arg Tyr Asn Pro Ala Xaa Asp Thr Gly Lys
1   5   10   15

Gly Asn Lys Xaa Trp Leu Xaa Ser Thr Leu Asn Gly Ser Xaa Trp Ala

```
            20                  25                  30

Pro Pro Thr Thr Asp Lys Asp Leu Ile Ile Glu Gly Leu Pro Leu Trp
        35                  40                  45

Leu Ala Leu Tyr Gly Tyr Trp Ser Tyr Xaa Lys Lys Val Lys Lys Asp
    50                  55                  60

Lys Gly Ile Leu Gln Ser His Met Phe Val Val Lys Ser Pro Ala Ile
65                  70                  75                  80

Gln Pro Leu Xaa Thr Ala Thr Thr Gln Xaa Thr Phe Tyr Pro Xaa Ile
                85                  90                  95

Asp Asn Ser Phe Ile Gln Gly Lys Xaa Pro Tyr Asp Glu Pro Xaa Thr
            100                 105                 110

Xaa Asn Gln Lys Lys Leu Trp Tyr Pro Thr Leu Glu His Gln Gln Glu
        115                 120                 125

Thr Ile Asn Ala Ile Val Glu Ser Gly Pro Tyr Val Pro Lys Leu Asp
    130                 135                 140

Asn Gln Lys Asn Ser Thr Trp Glu Leu Xaa Tyr Xaa Tyr Thr Phe Tyr
145                 150                 155                 160

Phe Lys
```

<210> 235
<211> 177
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<220>

<221> MOD_RES
<222> (16)..(16)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (26)..(26)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (33)..(33)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (73)..(73)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (81)..(82)
<223> Any amino acid

<220>
<221> SITE
<222> (81)..(82)
<223> This region may encompass 0-2 residues

<220>
<221> MOD_RES
<222> (90)..(90)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (94)..(94)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (119)..(124)
<223> Any amino acid

<220>
<221> SITE
<222> (119)..(124)
<223> This region may encompass 1-6 residues

<220>

<221> MOD_RES
<222> (168)..(177)
<223> Any amino acid

<220>
<221> SITE
<222> (168)..(177)
<223> This region may encompass 1-10 residues

<400> 235
Trp Gly Gly Pro Gln Ile Pro Asp Gln Pro Val Glu Asp Pro Lys Xaa
1   5   10   15

Gln Gly Thr Tyr Pro Val Pro Asp Thr Xaa Gln Gln Thr Ile Gln Ile
            20   25   30

Xaa Asn Pro Leu Lys Gln Lys Pro Glu Thr Met Phe His Asp Trp Asp
        35   40   45

Tyr Arg Arg Gly Ile Ile Thr Ser Thr Ala Leu Lys Arg Met Gln Glu
    50   55   60

Asn Leu Glu Thr Asp Ser Ser Phe Xaa Ser Asp Ser Glu Glu Thr Pro
65   70   75   80

Xaa Xaa Lys Lys Lys Lys Arg Leu Thr Xaa Glu Leu Pro Xaa Pro Gln
                85   90   95

Glu Glu Thr Glu Glu Ile Gln Ser Cys Leu Leu Ser Leu Cys Glu Glu
            100   105   110

Ser Thr Cys Gln Glu Glu Xaa Xaa Xaa Xaa Xaa Xaa Glu Asn Leu Gln
        115   120   125

Gln Leu Ile His Gln Gln Gln Gln Gln Gln Gln Gln Leu Lys His Asn
    130   135   140

Ile Leu Lys Leu Leu Ser Asp Leu Lys Glx Lys Gln Arg Leu Leu Gln
145   150   155   160

Leu Gln Thr Gly Ile Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
165 170 175

Xaa

<210> 236

<400> 236
000

<210> 237

<400> 237
000

<210> 238

<400> 238
000

<210> 239

<400> 239
000

<210> 240

<400> 240
000

<210> 241

<400> 241
000

<210> 242

<400> 242
000

<210> 243

<400> 243
000

<210> 244

<400> 244
000

<210> 245

<400> 245
000

<210> 246

<400> 246
000

<210> 247

<400> 247
000

<210> 248

<400> 248
000

<210> 249

<400> 249
000

<210> 250

<400> 250
000

<210> 251

<400> 251
000

<210> 252

<400> 252
000

<210> 253

<400> 253
000

<210> 254

<400> 254
000

<210> 255

<400> 255
000

<210> 256

<400> 256
000

<210> 257

<400> 257
000

<210> 258

<400> 258
000

<210> 259

<400> 259
000

<210> 260

<400> 260
000

<210> 261

<400> 261
000

<210> 262

<400> 262
000

<210> 263

<400> 263
000

<210> 264

<400> 264
000

<210> 265

<400> 265
000

<210> 266

<400> 266
000

<210> 267

<400> 267
000

<210> 268

<400> 268
000

<210> 269

<400> 269
000

<210> 270

<400> 270
000

<210> 271

<400> 271
000

<210> 272

<400> 272
000

<210> 273

<400> 273
000

<210> 274

<400> 274
000

<210> 275

<400> 275
000

<210> 276

<400> 276
000

<210> 277

<400> 277
000

<210> 278

<400> 278
000

<210> 279

<400> 279
000

<210> 280

<400> 280
000

<210> 281

<400> 281
000

<210> 282

<400> 282
000

<210> 283

<400> 283
000

<210> 284

<400> 284
000

<210> 285

<400> 285
000

<210> 286

<400> 286
000

<210> 287

<400> 287
000

<210> 288

<400> 288
000

<210> 289

<400> 289
000

<210> 290

<400> 290
000

<210> 291

<400> 291
000

<210> 292

<400> 292
000

<210> 293

<400> 293
000

<210> 294

<400> 294
000

<210> 295

<400> 295
000

<210> 296

<400> 296
000

<210> 297

<400> 297
000

<210> 298

<400> 298
000

<210> 299

<400> 299
000

<210> 300
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 300
gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 301
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 301
gcguacguca caagucacgu ggaggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 302
<211> 77
<212> RNA
<213> Alphatorquevirus sp.

<400> 302
gccauuuuaa guagcugacg ucaaggauug acgugaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggug 77

<210> 303
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 303
gcacacguca uaagucacgu gguggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugauuuguca cgugugua 78

<210> 304
<211> 66
<212> RNA
<213> Alphatorquevirus sp.

<400> 304
cuuccggguc auaggucaca ccuacgucac aagucacgug gggaggguug gcguauagcc 60

cggaag 66

<210> 305
<211> 68
<212> RNA

<213> Alphatorquevirus sp.

<400> 305
gccgggggggc ugccgccccc cccggggaaa gggggggggcc ccccccgggg gggggguuugc 60

cccccggc 68

<210> 306
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 306
auacgucauc agucacgugg gggaaggcgu gccuaaaccc ggaagcaucc ucguccacgu 60

gacugugacg uguguggc 78

<210> 307
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 307
cauuuuaagu aaggcggaag cagcucggcg uacacaaaau ggcggcggag cacuuccggc 60

uugcccaaaa ugg 73

<210> 308
<211> 71
<212> RNA
<213> Alphatorquevirus sp.

<400> 308
gucacaaguc acgugggggag gguuggcguu uaacccggaa gccaauccuc uuacguggcc 60

ugucacguga c 71

<210> 309
<211> 70
<212> RNA
<213> Alphatorquevirus sp.

<400> 309
cgaccgcguc ccgaaggcgg guacccgagg ugaguuuaca caccgagguu aagggccaau 60

ucgggcuugg 70

<210> 310
<211> 59
<212> RNA
<213> Alphatorquevirus sp.

<400> 310
cgcgguaucg uagccgacgc ggaccccguu uucggggccc ccgcggggcu cucggcgcg 59

<210> 311
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 311
cgccauuuug ugauacgcgc gucccccuccc ggcuuccgua caacgucagg cggggcgugg 60

ccguaucaga aaauggcg 78

<210> 312
<211> 77
<212> RNA
<213> Alphatorquevirus sp.

<400> 312
gcuacgucau aagucacgug acugggcagg uacuaaaccc ggaaguaucc ucggucacgu 60

ggccugucac guaguug 77

<210> 313
<211> 80
<212> RNA
<213> Alphatorquevirus sp.

<400> 313
ggcusugacg ucaaagucac gugggraggg uggcguuaaa cccggaaguc auccucguca 60

cgugaccuga cgucacagcc 80

<210> 314
<211> 66
<212> RNA
<213> Alphatorquevirus sp.

<400> 314

gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

gaaggu 66

<210> 315
<211> 80
<212> RNA
<213> Alphatorquevirus sp.

<400> 315
gguugugacg ucaaagucac gugggggaggg cggcguuaaa cccggaaguc auccucguca 60

cgugaccuga cgucacggcc 80

<210> 316
<211> 67
<212> RNA
<213> Alphatorquevirus sp.

<400> 316
gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

guaggug 67

<210> 317
<211> 67
<212> RNA
<213> Alphatorquevirus sp.

<400> 317
gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

guaggug 67

<210> 318
<211> 80
<212> RNA
<213> Alphatorquevirus sp.

<400> 318
ggcugugacg ucaaagucac gugggggaggg cggcguuaaa cccggaaguc auccucguca 60

cgugaccuga cgucacggcc 80

<210> 319

<211> 79
<212> RNA
<213> Alphatorquevirus sp.

<400> 319
agaccacgug guaagucacg ugggggcagc ugcuguaaac ccggaaguag cugacccgcg 60

ugacugguca cgugaccug 79

<210> 320
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 320
cgccauuuua uaauacgcgc gucccccucc ggcuuccgua cuacgucagg cggggcgugg 60

ccguauuaga aaauggug 78

<210> 321
<211> 72
<212> RNA
<213> Alphatorquevirus sp.

<400> 321
uaaguaaggc ggaaccaggc ugucacccug ugucaaaggu caagggacag ccuuccggcu 60

ugcacaaaau gg 72

<210> 322
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 322
ugccuacguc auaagucacg uggggacggc ugcuguaaac acggaaguag cugacccgcg 60

ugacuuguca cgugagca 78

<210> 323
<211> 72
<212> RNA
<213> Alphatorquevirus sp.

<400> 323
uuguguaagg cggaacaggc ugacaccccg ugucaaaggu caggggucag ccuccgcuuu 60

gcaccaaaug gu 72

<210> 324
<211> 79
<212> RNA
<213> Alphatorquevirus sp.

<400> 324
uaccuacguc auaagucacg ugggaagagc ugcugugaac cuggaaguag cugacccgcg 60

uggcuuguca cgugagugc 79

<210> 325
<211> 75
<212> RNA
<213> Alphatorquevirus sp.

<400> 325
uuuuccuggc ccguccgcgg cgagagcgcg agcgaagcga gcgaucgggc gucccgaggg 60

cgggugccgg aggug 75

<210> 326
<211> 68
<212> RNA
<213> Alphatorquevirus sp.

<400> 326
aaagugagug gggccagacu ucgccauagg gccuuuaacu uccgggugcg ucuggggggcc 60

gccauuuu 68

<210> 327
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 327
gugacguuac ucucacguga ugggggcgug cucuaacccg gaagcauccu cgaccacgug 60

acugugacgu cac 73

<210> 328
<211> 75

<212> RNA
<213> Alphatorquevirus sp.

<400> 328
agcgucuacu acguacacuu ccuggggugu guccugccac uguauauaaa ccagaggggu 60

gacgaauggu agagu 75

<210> 329
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 329
gugacgucaa agucacgugg ugacggccau uuuaacccgg aaguggcugu ugucacguga 60

cuugacguca cgg 73

<210> 330
<211> 62
<212> RNA
<213> Alphatorquevirus sp.

<400> 330
gcuuuagacg ccauuuuagg cccucgcggg cacccguagg cgcguuuuaa ugacgucacg 60

gc 62

<210> 331
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 331
cacccguagg cgcguuuuaa ugacgucacg gcagccauuu ugucgugacg uuugagacac 60

gugaugggggg cgu 73

<210> 332
<211> 80
<212> RNA
<213> Alphatorquevirus sp.

<400> 332
gucgugacgu uugagacacg ugauggggggc gugccuaaac ccggaagcau cccuggucac 60

gugacucuga cgucacggcg 80

<210> 333
<211> 77
<212> RNA
<213> Alphatorquevirus sp.

<400> 333
cgaaagugag uggggccaga cuucgccaua aggccuuuaa cuuccgggug cgugugggggg 60

ccgccauuuu agcuucg 77

<210> 334
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 334
cugugacguc aaagucacgu ggggagggcg gcguguaacc cggaagucau ccucgucacg 60

ugaccugacg ucacgg 76

<210> 335
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 335
cuguccgcca ucuugugacu uccuuccgcu uuuucaaaaa aaaagaggaa guaugacgua 60

gcggcggggg ggc 73

<210> 336
<211> 67
<212> RNA
<213> Alphatorquevirus sp.

<400> 336
gguagaguuu uuuccgcccg uccgcagcga ggacgcgagc gcagcgagcg gccgagcgac 60

ccguggg 67

<210> 337
<211> 80
<212> RNA

<213> Alphatorquevirus sp.

<400> 337
gcugugacgu uucagucacg uggggaggga acgccuaaac ccggaagcgu cccuggucac 60

gugauuguga cgucacggcc 80

<210> 338
<211> 63
<212> RNA
<213> Alphatorquevirus sp.

<400> 338
ccgccauuuu gugacuuccu uccgcuuuuu caaaaaaaaa gaggaagugu gacguagcgg 60

cgg 63

<210> 339
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 339
gacugugacg ucaaagucac gugggggaggg cggcguguaa cccggaaguc auccucguca 60

cgugaccuga cgucacgg 78

<210> 340
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 340
cuguccgcca ucuugugacu uccuuccgcu uuuucaaaaa aaaagaggaa guaugacgug 60

gcggcggggg ggc 73

<210> 341
<211> 80
<212> RNA
<213> Alphatorquevirus sp.

<400> 341
gguugugacg ucaaagucac gugggggaggg cggcguguaa cccggaaguc auccucguca 60

cgugaccuga cgucacggcc 80

<210> 342
<211> 65
<212> RNA
<213> Alphatorquevirus sp.

<400> 342
cccgccaucu ugugacuucc uuccgcuuuu ucaaaaaaaa agaggaagug ugacguagcg 60

gcggg 65

<210> 343
<211> 67
<212> RNA
<213> Alphatorquevirus sp.

<400> 343
gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

guaggug 67

<210> 344
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 344
gacugugacg ucaaagucac gugggggagga gggcguguaa cccggaaguc auccucguca 60

cgugaccuga cgucacgg 78

<210> 345
<211> 62
<212> RNA
<213> Alphatorquevirus sp.

<400> 345
ucgcgucuua gugacgucac ggcagccauc uugguccuga cgucacuguc acgugggggag 60

gg 62

<210> 346
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 346
ugacgucacu gucacguggg gagggaacac gugaacccgg aaguguccu ggucacguga 60

caugacguca cggccg 76

<210> 347
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 347
cgccauuuua aguaagcaug gcgggcggug augucaaaug uuaaagguca cagccgguca 60

ugcuugcaca aaauggcg 78

<210> 348
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 348
cgccauuuua aguaagcaug gcgggcggug acgugcaaug ucaaagguca cagccuguca 60

ugcuugcaca aaauggcg 78

<210> 349
<211> 72
<212> RNA
<213> Alphatorquevirus sp.

<400> 349
ccaucuuaag uaguugaggc ggacgguggc gucgguucaa aggucaccau cagccacacc 60

uacucaaaau gg 72

<210> 350
<211> 67
<212> RNA
<213> Alphatorquevirus sp.

<400> 350
gccugucaug cuugcacaaa auggcggacu uccgcuuccg ggucgccgcc auauuugguc 60

acgugac 67

<210> 351
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 351
gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 352
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 352
gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 353
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 353
gcauacguca caagucacgu gggggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 354
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 354
gccauuuuaa guagcugacg ucaaggauug acgugaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 355
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 355
gcacacguca uaagucacgu gguggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugauuuguca cgugugua 78

<210> 356
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 356
gccauuuuaa gucagcucug gggaggcgug acuuccaguu caaaggucau ccucaccaua 60

acuggcacaa aauggc 76

<210> 357
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 357
gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 358
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 358
gcauacguca caagucacgu ggaggggaca cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 359
<211> 79
<212> RNA
<213> Alphatorquevirus sp.

<400> 359
gcgccauguu aaguggcugu cgccgaggau ugacgucaca guucaaaggu cauccucgac 60

gguaaccgca aacauggcg 79

<210> 360
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 360
caugcgucau aagucacaug acaggggucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 361
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 361
uggcagcacu uccgaauggc ugaguuuucc acgcccgucc gcggagaggg agccacggag 60

gugaucccga acg 73

<210> 362
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 362
gccauuuuaa gucagcgcug gggaggcaug acuguaaguu caaaggucau ccucaccgga 60

acugacacaa aauggccg 78

<210> 363
<211> 80
<212> RNA
<213> Alphatorquevirus sp.

<400> 363
gccaucuuaa guggcugucg ccgaggauug acgucacagu ucaaagguca uccucggcgg 60

uaaccgcaaa gauggcgguc 80

<210> 364
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 364

auacgucaua agucacaugu cuaggggucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 365
<211> 77
<212> RNA
<213> Alphatorquevirus sp.

<400> 365
ccauuuuaag uaaggcggaa gcagcugucc cuguaacaaa auggcggcga cagccuuccg 60

cuuugcacaa aauggag 77

<210> 366
<211> 80
<212> RNA
<213> Alphatorquevirus sp.

<400> 366
gccaucuuaa guggcugucg cugaggauug acgucacagu ucaaagguca uccucggcgg 60

uaaccgcaaa gauggcgguc 80

<210> 367
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 367
cauacgucau aagucacaug acaggagucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 368
<211> 80
<212> RNA
<213> Alphatorquevirus sp.

<400> 368
cgccaucuua aguggcuguc gccgaggauu ggcgucacag uucaaagguc auccucggcg 60

guaaccgcaa agauggcggu 80

<210> 369

<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 369
cauacgucau aagucacaug acaggggucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 370
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 370
gcauacguca caagucacgu gggggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cguggugu 78

<210> 371
<211> 77
<212> RNA
<213> Alphatorquevirus sp.

<400> 371
ccgccauuuu aggcuguugc cgggcguuug acuuccgugu uaaaggucaa acacccagcg 60

acaccaaaaa auggccg 77

<210> 372
<211> 77
<212> RNA
<213> Alphatorquevirus sp.

<400> 372
cuacgucaua agucacguga cagggagggg cgacaaaccc ggaagucauc cucgcccacg 60

ugacuuacca cguggug 77

<210> 373
<211> 77
<212> RNA
<213> Alphatorquevirus sp.

<400> 373
gccauuuuaa guaggugacg uccaggacug acguaaaguu caaaggucau ccucggcgga 60

accuauacaa aauggcg 77

<210> 374
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 374
cuacgucaua agucacgugg ggacggcugu acuuaaacac ggaaguaggc cccgucacgu 60

gauuuaccac guggug 76

<210> 375
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 375
gccauuuuaa guaaggcgga agagcucuag cuauacaaaa uggcggcgga gcacuuccgc 60

uuugcccaaa aug 73

<210> 376
<211> 77
<212> RNA
<213> Alphatorquevirus sp.

<400> 376
gccauuuuaa guagcugacg ucaaggauug acguagaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggug 77

<210> 377
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 377
gcauacguca caagucacgu gggggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 378
<211> 80

<212> RNA
<213> Alphatorquevirus sp.

<400> 378
ggcgccauuu uaaguaagca uggcgggcgg cgacgucaca ugucaaaggu caccgcacuu 60

ccgugcuugc acaaaauggc 80

<210> 379
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 379
ugcuacguca ucgagacacg uggugccagc agcuguaaac ccggaagucg cugacacacg 60

ugucuuguca cgu 73

<210> 380
<211> 78
<212> RNA
<213> Alphatorquevirus sp.

<400> 380
gccauuuuaa guaagcaccg ccuagggaug acguauaagu ucaaagguca uccucagccg 60

gaacuuacac aaaauggu 78

<210> 381
<211> 72
<212> RNA
<213> Alphatorquevirus sp.

<400> 381
acgucauaug ucacgugggg aggcccugcu gcgcaaacgc ggaaguaggc cccgucacgu 60

gucauaccac gu 72

<210> 382
<211> 77
<212> RNA
<213> Alphatorquevirus sp.

<400> 382
ccauuuuaag uaaggcggaa gcagcuccac uuucucacaa aauggcggcg gggcacuucc 60

ggcuugccca aaauggc 77

<210> 383
<211> 72
<212> RNA
<213> Alphatorquevirus sp.

<400> 383
ccauuuuaag uaaggcggaa guuucuccac uauacaaaau ggcggcggag cacuuccggc 60

uugcccaaaa ug 72

<210> 384
<211> 72
<212> RNA
<213> Alphatorquevirus sp.

<400> 384
ccaucuuaag uaguugaggc ggacgguggc gugaguucaa aggucaccau cagccacacc 60

uacucaaaau gg 72

<210> 385
<211> 76
<212> RNA
<213> Alphatorquevirus sp.

<400> 385
cgccaucuua aguaguugag gcggacggug gcgugaguuc aaaggucacc aucagccaca 60

ccuacucaaa auggug 76

<210> 386
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 386
uuucggaccu ucggcgucgg gggggucggg ggcuuuacua aacagacucc gagaugccau 60

uggacacuga ggg 73

<210> 387
<211> 76
<212> RNA

<213> Alphatorquevirus sp.

<400> 387
ccauuuuaag uaggugccgu ccagcacugc uguuccgggu uaaagggcau ccucggcgga 60

accuauacaa aauggc 76

<210> 388
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 388
cuacgucauc gaugacgugg ggaggcguac uaugaaacgc ggaaguaggc cccgcuacgu 60

caucaucacg ugg 73

<210> 389
<211> 73
<212> RNA
<213> Alphatorquevirus sp.

<400> 389
ccauuuuaag uaaggcggaa gagcugcucu auauacaaaa uggcggagga gcacuuccgg 60

cuugcccaaa aug 73

<210> 390
<211> 75
<212> RNA
<213> Alphatorquevirus sp.

<400> 390
ugccuacgua acaagucacg uggggagggu uggcguauaa cccggaaguc aauccuccca 60

cguggccugu cacgu 75

<210> 391
<211> 72
<212> RNA
<213> Alphatorquevirus sp.

<400> 391
uaaguaaggc ggaaccaggc ugucaccccg ugucaaaggu caggggucag ccuuccgcuu 60

uacacaaaau gg 72

<210> 392
<211> 72
<212> RNA
<213> Alphatorquevirus sp.

<400> 392
uaaguaaggc ggaaccaggc ugucaccccg ugucaaaggu caggggucag ccuuccgcuu 60

uacacaaaau gg 72

<210> 393
<211> 80
<212> RNA
<213> Alphatorquevirus sp.

<400> 393
gcagccauuu uaagucagcu ucggggaggg ucacgcaaag uucaaagguc auccucaccg 60

gaacugguac aaaauggccg 80

<210> 394
<211> 74
<212> RNA
<213> Alphatorquevirus sp.

<400> 394
ugcuacguca uaagugacgu agcugguguc ugcuguaaac acggaaguag gccccgccac 60

gucacuuguc acgu 74

<210> 395
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 395
aguagcugac gucaaggauu gac 23

<210> 396
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 396

caagucacgu ggaggggacc cg 22

<210> 397
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 397
aaguagcuga cgucaaggau ugacg 25

<210> 398
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 398
auaagucacg uggugggac ccg 23

<210> 399
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 399
uggggagggu uggcguauag cccgga 26

<210> 400
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 400
ccccccccgg gggggguuu gccc 24

<210> 401
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 401
aucagucacg ugggggaagg cgugc 25

<210> 402
<211> 22

<212> RNA
<213> Alphatorquevirus sp.

<400> 402
aaguaaggcg gaagcagcuc gg 22

<210> 403
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 403
agucacgugg ggaggguugg c 21

<210> 404
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 404
cccgaaggcg gguacccgag gu 22

<210> 405
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 405
uaucguagcc gacgcggacc ccg 23

<210> 406
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 406
auuuugugau acgcgcgucc ccuccc 26

<210> 407
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 407
aagucacgug acugggcagg u 21

<210> 408
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 408
ugacgucaaa gucacguggg ragggu 26

<210> 409
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 409
gaucgagcgu cccgugggcg ggu 23

<210> 410
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 410
ugacgucaaa gucacguggg gagggcgg 28

<210> 411
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 411
gaucgagcgu cccgugggcg ggu 23

<210> 412
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 412
gaucgagcgu cccgugggcg ggu 23

<210> 413
<211> 28
<212> RNA

<213> Alphatorquevirus sp.

<400> 413
ugacgucaaa gucacguggg gagggcgg 28

<210> 414
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 414
acgugguaag ucacgugggg gcagcu 26

<210> 415
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 415
auuuuauaau acgcgcgucc ccucc 25

<210> 416
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 416
aagggacagc cuuccggcuu gc 22

<210> 417
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 417
cauaagucac gugggacgg cugcu 25

<210> 418
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 418
uaaggcggaa caggcugaca cccc 24

<210> 419
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 419
uacgucauaa gucacguggg aagagcug 28

<210> 420
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 420
ucgggcgucc cgagggcggg ug 22

<210> 421
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 421
aaagugagug gggccagacu ucgcc 25

<210> 422
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 422
cucucacgug auggggggcgu gc 22

<210> 423
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 423
ucuacuacgu acacuuccug gggugugu 28

<210> 424
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 424
uggcuguugu cacgugacuu ga 22

<210> 425
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 425
agacgccauu uuaggcccuc gcgg 24

<210> 426
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 426
ugucgugacg uuugagacac gugau 25

<210> 427
<211> 30
<212> RNA
<213> Alphatorquevirus sp.

<400> 427
ugacguuuga gacacgugau gggggcgugc 30

<210> 428
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 428
agugaguggg gccagacuuc gc 22

<210> 429
<211> 30
<212> RNA
<213> Alphatorquevirus sp.

<400> 429
ugugacguca aagucacgug gggagggcgg 30

<210> 430
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 430
aaaagaggaa guaugacgua gcggcgg 27

<210> 431
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 431
agcgagcggc cgagcgaccc g 21

<210> 432
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 432
uucagucacg uggggaggga acgc 24

<210> 433
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 433
aaaagaggaa gugugacgua gcgg 24

<210> 434
<211> 30
<212> RNA
<213> Alphatorquevirus sp.

<400> 434
ugugacguca aagucacgug gggagggcgg 30

<210> 435
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 435
aaaagaggaa guaugacgug gcgg 24

<210> 436
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 436
ugacgucaaa gucacguggg gagggcgg 28

<210> 437
<211> 29
<212> RNA
<213> Alphatorquevirus sp.

<400> 437
aaaaaagagg aagugugacg uagcggcgg 29

<210> 438
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 438
gaucgagcgu cccgugggcg ggu 23

<210> 439
<211> 30
<212> RNA
<213> Alphatorquevirus sp.

<400> 439
ugugacguca aagucacgug gggaggaggg 30

<210> 440
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 440
uugguccuga cgucacuguc a 21

<210> 441

<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 441
cgucacuguc acguggggag ggaacac 27

<210> 442
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 442
uaaguaagca uggcgggcgg ugau 24

<210> 443
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 443
aaguaagcau ggcgggcggu ga 22

<210> 444
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 444
uaaguaguug aggcggacgg uggc 24

<210> 445
<211> 29
<212> RNA
<213> Alphatorquevirus sp.

<400> 445
ucaugcuugc acaaaauggc ggacuuccg 29

<210> 446
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 446

aguagcugac gucaaggauu gac 23

<210> 447
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 447
aguagcugac gucaaggauu gac 23

<210> 448
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 448
acaagucacg ugggggggac ccg 23

<210> 449
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 449
aaguagcuga cgucaaggau ugacg 25

<210> 450
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 450
auaagucacg ugguggggac ccg 23

<210> 451
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 451
aagucagcuc uggggaggcg ugacuu 26

<210> 452
<211> 23

<212> RNA
<213> Alphatorquevirus sp.

<400> 452
aguagcugac gucaaggauu gac 23

<210> 453
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 453
caagucacgu ggaggggaca cg 22

<210> 454
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 454
uguuaagugg cugucgccga ggauuga 27

<210> 455
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 455
uaagucacau gacaggguc ca 22

<210> 456
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 456
cggagaggga gccacggagg ug 22

<210> 457
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 457
aagucagcgc uggggaggca uga 23

<210> 458
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 458
ucuuaagugg cugucgccga ggauugac 28

<210> 459
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 459
aagucacaug ucuaggguc cacu 24

<210> 460
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 460
aaguaaggcg gaagcagcug ucc 23

<210> 461
<211> 29
<212> RNA
<213> Alphatorquevirus sp.

<400> 461
aucuuaagug gcugucgcug aggauugac 29

<210> 462
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 462
uaagucacau gacaggaguc cacu 24

<210> 463
<211> 22
<212> RNA

<213> Alphatorquevirus sp.

<400> 463
aaguggcugu cgccgaggau ug 22

<210> 464
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 464
uaagucacau gacagggguc ca 22

<210> 465
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 465
acaagucacg ugggggggac ccg 23

<210> 466
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 466
auuuuaggcu guugccgggc guuugacu 28

<210> 467
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 467
auaagucacg ugacagggag ggg 23

<210> 468
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 468
aaguagguga cguccaggac u 21

<210> 469
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 469
cauaagucac gugggacgg cugu 24

<210> 470
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 470
uaaguaaggc ggaagagcuc uagcua 26

<210> 471
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 471
aguagcugac gucaaggauu gac 23

<210> 472
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 472
acaagucacg ugggggggac ccg 23

<210> 473
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 473
uaaguaagca uggcgggcgg cgac 24

<210> 474
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 474
aucgagacac guggugccag cagcu 25

<210> 475
<211> 30
<212> RNA
<213> Alphatorquevirus sp.

<400> 475
ucauccucag ccggaacuua cacaaaaugg 30

<210> 476
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 476
auaugucacg uggggaggcc cugcug 26

<210> 477
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 477
aaguaaggcg gaagcagcuc cacuuu 26

<210> 478
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 478
aaguaaggcg gaaguuucuc cacu 24

<210> 479
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 479
uaaguaguug aggcggacgg uggc 24

<210> 480
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 480
uaaguaguug aggcggacgg ugg 23

<210> 481
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 481
gaccuucggc gucgggggg ucggggg 27

<210> 482
<211> 19
<212> RNA
<213> Alphatorquevirus sp.

<400> 482
auccucggcg gaaccuaua 19

<210> 483
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 483
aucgaugacg uggggaggcg uacuau 26

<210> 484
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 484
uggcggagga gcacuuccgg cuug 24

<210> 485
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 485
aacaagucac gugggggaggg uuggc 25

<210> 486
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 486
aggggucagc cuuccgcuuu a 21

<210> 487
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 487
aggggucagc cuuccgcuuu a 21

<210> 488
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 488
uaagucagcu ucggggaggg ucac 24

<210> 489
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 489
ucauaaguga cguagcuggu gucugcu 27

<210> 490
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 490
cauccucggc ggaagcuaca caa 23

<210> 491

<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 491
ggccccguca cgugacuuac cac 23

<210> 492
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 492
ucauccucgg cggaagcuac acaa 24

<210> 493
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 493
ggccccguca cgugauuugu cac 23

<210> 494
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 494
ccgggucaua ggucacaccu acgucac 27

<210> 495
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 495
ggcugccgcc ccccccgggg aaaggggg 28

<210> 496
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 496

auccucgucc acgugacugu ga 22

<210> 497
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 497
gagcacuucc ggcuugccca a 21

<210> 498
<211> 20
<212> RNA
<213> Alphatorquevirus sp.

<400> 498
caauccucuu acguggccug 20

<210> 499
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 499
cgagguuaag ggccaauucg ggcu 24

<210> 500
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 500
gggccccgc ggggcucucg gcg 23

<210> 501
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 501
gcggggcgug gccguaucag aaaaugg 27

<210> 502
<211> 19

<212> RNA
<213> Alphatorquevirus sp.

<400> 502
ccucggucac guggccugu 19

<210> 503
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 503
ccucgucacg ugaccugacg ucacag 26

<210> 504
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 504
ccguccgcgg cgagagcgcg agcga 25

<210> 505
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 505
auccucguca cgugaccuga cgucacg 27

<210> 506
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 506
ccguccgcgg cgagagcgcg agcga 25

<210> 507
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 507
ccguccgcgg cgagagcgcg agcga 25

<210> 508
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 508
auccucguca cgugaccuga cgucacg 27

<210> 509
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 509
cugacccgcg ugacugguca cguga 25

<210> 510
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 510
cggggcgugg ccguauuaga aaaugg 26

<210> 511
<211> 29
<212> RNA
<213> Alphatorquevirus sp.

<400> 511
aguaaggcgg aaccaggcug ucacccugu 29

<210> 512
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 512
uagcugaccc gcgugacuug ucac 24

<210> 513
<211> 19
<212> RNA

<213> Alphatorquevirus sp.

<400> 513
ggucagccuc cgcuuugca 19

<210> 514
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 514
gcugacccgc guggcuuguc acgugagu 28

<210> 515
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 515
ggcccguccg cggcgagagc gcgag 25

<210> 516
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 516
uccgggugcg ucuggggggcc gccauuu 27

<210> 517
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 517
auccucgacc acgugacugu g 21

<210> 518
<211> 30
<212> RNA
<213> Alphatorquevirus sp.

<400> 518
auaaaccaga ggggugacga augguagagu 30

<210> 519
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 519
caaagucacg uggugacggc cau 23

<210> 520
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 520
guaggcgcgu uuuaaugacg ucacgg 26

<210> 521
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 521
uaggcgcguu uuaaugacgu cacggcag 28

<210> 522
<211> 29
<212> RNA
<213> Alphatorquevirus sp.

<400> 522
aucccugguc acgugacucu gacgucacg 29

<210> 523
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 523
gcgugugggg gccgccauuu uagcuu 26

<210> 524
<211> 29
<212> RNA
<213> Alphatorquevirus sp.

<400> 524
ucauccucgu cacgugaccu gacgucacg 29

<210> 525
<211> 29
<212> RNA
<213> Alphatorquevirus sp.

<400> 525
cgccaucuug ugacuuccuu ccgcuuuuu 29

<210> 526
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 526
uagaguuuuu uccgcccguc cg 22

<210> 527
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 527
gucccugguc acgugauugu gac 23

<210> 528
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 528
cauuuuguga cuuccuuccg cuuuuu 26

<210> 529
<211> 29
<212> RNA
<213> Alphatorquevirus sp.

<400> 529
ucauccucgu cacgugaccu gacgucacg 29

<210> 530
<211> 30
<212> RNA
<213> Alphatorquevirus sp.

<400> 530
ccgccaucuu gugacuuccu uccgcuuuuu 30

<210> 531
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 531
auccucguca cgugaccuga cgucacg 27

<210> 532
<211> 30
<212> RNA
<213> Alphatorquevirus sp.

<400> 532
cgccaucuug ugacuuccuu ccgcuuuuuc 30

<210> 533
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 533
ccguccgcgg cgagagcgcg agcga 25

<210> 534
<211> 29
<212> RNA
<213> Alphatorquevirus sp.

<400> 534
ucauccucgu cacgugaccu gacgucacg 29

<210> 535
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 535
cuuagugacg ucacggcagc cau 23

<210> 536
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 536
gucccugguc acgugacaug acguc 25

<210> 537
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 537
cacagccggu caugcuugca caaa 24

<210> 538
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 538
acagccuguc augcuugcac aa 22

<210> 539
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 539
caccaucagc cacaccuacu caaa 24

<210> 540
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 540
cgggucgccg ccauauuugg ucacguga 28

<210> 541

<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 541
cauccucggc ggaagcuaca caa 23

<210> 542
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 542
cauccucggc ggaagcuaca caa 23

<210> 543
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 543
ggccccguca cgugacuuac cac 23

<210> 544
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 544
ucauccucgg cggaagcuac acaa 24

<210> 545
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 545
ggccccguca cgugauuugu cac 23

<210> 546
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 546

gucauccuca ccauaacugg cacaa 25

<210> 547
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 547
cauccucggc ggaagcuaca caa 23

<210> 548
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 548
ggccccguca cgugacuuac cac 23

<210> 549
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 549
auccucgacg guaaccgcaa acaug 25

<210> 550
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 550
ggccccgaca ugugacucgu c 21

<210> 551
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 551
agcacuuccg aauggcugag uuuucca 27

<210> 552
<211> 22

<212> RNA
<213> Alphatorquevirus sp.

<400> 552
auccucaccg gaacugacac aa 22

<210> 553
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 553
cauccucggc gguaaccgca aagaug 26

<210> 554
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 554
uaggccccga caugugacuc gu 22

<210> 555
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 555
acagccuucc gcuuugcaca a 21

<210> 556
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 556
cauccucggc gguaaccgca aagaugg 27

<210> 557
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 557
uaggccccga caugugacuc guc 23

<210> 558
<211> 20
<212> RNA
<213> Alphatorquevirus sp.

<400> 558
uccucggcgg uaaccgcaaa 20

<210> 559
<211> 21
<212> RNA
<213> Alphatorquevirus sp.

<400> 559
ggccccgaca ugugacucgu c 21

<210> 560
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 560
ggccccguca cgugacuuac cac 23

<210> 561
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 561
ucaaacaccc agcgacacca aaaaaugg 28

<210> 562
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 562
ccucgcccac gugacuuacc ac 22

<210> 563
<211> 20
<212> RNA

<213> Alphatorquevirus sp.

<400> 563
ccucggcgga accuauacaa 20

<210> 564
<211> 22
<212> RNA
<213> Alphatorquevirus sp.

<400> 564
gccccgucac gugauuuacc ac 22

<210> 565
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 565
gcggcggagc acuuccgcuu ugcccaaa 28

<210> 566
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 566
cauccucggc ggaagcuaca caa 23

<210> 567
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 567
ggccccguca cgugacuuac cac 23

<210> 568
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 568
caccgcacuu ccgugcuugc acaaa 25

<210> 569
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 569
ucgcugacac acgugucuug ucac 24

<210> 570
<211> 30
<212> RNA
<213> Alphatorquevirus sp.

<400> 570
cauuuuaagu aagcaccgcc uagggaugac 30

<210> 571
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 571
guaggccccg ucacguguca uaccac 26

<210> 572
<211> 25
<212> RNA
<213> Alphatorquevirus sp.

<400> 572
ggcggggcac uuccggcuug cccaa 25

<210> 573
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 573
cggcggagca cuuccggcuu gcccaa 26

<210> 574
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 574
caccaucagc cacaccuacu caaa 24

<210> 575
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 575
accaucagcc acaccuacuc aaa 23

<210> 576
<211> 28
<212> RNA
<213> Alphatorquevirus sp.

<400> 576
gacuccgaga ugccauugga cacugagg 28

<210> 577
<211> 18
<212> RNA
<213> Alphatorquevirus sp.

<400> 577
aguaggugcc guccagca 18

<210> 578
<211> 27
<212> RNA
<213> Alphatorquevirus sp.

<400> 578
aaguaggccc cgcuacguca ucaucac 27

<210> 579
<211> 23
<212> RNA
<213> Alphatorquevirus sp.

<400> 579
aaggcggaag agcugcucua uau 23

<210> 580
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 580
caauccuccc acguggccug ucac 24

<210> 581
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 581
aaggcggaac caggcuguca ccccgu 26

<210> 582
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 582
aaggcggaac caggcuguca ccccgu 26

<210> 583
<211> 24
<212> RNA
<213> Alphatorquevirus sp.

<400> 583
cauccucacc ggaacuggua caaa 24

<210> 584
<211> 26
<212> RNA
<213> Alphatorquevirus sp.

<400> 584
uaggccccgc cacgucacuu gucacg 26

<210> 585

<400> 585
000

<210> 586

<400> 586
000

<210> 587

<400> 587
000

<210> 588

<400> 588
000

<210> 589

<400> 589
000

<210> 590

<400> 590
000

<210> 591

<400> 591
000

<210> 592

<400> 592
000

<210> 593

<400> 593
000

<210> 594

<400> 594
000

<210> 595

<400> 595
000

<210> 596

<400> 596
000

<210> 597

<400> 597
000

<210> 598

<400> 598
000

<210> 599

<400> 599
000

<210> 600

<400> 600
000

<210> 601

<400> 601
000

<210> 602

<400> 602

000

<210> 603

<400> 603
000

<210> 604

<400> 604
000

<210> 605

<400> 605
000

<210> 606

<400> 606
000

<210> 607

<400> 607
000

<210> 608

<400> 608
000

<210> 609

<400> 609
000

<210> 610

<400> 610
000

<210> 611

<400> 611
000

<210> 612

<400> 612
000

<210> 613

<400> 613
000

<210> 614

<400> 614
000

<210> 615

<400> 615
000

<210> 616

<400> 616
000

<210> 617

<400> 617
000

<210> 618

<400> 618
000

<210> 619

<400> 619
000

<210> 620

<400> 620
000

<210> 621

<400> 621
000

<210> 622

<400> 622
000

<210> 623

<400> 623
000

<210> 624

<400> 624
000

<210> 625

<400> 625
000

<210> 626

<400> 626
000

<210> 627

<400> 627

000

<210> 628

<400> 628
000

<210> 629

<400> 629
000

<210> 630

<400> 630
000

<210> 631

<400> 631
000

<210> 632

<400> 632
000

<210> 633

<400> 633
000

<210> 634

<400> 634
000

<210> 635

<400> 635
000

<210> 636

<400> 636
000

<210> 637

<400> 637
000

<210> 638

<400> 638
000

<210> 639

<400> 639
000

<210> 640

<400> 640
000

<210> 641

<400> 641
000

<210> 642

<400> 642
000

<210> 643

<400> 643
000

<210> 644

<400> 644
000

<210> 645

<400> 645
000

<210> 646

<400> 646
000

<210> 647

<400> 647
000

<210> 648

<400> 648
000

<210> 649

<400> 649
000

<210> 650

<400> 650
000

<210> 651

<400> 651
000

<210> 652

<400> 652

000

<210> 653

<400> 653
000

<210> 654

<400> 654
000

<210> 655

<400> 655
000

<210> 656

<400> 656
000

<210> 657

<400> 657
000

<210> 658

<400> 658
000

<210> 659

<400> 659
000

<210> 660

<400> 660
000

<210> 661

<400> 661
000

<210> 662

<400> 662
000

<210> 663

<400> 663
000

<210> 664

<400> 664
000

<210> 665

<400> 665
000

<210> 666

<400> 666
000

<210> 667

<400> 667
000

<210> 668

<400> 668
000

<210> 669

<400> 669
000

<210> 670

<400> 670
000

<210> 671

<400> 671
000

<210> 672

<400> 672
000

<210> 673

<400> 673
000

<210> 674

<400> 674
000

<210> 675

<400> 675
000

<210> 676

<400> 676
000

<210> 677

<400> 677

000

<210> 678

<400> 678
000

<210> 679

<400> 679
000

<210> 680

<400> 680
000

<210> 681

<400> 681
000

<210> 682

<400> 682
000

<210> 683

<400> 683
000

<210> 684

<400> 684
000

<210> 685

<400> 685
000

<210> 686

<400> 686
000

<210> 687

<400> 687
000

<210> 688

<400> 688
000

<210> 689

<400> 689
000

<210> 690

<400> 690
000

<210> 691

<400> 691
000

<210> 692

<400> 692
000

<210> 693

<400> 693
000

<210> 694

<400> 694
000

<210> 695

<400> 695
000

<210> 696

<400> 696
000

<210> 697

<400> 697
000

<210> 698

<400> 698
000

<210> 699

<400> 699
000

<210> 700

<400> 700
000

<210> 701

<400> 701
000

<210> 702

<400> 702

000

<210> 703

<400> 703
000

<210> 704

<400> 704
000

<210> 705

<400> 705
000

<210> 706

<400> 706
000

<210> 707

<400> 707
000

<210> 708

<400> 708
000

<210> 709

<400> 709
000

<210> 710

<400> 710
000

<210> 711

<400> 711
000

<210> 712

<400> 712
000

<210> 713

<400> 713
000

<210> 714

<400> 714
000

<210> 715

<400> 715
000

<210> 716

<400> 716
000

<210> 717

<400> 717
000

<210> 718

<400> 718
000

<210> 719

<400> 719
000

<210> 720

<400> 720
000

<210> 721

<400> 721
000

<210> 722

<400> 722
000

<210> 723

<400> 723
000

<210> 724

<400> 724
000

<210> 725

<400> 725
000

<210> 726

<400> 726
000

<210> 727

<400> 727

000

<210> 728

<400> 728
000

<210> 729

<400> 729
000

<210> 730

<400> 730
000

<210> 731

<400> 731
000

<210> 732

<400> 732
000

<210> 733

<400> 733
000

<210> 734

<400> 734
000

<210> 735

<400> 735
000

<210> 736

<400> 736
000

<210> 737

<400> 737
000

<210> 738

<400> 738
000

<210> 739

<400> 739
000

<210> 740

<400> 740
000

<210> 741

<400> 741
000

<210> 742

<400> 742
000

<210> 743

<400> 743
000

<210> 744

<400> 744
000

<210> 745

<400> 745
000

<210> 746

<400> 746
000

<210> 747

<400> 747
000

<210> 748

<400> 748
000

<210> 749

<400> 749
000

<210> 750

<400> 750
000

<210> 751

<400> 751
000

<210> 752

<400> 752

000

<210> 753

<400> 753
000

<210> 754

<400> 754
000

<210> 755

<400> 755
000

<210> 756

<400> 756
000

<210> 757

<400> 757
000

<210> 758

<400> 758
000

<210> 759

<400> 759
000

<210> 760

<400> 760
000

<210> 761

<400> 761
000

<210> 762

<400> 762
000

<210> 763

<400> 763
000

<210> 764

<400> 764
000

<210> 765

<400> 765
000

<210> 766

<400> 766
000

<210> 767

<400> 767
000

<210> 768

<400> 768
000

<210> 769

<400> 769
000

<210> 770

<400> 770
000

<210> 771

<400> 771
000

<210> 772

<400> 772
000

<210> 773

<400> 773
000

<210> 774

<400> 774
000

<210> 775

<400> 775
000

<210> 776

<400> 776
000

<210> 777

<400> 777

000

<210> 778

<400> 778
000

<210> 779

<400> 779
000

<210> 780

<400> 780
000

<210> 781

<400> 781
000

<210> 782

<400> 782
000

<210> 783

<400> 783
000

<210> 784

<400> 784
000

<210> 785

<400> 785
000

<210> 786

<400> 786
000

<210> 787

<400> 787
000

<210> 788

<400> 788
000

<210> 789

<400> 789
000

<210> 790

<400> 790
000

<210> 791

<400> 791
000

<210> 792

<400> 792
000

<210> 793

<400> 793
000

<210> 794

<400> 794
000

<210> 795

<400> 795
000

<210> 796

<400> 796
000

<210> 797

<400> 797
000

<210> 798

<400> 798
000

<210> 799

<400> 799
000

<210> 800

<400> 800
000

<210> 801
<211> 156
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 801
gcggcggggg ggcggccgcg ttcgcgcgcc gcccaccagg gggtgctgcg cgccccccc 60

cgcgcatgcg cggggccccc ccccgggggg gctccgcccc cccggccccc ccccgtgcta 120

aacccaccgc gcatgcgcga ccacgccccc gccgcc 156

<210> 802
<211> 150
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 802
ccgagcgtta gcgaggagtg cgaccctacc ccctgggccc acttcttcgg agccgcgcgc 60

tacgccttcg gctgcgcgcg gcacctcaga cccccgctcg tgctgacacg cttgcgcgtg 120

tcagaccact tcgggctcgc gggggtcggg 150

<210> 803
<211> 122
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 803
gccgccgcgg cggcgggggg cggcgcgctg cgcgcgccgc ccagtagggg gagccatgcg 60

ccccccccg cgcatgcgcg gggcccccc ccgcgggggg ctccgccccc cggccccccc 120

cg 122

<210> 804
<211> 111
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 804
cggcccagcg gcggcgcgcg cgcttcgcgc gcgcgccggg gggctccgcc ccccccgcg 60

catgcgcggg gccccccccc gcgggggggct ccgcccccg gtcccccccc g 111

<210> 805
<211> 115
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 805
cggccgtgcg gcggcgcgcg cgcttcgcgc gcgcgccggg ggctgccgcc ccccccgcg 60

catgcgcgcg gggcccccc ccgcgggggg ctccgccccc cggcccccc ccccg 115

<210> 806
<211> 104
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 806
cggcggcggc gcgcgcgcta cgcgcgcgcg ccgggggggct gccgcccccc ccccgcgcat 60

gcgcggggcc ccccccgcg gggggctccg ccccccggcc cccc 104

<210> 807
<211> 108
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 807
ggcggcggcg cgcgcgctac gcgcgcgcgc cggggagctc tgcccccccc cgcgcatgcg 60

cgcgggtccc ccccccgcgg ggggctccgc cccccggtcc cccccccg 108

<210> 808

<400> 808
000

<210> 809

<400> 809
000

<210> 810

<400> 810
000

<210> 811

<400> 811
000

<210> 812

<400> 812
000

<210> 813

<400> 813
000

<210> 814

<400> 814
000

<210> 815

<400> 815
000

<210> 816

<400> 816
000

<210> 817

<400> 817
000

<210> 818

<400> 818
000

<210> 819

<400> 819
000

<210> 820

<400> 820
000

<210> 821

<400> 821
000

<210> 822

<400> 822
000

<210> 823

<400> 823
000

<210> 824

<400> 824
000

<210> 825

<400> 825
000

<210> 826

<400> 826
000

<210> 827

<400> 827
000

<210> 828

<400> 828
000

<210> 829
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<220>
<221> MOD_RES
<222> (4)..(5)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (7)..(7)
<223> Any amino acid

<220>
<221> MOD_RES

<222> (9)..(10)
<223> Any amino acid

<400> 829
Tyr Asn Pro Xaa Xaa Asp Xaa Gly Xaa Xaa Asn
1   5   10

<210> 830

<400> 830
000

<210> 831

<400> 831
000

<210> 832

<400> 832
000

<210> 833

<400> 833
000

<210> 834

<400> 834
000

<210> 835

<400> 835
000

<210> 836

<400> 836
000

<210> 837

<400> 837
000

<210> 838

<400> 838
000

<210> 839

<400> 839
000

<210> 840

<400> 840
000

<210> 841

<400> 841
000

<210> 842

<400> 842
000

<210> 843

<400> 843
000

<210> 844

<400> 844
000

<210> 845

<400> 845
000

<210> 846

<400> 846
000

<210> 847

<400> 847
000

<210> 848

<400> 848
000

<210> 849

<400> 849
000

<210> 850

<400> 850
000

<210> 851

<400> 851
000

<210> 852

<400> 852
000

<210> 853

<400> 853
000

<210> 854

<400> 854
000

<210> 855

<400> 855
000

<210> 856

<400> 856
000

<210> 857

<400> 857
000

<210> 858

<400> 858
000

<210> 859

<400> 859
000

<210> 860

<400> 860
000

<210> 861

<400> 861
000

<210> 862

<400> 862
000

<210> 863

<400> 863
000

<210> 864

<400> 864
000

<210> 865

<400> 865
000

<210> 866

<400> 866
000

<210> 867

<400> 867
000

<210> 868

<400> 868
000

<210> 869

<400> 869
000

<210> 870

<400> 870
000

<210> 871

<400> 871
000

<210> 872

<400> 872
000

<210> 873

<400> 873
000

<210> 874

<400> 874
000

<210> 875

<400> 875
000

<210> 876

<400> 876
000

<210> 877

<400> 877
000

<210> 878
<211> 3264
<212> DNA
<213> Gammatorquevirus sp.

<400> 878
taaaatggcg gcaaccaatc attttatact ttcactttcc aattacaagc cgccacgtca 60

cagaacaggg gtggagactt taaaactata taaccaagtg atgtgacgaa tggctgagtt 120

taccccgcta gacggtgcag ggaccggatc gagcgcagcg aggaggtccc cggctgcccg 180

tgggcgggag cccgaggtga gtgaaaccac cgaggtctag gggcaattcg ggctagggca 240

gtctagcgga acgggcaaga aacttaaaat atgttttgtt tcagatgcag acacctgctt 300

cacagataag ctcagacgac ttctttgtac acactccatt taatgcagta actaaacagc 360

aaatatggat gtctcaaatt gctgatggac atgacaacat ttgtcactgc caccgtcctt 420

ttgctcacct gcttgctaat atttttcctc ctggtcataa agacagggat cttaccatta 480

atcaaatact tgctagagat cttacagaaa catgccattc tggtggagac gaaggaacaa 540

gcggtggtgg ggtcgccgct tccgctaccg ccgctacaac aaatataaaa ccagaaggag 600

acgcagaata cccagaagac gaaatagaag atttactaag acacgcagga gaagaaaaag 660

aaagaaggta agaagaaaac ttaaaaaaat tactattaaa caatggcagc cagattcagt 720

gaaaaaatgt aaaattaaag gatatagtac tttagttatg ggtgcacaag gaaaacaata 780

caactgttac acaaaccaag caagtgacta tgttcagcct aaagcaccac aaggtggggg 840

ctttggctgt gaagtattta atttaaaatg gctataccaa gaatatactg cacacagaaa 900

tatttggaca aaaacaaatg aatatacaga cctttgtaga tacactggag ctcaaataat 960

tttatacagg cacccagatg ttgattttat agtcagctgg gacaatcagc cacctttttt 1020

acttaacaaa tatacatatc cagaactgca accacaaaac cttttactag ctagaaggaa 1080

aagaattatt cttagtcaaa aatcaaaccc caaaggaaaa ctaagaatta aactaagaat 1140

accaccacca aaacaaatga taacaaaatg gttttttcaa agagactttt gtgatgtgaa 1200

tctgtttaaa ctatgtgctt ctgctgcttc tttccgctac ccaggtatca gtcatggagc 1260

tcaaagtact atttttttctg catatgcttt aaacactgac ttttatcaat gcagtgactg 1320

gtgccaaact aacacagaaa ctggctacct aaacattaaa acacaacaaa tgccactatg 1380

gtttcattac agagagggtg gcaaagagaa atggtataaa tacaccaaca aagaacacag 1440

accatataca aatacatatc ttaaaagtat tagctataat gatggattgt tttctcctaa 1500

agccatgttt gcatttgaag taaaagcggg gggtgaagga acaacagaac caccacaagg 1560

cgcccaatta attgctaacc ttccactcat tgcactaaga tataatccac atgaagacac 1620

aggccatggc aatgaaattt accttacatc aacttttaaa ggtacatatg acaaacctaa 1680

agttactgat gctctatact ttaacaatgt acccctgtgg atgggatttt atggctactg 1740

ggactttata ttacaagaaa caaaaaacaa aggtgtcttt gatcaacata tgtttgttgt 1800

taaatgtcct gccttaaggc ccatatcaca agtcacaaaa caagtatact acccacttgt 1860

agacatggac ttttgttcag ggagactgcc atttgatgaa tatttatcca aagacattaa 1920

aagtcattgg tatcccactg cagaaagaca aacagttaca ataaataatt ttgttacagc 1980

aggtccatac atgcctaaat ttgaacccac agacaaagac agtacatggc aattaaacta 2040

tcactataaa tttttttta agtggggtgg tccacaagtc acagacccaa ctgttgaaga 2100

cccatgcagc agaaacaaat atcctgtccc cgatacaatg caacaaacaa tacaaattaa 2160

aaaccctgaa aagctgcacc cagcaaccct cttccatgac tgggacctta gaaggggctt 2220

cattacacaa gcagctatta aaagaatgtc agaaaacctc caaattgatt catctttcga 2280

atctgatggc acagaatcac ccaaaaaaaa gaaaagatgc accaaagaaa tcccaacaca 2340

aaaccaaaag caagaagaga tccaagaatg tctcctctca ctctgcgaag agcctacatg 2400

ccaagaagaa acagaggacc tccagctctt catccagcag cagcagcagc agcagtacaa 2460

gctcagaaaa aacctcttca aactcctcac tcacctgaaa aaaggacaga gaataagtca 2520

actacaaacg ggacttttag agtaatacca tttaaaccag gttttgaaca agaaacagaa 2580

aaagaacttg ccatagcttt ctgcagacca cctagaaaat ataaaaatga tccccctttt 2640

tatccctggt taccatggac acccccttgta cactttaacc ttaattacaa aggctaggcc 2700

aacactgttc acttagtggt gtatgtttaa taaagtttca cccccaaaaa aaaaaaaaaa 2760

aaaaaaaaaa aaaaaaaaaa taaaaaattg caaaaattcg gcgctcgcgc gcgctgcgcg 2820

cgcgcgagcg ccgtcacgcg ccggcgctcg cgcgccgcgc gtatgtgcta acacaccacg 2880

cacctagatt ggggtgcgcg cgctagcgcg cgcaccccaa tgcgccccgc cctcgttccg 2940

acccgcttgc gcgggtcgga ccacttcggg ctcgggggggg cgcgcctgcg gcgctttttt 3000

actaaacaga ctccgagccg ccatttggcc ccccctaagc tccgcccccc tcatgaatat 3060

tcataaagga aaccacataa ttagaattgc cgaccacaaa ctgccatatg ctaattagtt 3120

ccccttttac acagtaaaaa ggggaagtgg gggggcatag ccccccccaca ccccccgcgg 3180

ggggggcaga gccccccccc gcacccccccc cctacgtcac aatccacgcc cccgccgcca 3240

tcttgggtgc ggcagggcgg gggc 3264

<210> 879
<211> 128
<212> PRT
<213> Gammatorquevirus sp.

<400> 879
Met Gln Thr Pro Ala Ser Gln Ile Ser Ser Asp Asp Phe Phe Val His
1 5 10 15

Thr Pro Phe Asn Ala Val Thr Lys Gln Gln Ile Trp Met Ser Gln Ile
20 25 30

Ala Asp Gly His Asp Asn Ile Cys His Cys His Arg Pro Phe Ala His
35 40 45

Leu Leu Ala Asn Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr
50 55 60

Ile Asn Gln Ile Leu Ala Arg Asp Leu Thr Glu Thr Cys His Ser Gly
65 70 75 80

Gly Asp Glu Gly Thr Ser Gly Gly Gly Val Ala Ala Ser Ala Thr Ala
85 90 95

Ala Thr Thr Asn Ile Lys Pro Glu Gly Asp Ala Glu Tyr Pro Glu Asp
100 105 110

Glu Ile Glu Asp Leu Leu Arg His Ala Gly Glu Glu Lys Glu Arg Arg
115 120 125

<210> 880
<211> 272
<212> PRT
<213> Gammatorquevirus sp.

<400> 880
Met Gln Thr Pro Ala Ser Gln Ile Ser Ser Asp Asp Phe Phe Val His
1   5   10   15

Thr Pro Phe Asn Ala Val Thr Lys Gln Gln Ile Trp Met Ser Gln Ile
            20   25   30

Ala Asp Gly His Asp Asn Ile Cys His Cys His Arg Pro Phe Ala His
        35   40   45

Leu Leu Ala Asn Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr
    50   55   60

Ile Asn Gln Ile Leu Ala Arg Asp Leu Thr Glu Thr Cys His Ser Gly
65   70   75   80

Gly Asp Glu Gly Thr Ser Gly Gly Gly Val Ala Ala Ser Ala Thr Ala
                85   90   95

Ala Thr Thr Asn Ile Lys Pro Glu Gly Asp Ala Glu Tyr Pro Glu Asp
            100   105   110

Glu Ile Glu Asp Leu Leu Arg His Ala Gly Glu Glu Lys Glu Arg Ser
        115   120   125

Gly Val Val His Lys Ser Gln Thr Gln Leu Leu Lys Thr His Ala Ala
    130   135   140

Glu Thr Asn Ile Leu Ser Pro Ile Gln Cys Asn Lys Gln Tyr Lys Leu
145   150   155   160

Lys Thr Leu Lys Ser Cys Thr Gln Gln Pro Ser Ser Met Thr Gly Thr
                165   170   175

Leu Glu Gly Ala Ser Leu His Lys Gln Leu Leu Lys Glu Cys Gln Lys
180 185 190

Thr Ser Lys Leu Ile His Leu Ser Asn Leu Met Ala Gln Asn His Pro
195 200 205

Lys Lys Arg Lys Asp Ala Pro Lys Lys Ser Gln His Lys Thr Lys Ser
210 215 220

Lys Lys Arg Ser Lys Asn Val Ser Ser His Ser Ala Lys Ser Leu His
225 230 235 240

Ala Lys Lys Lys Gln Arg Thr Ser Ser Ser Ser Ser Ser Ser Ser Ser
245 250 255

Ser Ser Ser Thr Ser Ser Glu Lys Thr Ser Ser Asn Ser Ser Leu Thr
260 265 270

<210> 881
<211> 261
<212> PRT
<213> Gammatorquevirus sp.

<400> 881
Met Gln Thr Pro Ala Ser Gln Ile Ser Ser Asp Asp Phe Phe Val His
1 5 10 15

Thr Pro Phe Asn Ala Val Thr Lys Gln Gln Ile Trp Met Ser Gln Ile
20 25 30

Ala Asp Gly His Asp Asn Ile Cys His Cys His Arg Pro Phe Ala His
35 40 45

Leu Leu Ala Asn Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr
50 55 60

Ile Asn Gln Ile Leu Ala Arg Asp Leu Thr Glu Thr Cys His Ser Gly
65 70 75 80

```
Gly Asp Glu Gly Thr Ser Gly Gly Gly Val Ala Ala Ser Ala Thr Ala
                85                  90                  95

Ala Thr Thr Asn Ile Lys Pro Glu Gly Asp Ala Glu Tyr Pro Glu Asp
            100                 105                 110

Glu Ile Glu Asp Leu Leu Arg His Ala Gly Glu Glu Lys Glu Arg Arg
        115                 120                 125

Ile Thr Gln Lys Lys Glu Lys Met His Gln Arg Asn Pro Asn Thr Lys
    130                 135                 140

Pro Lys Ala Arg Arg Asp Pro Arg Met Ser Pro Leu Thr Leu Arg Arg
145                 150                 155                 160

Ala Tyr Met Pro Arg Arg Asn Arg Gly Pro Pro Ala Leu His Pro Ala
                165                 170                 175

Ala Ala Ala Ala Ala Val Gln Ala Gln Lys Lys Pro Leu Gln Thr Pro
            180                 185                 190

His Ser Pro Glu Lys Arg Thr Glu Asn Lys Ser Thr Thr Asn Gly Thr
        195                 200                 205

Phe Arg Val Ile Pro Phe Lys Pro Gly Phe Glu Gln Glu Thr Glu Lys
    210                 215                 220

Glu Leu Ala Ile Ala Phe Cys Arg Pro Pro Arg Lys Tyr Lys Asn Asp
225                 230                 235                 240

Pro Pro Phe Tyr Pro Trp Leu Pro Trp Thr Pro Leu Val His Phe Asn
                245                 250                 255

Leu Asn Tyr Lys Gly
            260
```

<210> 882

<211> 67
<212> PRT
<213> Gammatorquevirus sp.

<400> 882
Met Asp Met Thr Thr Phe Val Thr Ala Thr Val Leu Leu Leu Thr Cys
1   5   10   15

Leu Leu Ile Phe Phe Leu Leu Val Ile Lys Thr Gly Ile Leu Pro Leu
            20   25   30

Ile Lys Tyr Leu Leu Glu Ile Leu Gln Lys His Ala Ile Leu Val Glu
        35   40   45

Thr Lys Glu Gln Ala Val Val Gly Ser Pro Leu Pro Leu Pro Pro Leu
    50   55   60

Gln Gln Ile
65

<210> 883
<211> 677
<212> PRT
<213> Gammatorquevirus sp.

<400> 883
Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg
1   5   10   15

Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg
            20   25   30

Ile Pro Arg Arg Arg Asn Arg Arg Phe Thr Lys Thr Arg Arg Arg Arg
        35   40   45

Lys Arg Lys Lys Val Arg Arg Lys Leu Lys Lys Ile Thr Ile Lys Gln
    50   55   60

Trp Gln Pro Asp Ser Val Lys Lys Cys Lys Ile Lys Gly Tyr Ser Thr
65   70   75   80

```
Leu Val Met Gly Ala Gln Gly Lys Gln Tyr Asn Cys Tyr Thr Asn Gln
                85                  90                  95

Ala Ser Asp Tyr Val Gln Pro Lys Ala Pro Gln Gly Gly Gly Phe Gly
            100                 105                 110

Cys Glu Val Phe Asn Leu Lys Trp Leu Tyr Gln Glu Tyr Thr Ala His
        115                 120                 125

Arg Asn Ile Trp Thr Lys Thr Asn Glu Tyr Thr Asp Leu Cys Arg Tyr
    130                 135                 140

Thr Gly Ala Gln Ile Ile Leu Tyr Arg His Pro Asp Val Asp Phe Ile
145                 150                 155                 160

Val Ser Trp Asp Asn Gln Pro Pro Phe Leu Leu Asn Lys Tyr Thr Tyr
                165                 170                 175

Pro Glu Leu Gln Pro Gln Asn Leu Leu Leu Ala Arg Arg Lys Arg Ile
            180                 185                 190

Ile Leu Ser Gln Lys Ser Asn Pro Lys Gly Lys Leu Arg Ile Lys Leu
        195                 200                 205

Arg Ile Pro Pro Pro Lys Gln Met Ile Thr Lys Trp Phe Phe Gln Arg
    210                 215                 220

Asp Phe Cys Asp Val Asn Leu Phe Lys Leu Cys Ala Ser Ala Ala Ser
225                 230                 235                 240

Phe Arg Tyr Pro Gly Ile Ser His Gly Ala Gln Ser Thr Ile Phe Ser
                245                 250                 255

Ala Tyr Ala Leu Asn Thr Asp Phe Tyr Gln Cys Ser Asp Trp Cys Gln
            260                 265                 270

Thr Asn Thr Glu Thr Gly Tyr Leu Asn Ile Lys Thr Gln Gln Met Pro
```

```
        275                 280                 285

Leu Trp Phe His Tyr Arg Glu Gly Gly Lys Glu Lys Trp Tyr Lys Tyr
    290                 295                 300

Thr Asn Lys Glu His Arg Pro Tyr Thr Asn Thr Tyr Leu Lys Ser Ile
305                 310                 315                 320

Ser Tyr Asn Asp Gly Leu Phe Ser Pro Lys Ala Met Phe Ala Phe Glu
                325                 330                 335

Val Lys Ala Gly Gly Glu Gly Thr Thr Glu Pro Pro Gln Gly Ala Gln
            340                 345                 350

Leu Ile Ala Asn Leu Pro Leu Ile Ala Leu Arg Tyr Asn Pro His Glu
        355                 360                 365

Asp Thr Gly His Gly Asn Glu Ile Tyr Leu Thr Ser Thr Phe Lys Gly
    370                 375                 380

Thr Tyr Asp Lys Pro Lys Val Thr Asp Ala Leu Tyr Phe Asn Asn Val
385                 390                 395                 400

Pro Leu Trp Met Gly Phe Tyr Gly Tyr Trp Asp Phe Ile Leu Gln Glu
                405                 410                 415

Thr Lys Asn Lys Gly Val Phe Asp Gln His Met Phe Val Val Lys Cys
            420                 425                 430

Pro Ala Leu Arg Pro Ile Ser Gln Val Thr Lys Gln Val Tyr Tyr Pro
        435                 440                 445

Leu Val Asp Met Asp Phe Cys Ser Gly Arg Leu Pro Phe Asp Glu Tyr
    450                 455                 460

Leu Ser Lys Asp Ile Lys Ser His Trp Tyr Pro Thr Ala Glu Arg Gln
465                 470                 475                 480
```

```
Thr Val Thr Ile Asn Asn Phe Val Thr Ala Gly Pro Tyr Met Pro Lys
                485                 490                 495

Phe Glu Pro Thr Asp Lys Asp Ser Thr Trp Gln Leu Asn Tyr His Tyr
            500                 505                 510

Lys Phe Phe Phe Lys Trp Gly Gly Pro Gln Val Thr Asp Pro Thr Val
        515                 520                 525

Glu Asp Pro Cys Ser Arg Asn Lys Tyr Pro Val Pro Asp Thr Met Gln
    530                 535                 540

Gln Thr Ile Gln Ile Lys Asn Pro Glu Lys Leu His Pro Ala Thr Leu
545                 550                 555                 560

Phe His Asp Trp Asp Leu Arg Arg Gly Phe Ile Thr Gln Ala Ala Ile
                565                 570                 575

Lys Arg Met Ser Glu Asn Leu Gln Ile Asp Ser Ser Phe Glu Ser Asp
            580                 585                 590

Gly Thr Glu Ser Pro Lys Lys Lys Lys Arg Cys Thr Lys Glu Ile Pro
        595                 600                 605

Thr Gln Asn Gln Lys Gln Glu Glu Ile Gln Glu Cys Leu Leu Ser Leu
    610                 615                 620

Cys Glu Glu Pro Thr Cys Gln Glu Glu Thr Glu Asp Leu Gln Leu Phe
625                 630                 635                 640

Ile Gln Gln Gln Gln Gln Gln Gln Tyr Lys Leu Arg Lys Asn Leu Phe
                645                 650                 655

Lys Leu Leu Thr His Leu Lys Lys Gly Gln Arg Ile Ser Gln Leu Gln
            660                 665                 670

Thr Gly Leu Leu Glu
```

675

<210> 884
<211> 212
<212> PRT
<213> Gammatorquevirus sp.

<400> 884
Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg
1 5 10 15

Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg
20 25 30

Ile Pro Arg Arg Arg Asn Arg Arg Phe Thr Lys Thr Arg Arg Arg Arg
35 40 45

Lys Arg Lys Lys Trp Gly Gly Pro Gln Val Thr Asp Pro Thr Val Glu
50 55 60

Asp Pro Cys Ser Arg Asn Lys Tyr Pro Val Pro Asp Thr Met Gln Gln
65 70 75 80

Thr Ile Gln Ile Lys Asn Pro Glu Lys Leu His Pro Ala Thr Leu Phe
85 90 95

His Asp Trp Asp Leu Arg Arg Gly Phe Ile Thr Gln Ala Ala Ile Lys
100 105 110

Arg Met Ser Glu Asn Leu Gln Ile Asp Ser Ser Phe Glu Ser Asp Gly
115 120 125

Thr Glu Ser Pro Lys Lys Lys Lys Arg Cys Thr Lys Glu Ile Pro Thr
130 135 140

Gln Asn Gln Lys Gln Glu Glu Ile Gln Glu Cys Leu Leu Ser Leu Cys
145 150 155 160

Glu Glu Pro Thr Cys Gln Glu Glu Thr Glu Asp Leu Gln Leu Phe Ile

```
                165                 170                 175

Gln Gln Gln Gln Gln Gln Gln Tyr Lys Leu Arg Lys Asn Leu Phe Lys
            180                 185                 190

Leu Leu Thr His Leu Lys Lys Gly Gln Arg Ile Ser Gln Leu Gln Thr
        195                 200                 205

Gly Leu Leu Glu
    210
```

<210> 885
<211> 119
<212> PRT
<213> Gammatorquevirus sp.

<400> 885

```
Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg
1               5                   10                  15

Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg
            20                  25                  30

Ile Pro Arg Arg Arg Asn Arg Arg Phe Thr Lys Thr Arg Arg Arg Arg
        35                  40                  45

Lys Arg Lys Lys Asn His Pro Lys Lys Arg Lys Asp Ala Pro Lys Lys
    50                  55                  60

Ser Gln His Lys Thr Lys Ser Lys Lys Arg Ser Lys Asn Val Ser Ser
65                  70                  75                  80

His Ser Ala Lys Ser Leu His Ala Lys Lys Lys Gln Arg Thr Ser Ser
                85                  90                  95

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Thr Ser Ser Glu Lys Thr
            100                 105                 110

Ser Ser Asn Ser Ser Leu Thr
```

<210> 886
<211> 3176
<212> DNA
<213> Gammatorquevirus sp.

<400> 886
taaaatggcg ggagccaatc attttatact ttcactttcc aattaaaaat ggccacgtca 60

caaacaaggg gtggagccat ttaaactata taactaagtg gggtggcgaa tggctgagtt 120

taccccgcta gacggtgcag ggaccggatc gagcgcagcg aggaggtccc cggctgccca 180

tgggcgggag ccgaggtgag tgaaaccacc gaggtctagg ggcaattcgg gctagggcag 240

tctagcggaa cgggcaagaa acttaaaaca atatttgttt tacagatggt tagtatatcc 300

tcaagtgatt tttttaagaa aacgaaattt aatgaggaga cgcagaacca agtatggatg 360

tctcaaattg ctgactctca tgataatatc tgcagttgct ggcatccatt tgctcacctt 420

cttgcttcca tatttcctcc tggccacaaa gatcgtgatc ttactattaa ccaaattctt 480

ctaagagatt ataaagaaaa atgccattct ggtggagaag aaggagaaaa ttctggacca 540

acaacaggtt taattacacc aaaagaagaa gatatagaaa aagatggccc agaaggcgcc 600

gcagaagaag accatacaga cgccctgttc gccgccgccg tagaaaactt cgaaaggtaa 660

agagaaaaaa aaaatcttta attgttagac aatggcaacc agacagtata agaacttgta 720

aaattatagg acagtcagct atagttgttg gggctgaagg aaagcaaatg tactgttata 780

ctgtcaataa gttaattaat gtgcccccaa aaacaccata ggggggaggc tttggagtag 840

accaatacac actgaaatac ttatatgaag aatacagatt tgcacaaaac atttggacac 900

aatctaatgt actgaaagac ttatgcagat acataaatgt taagctaata ttctacagag 960

acaacaaaac agactttgtc ctttcctatg acagaaaccc acctttcaa ctaacaaaat 1020

ttacataccc aggagcacac ccacaacaaa tcatgcttca aaaacaccac aaattcatac 1080

tatcacaaat gacaaagcct aatggaagac taacaaaaaa actcaaaatt aaacctccta 1140

aacaaatgct ttctaaatgg ttcttttcaa aacaattctg taaataccct ttactatctc 1200

ttaaagcttc tgcactagac cttaggcact cttacctagg ctgctgtaat gaaaatccac 1260

aggtattttt ttattattta aaccatggat actacacaat aacaaactgg ggagcacaat 1320

cctcaacagc atacagacct aactccaagg tgacagacac aacatactac agatacaaaa 1380

atgacagaaa aaatattaac attaaaagcc atgaatacga aaaaagtata tcatatgaaa 1440

acggttattt tcaatctagt ttcttacaaa cacagtgcat atataccagt gagcgtggtg 1500

aagcctgtat agcagaaaaa ccactaggaa tagctattta caatccagta aaagacaatg 1560

gagatggtaa tatgatatac cttgtaagca ctctagcaaa cacttgggac cagcctccaa 1620

aagacagtgc tattttaata caaggagtac ccatatggct aggcttattt ggatatttag 1680

actactgtag acaaattaaa gctgacaaaa catggctaga cagtcatgta ctagtaattc 1740

aaagtcctgc tatttttact tacccaaatc caggagcagg caaatggtat tgtccactat 1800

cacaaagttt tataaatggc aatggtccgt ttaatcaacc acctacactg ctacaaaaag 1860

caaagtggtt tccacaaata caataccaac aagaaattat taatagcttt gtagaatcag 1920

gaccatttgt tcccaaatat gcaaatcaaa ctgaaagcaa ctgggaacta aaatataaat 1980

atgtttttac atttaagtgg ggtggaccac aattccatga accagaaatt gctgacccta 2040

gcaaacaaga gcagtatgat gtccccgata ctttctacca aacaatacaa attgaagatc 2100

cagaaggaca agaccccaga tctctcatcc atgattggga ctacagacga ggctttatta 2160

aagaaagatc tcttaaaaga atgtcaactt acttctcaac tcatacagat cagcaagcaa 2220

cttcagagga agacattccc aaaaagaaaa agagaattgg accccaactc acagtcccac 2280

aacaaaaaga agaggagaca ctgtcatgtc tcctctctct ctgcaaaaaa gataccttcc 2340

aagaaacaga gacacaagaa gacctccagc agctcatcaa gcagcagcag gagcagcagc 2400

tcctcctcaa gagaaacatc ctccagctca tccacaaact aaaagagaat caacaaatgc 2460

ttcagcttca cacaggcatg ttaccttaac cagatttaaa cctggatttg aagagcaaac 2520

agagagagaa ttagcaatta tatttcatag gccccctaga acctacaaag aggaccttcc 2580

attctatccc tggctaccac ctgcacccct tgtacaattt aaccttaact tcaaaggcta 2640

ggccaacaat gtacacttag taaagcatgt ttattaaagc acaaccccca aaataaatgt 2700

aaaaataaaa aaaaaaaaaa aaaaataaaa aattgcaaaa attcggcgct cgcgcgcatg 2760

tgcgcctctg gcgcaaatca cgcaacgctc gcgcgcccgc gtatgtctct ttaccacgca 2820

cctagattgg ggtgcgcgcg ctagcgcgcg caccccaatg cgccccgccc tcgttccgac 2880

ccgcttgcgc gggtcggacc acttcgggct cgggggggcg cgcctgcggc gctttttac 2940

taaacagact ccgagccgcc atttggcccc ctaagctccg ccccctcat gaatattcat 3000

aaaggaaacc acataattag aattgccgac cacaaactgc catatgctaa ttagttcccc 3060

ttttacaaag taaaagggga agtgaacata gccccacacc cgcaggggca aggccccgca 3120

cccctacgtc actaaccacg cccccgccgc catcttgggt gcggcagggc gggggc 3176

<210> 887
<211> 124
<212> PRT
<213> Gammatorquevirus sp.

<400> 887
Met Val Ser Ile Ser Ser Ser Asp Phe Phe Lys Lys Thr Lys Phe Asn
1 5 10 15

Glu Glu Thr Gln Asn Gln Val Trp Met Ser Gln Ile Ala Asp Ser His
20 25 30

Asp Asn Ile Cys Ser Cys Trp His Pro Phe Ala His Leu Leu Ala Ser
35 40 45

Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr Ile Asn Gln Ile
50 55 60

Leu Leu Arg Asp Tyr Lys Glu Lys Cys His Ser Gly Gly Glu Glu Gly
65 70 75 80

Glu Asn Ser Gly Pro Thr Thr Gly Leu Ile Thr Pro Lys Glu Glu Asp
85 90 95

Ile Glu Lys Asp Gly Pro Glu Gly Ala Ala Glu Glu Asp His Thr Asp
100 105 110

Ala Leu Phe Ala Ala Ala Val Glu Asn Phe Glu Arg
        115   120

<210> 888
<211> 271
<212> PRT
<213> Gammatorquevirus sp.

<400> 888
Met Val Ser Ile Ser Ser Ser Asp Phe Phe Lys Lys Thr Lys Phe Asn
1   5   10   15

Glu Glu Thr Gln Asn Gln Val Trp Met Ser Gln Ile Ala Asp Ser His
            20   25   30

Asp Asn Ile Cys Ser Cys Trp His Pro Phe Ala His Leu Leu Ala Ser
        35   40   45

Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr Ile Asn Gln Ile
    50   55   60

Leu Leu Arg Asp Tyr Lys Glu Lys Cys His Ser Gly Gly Glu Glu Gly
65   70   75   80

Glu Asn Ser Gly Pro Thr Thr Gly Leu Ile Thr Pro Lys Glu Glu Asp
                85   90   95

Ile Glu Lys Asp Gly Pro Glu Gly Ala Ala Glu Glu Asp His Thr Asp
            100   105   110

Ala Leu Phe Ala Ala Ala Val Glu Asn Phe Glu Ser Gly Val Asp His
        115   120   125

Asn Ser Met Asn Gln Lys Leu Leu Thr Leu Ala Asn Lys Ser Ser Met
    130   135   140

Met Ser Pro Ile Leu Ser Thr Lys Gln Tyr Lys Leu Lys Ile Gln Lys
145   150   155   160

```
Asp Lys Thr Pro Asp Leu Ser Ser Met Ile Gly Thr Thr Asp Glu Ala
                165                 170                 175

Leu Leu Lys Lys Asp Leu Leu Lys Glu Cys Gln Leu Thr Ser Gln Leu
            180                 185                 190

Ile Gln Ile Ser Lys Gln Leu Gln Arg Lys Thr Phe Pro Lys Arg Lys
        195                 200                 205

Arg Glu Leu Asp Pro Asn Ser Gln Ser His Asn Lys Lys Lys Arg Arg
    210                 215                 220

His Cys His Val Ser Ser Leu Ser Ala Lys Lys Ile Pro Ser Lys Lys
225                 230                 235                 240

Gln Arg His Lys Lys Thr Ser Ser Ser Ser Ser Ser Ser Ser Arg Ser
                245                 250                 255

Ser Ser Ser Ser Ser Arg Glu Thr Ser Ser Ser Ser Ser Thr Asn
            260                 265                 270
```

<210> 889
<211> 267
<212> PRT
<213> Gammatorquevirus sp.

<400> 889

```
Met Val Ser Ile Ser Ser Ser Asp Phe Phe Lys Lys Thr Lys Phe Asn
1               5                   10                  15

Glu Glu Thr Gln Asn Gln Val Trp Met Ser Gln Ile Ala Asp Ser His
            20                  25                  30

Asp Asn Ile Cys Ser Cys Trp His Pro Phe Ala His Leu Leu Ala Ser
        35                  40                  45

Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr Ile Asn Gln Ile
    50                  55                  60
```

```
Leu Leu Arg Asp Tyr Lys Glu Lys Cys His Ser Gly Gly Glu Glu Gly
65                  70                  75                  80

Glu Asn Ser Gly Pro Thr Thr Gly Leu Ile Thr Pro Lys Glu Glu Asp
                85                  90                  95

Ile Glu Lys Asp Gly Pro Glu Gly Ala Ala Glu Glu Asp His Thr Asp
            100                 105                 110

Ala Leu Phe Ala Ala Ala Val Glu Asn Phe Glu Arg Ser Ala Ser Asn
        115                 120                 125

Phe Arg Gly Arg His Ser Gln Lys Glu Lys Glu Asn Trp Thr Pro Thr
    130                 135                 140

His Ser Pro Thr Thr Lys Arg Arg Gly Asp Thr Val Met Ser Pro Leu
145                 150                 155                 160

Ser Leu Gln Lys Arg Tyr Leu Pro Arg Asn Arg Asp Thr Arg Arg Pro
                165                 170                 175

Pro Ala Ala His Gln Ala Ala Ala Gly Ala Ala Ala Pro Pro Gln Glu
            180                 185                 190

Lys His Pro Pro Ala His Pro Gln Thr Lys Arg Glu Ser Thr Asn Ala
        195                 200                 205

Ser Ala Ser His Arg His Val Thr Leu Thr Arg Phe Lys Pro Gly Phe
    210                 215                 220

Glu Glu Gln Thr Glu Arg Glu Leu Ala Ile Ile Phe His Arg Pro Pro
225                 230                 235                 240

Arg Thr Tyr Lys Glu Asp Leu Pro Phe Tyr Pro Trp Leu Pro Pro Ala
                245                 250                 255

Pro Leu Val Gln Phe Asn Leu Asn Phe Lys Gly
            260                 265
```

<210> 890
<211> 50
<212> PRT
<213> Gammatorquevirus sp.

<400> 890
Met Arg Arg Arg Arg Thr Lys Tyr Gly Cys Leu Lys Leu Leu Thr Leu
1   5   10   15

Met Ile Ile Ser Ala Val Ala Gly Ile His Leu Leu Thr Phe Leu Leu
            20   25   30

Pro Tyr Phe Leu Leu Ala Thr Lys Ile Val Ile Leu Leu Leu Thr Lys
        35   40   45

Phe Phe
    50

<210> 891
<211> 662
<212> PRT
<213> Gammatorquevirus sp.

<400> 891
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1   5   10   15

Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
            20   25   30

Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35   40   45

Lys Leu Arg Lys Val Lys Arg Lys Lys Lys Ser Leu Ile Val Arg Gln
    50   55   60

Trp Gln Pro Asp Ser Ile Arg Thr Cys Lys Ile Ile Gly Gln Ser Ala
65   70   75   80

```
Ile Val Val Gly Ala Glu Gly Lys Gln Met Tyr Cys Tyr Thr Val Asn
                85                  90                  95

Lys Leu Ile Asn Val Pro Pro Lys Thr Pro Tyr Gly Gly Gly Phe Gly
            100                 105                 110

Val Asp Gln Tyr Thr Leu Lys Tyr Leu Tyr Glu Glu Tyr Arg Phe Ala
        115                 120                 125

Gln Asn Ile Trp Thr Gln Ser Asn Val Leu Lys Asp Leu Cys Arg Tyr
    130                 135                 140

Ile Asn Val Lys Leu Ile Phe Tyr Arg Asp Asn Lys Thr Asp Phe Val
145                 150                 155                 160

Leu Ser Tyr Asp Arg Asn Pro Pro Phe Gln Leu Thr Lys Phe Thr Tyr
                165                 170                 175

Pro Gly Ala His Pro Gln Gln Ile Met Leu Gln Lys His His Lys Phe
            180                 185                 190

Ile Leu Ser Gln Met Thr Lys Pro Asn Gly Arg Leu Thr Lys Lys Leu
        195                 200                 205

Lys Ile Lys Pro Pro Lys Gln Met Leu Ser Lys Trp Phe Phe Ser Lys
    210                 215                 220

Gln Phe Cys Lys Tyr Pro Leu Leu Ser Leu Lys Ala Ser Ala Leu Asp
225                 230                 235                 240

Leu Arg His Ser Tyr Leu Gly Cys Cys Asn Glu Asn Pro Gln Val Phe
                245                 250                 255

Phe Tyr Tyr Leu Asn His Gly Tyr Tyr Thr Ile Thr Asn Trp Gly Ala
            260                 265                 270

Gln Ser Ser Thr Ala Tyr Arg Pro Asn Ser Lys Val Thr Asp Thr Thr
        275                 280                 285
```

```
Tyr Tyr Arg Tyr Lys Asn Asp Arg Lys Asn Ile Asn Ile Lys Ser His
    290                 295                 300

Glu Tyr Glu Lys Ser Ile Ser Tyr Glu Asn Gly Tyr Phe Gln Ser Ser
305                 310                 315                 320

Phe Leu Gln Thr Gln Cys Ile Tyr Thr Ser Glu Arg Gly Glu Ala Cys
                325                 330                 335

Ile Ala Glu Lys Pro Leu Gly Ile Ala Ile Tyr Asn Pro Val Lys Asp
            340                 345                 350

Asn Gly Asp Gly Asn Met Ile Tyr Leu Val Ser Thr Leu Ala Asn Thr
        355                 360                 365

Trp Asp Gln Pro Pro Lys Asp Ser Ala Ile Leu Ile Gln Gly Val Pro
    370                 375                 380

Ile Trp Leu Gly Leu Phe Gly Tyr Leu Asp Tyr Cys Arg Gln Ile Lys
385                 390                 395                 400

Ala Asp Lys Thr Trp Leu Asp Ser His Val Leu Val Ile Gln Ser Pro
                405                 410                 415

Ala Ile Phe Thr Tyr Pro Asn Pro Gly Ala Gly Lys Trp Tyr Cys Pro
            420                 425                 430

Leu Ser Gln Ser Phe Ile Asn Gly Asn Gly Pro Phe Asn Gln Pro Pro
        435                 440                 445

Thr Leu Leu Gln Lys Ala Lys Trp Phe Pro Gln Ile Gln Tyr Gln Gln
    450                 455                 460

Glu Ile Ile Asn Ser Phe Val Glu Ser Gly Pro Phe Val Pro Lys Tyr
465                 470                 475                 480
```

```
Ala Asn Gln Thr Glu Ser Asn Trp Glu Leu Lys Tyr Lys Tyr Val Phe
                485                 490                 495

Thr Phe Lys Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala Asp
            500                 505                 510

Pro Ser Lys Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln Thr
        515                 520                 525

Ile Gln Ile Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile His
    530                 535                 540

Asp Trp Asp Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys Arg
545                 550                 555                 560

Met Ser Thr Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser Glu
                565                 570                 575

Glu Asp Ile Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr Val
            580                 585                 590

Pro Gln Gln Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu Cys
        595                 600                 605

Lys Lys Asp Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln Gln
    610                 615                 620

Leu Ile Lys Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn Ile
625                 630                 635                 640

Leu Gln Leu Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln Leu
                645                 650                 655

His Thr Gly Met Leu Pro
            660
```

<210> 892
<211> 215

<212> PRT
<213> Gammatorquevirus sp.

<400> 892
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1   5   10   15

Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
            20   25   30

Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35   40   45

Lys Leu Arg Lys Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala
    50   55   60

Asp Pro Ser Lys Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln
65   70   75   80

Thr Ile Gln Ile Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile
                85   90   95

His Asp Trp Asp Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys
            100   105   110

Arg Met Ser Thr Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser
        115   120   125

Glu Glu Asp Ile Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr
    130   135   140

Val Pro Gln Gln Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu
145   150   155   160

Cys Lys Lys Asp Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln
                165   170   175

Gln Leu Ile Lys Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn
            180   185   190

```
Ile Leu Gln Leu Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln
        195                 200                 205


Leu His Thr Gly Met Leu Pro
    210                 215
```

<210> 893
<211> 129
<212> PRT
<213> Gammatorquevirus sp.

<400> 893

```
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1               5                   10                  15


Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
            20                  25                  30


Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35                  40                  45


Lys Leu Arg Lys Ile Ser Lys Gln Leu Gln Arg Lys Thr Phe Pro Lys
    50                  55                  60


Arg Lys Arg Glu Leu Asp Pro Asn Ser Gln Ser His Asn Lys Lys Lys
65                  70                  75                  80


Arg Arg His Cys His Val Ser Ser Leu Ser Ala Lys Lys Ile Pro Ser
                85                  90                  95


Lys Lys Gln Arg His Lys Lys Thr Ser Ser Ser Ser Ser Ser Ser Ser
            100                 105                 110


Arg Ser Ser Ser Ser Ser Ser Arg Glu Thr Ser Ser Ser Ser Ser Thr
        115                 120                 125


Asn
```

<210> 894
<211> 3696
<212> DNA
<213> Alphatorquevirus sp.

<400> 894
attttgttca gcccgccaat ttctctttca aacaggccaa tcagctacta cttcgtgcac 60

ttcctggggc gtgtcctgcc gctctatata agcagaggcg gtgacgaatg gtagagtttt 120

tcttggcccg tccgcggcga gagcgcgagc gaagcgagcg atcgagcgtc ccgagggcgg 180

gtgccggagg tgagtttaca caccgcagtc aaggggcaat tcgggctcgg gactggccgg 240

gctatgggca agattcttaa aaaattcccc cgatcccttt gccgccagga cataaaaaca 300

tgccgtggag accgccggtc catagtgtcc aggggcgaga ggatcagtgg ttcgcaagct 360

tttttcacgg ccacgattcg ttttgcggct gcggtgaccc tcttggccat attaatagca 420

ttgctcatcg ctttcctcgc gccggtccac caaggccccc tccggggcta gatcagccta 480

acccccggga gcagggcccg gccggacccg gagggccgcc cgccatcttg gccctgccgg 540

ctccgcccgc ggagcctgac gacccgcagc cacggcgtgg tggtggggac ggtggcgccg 600

ccgctggcgc cgcagacgac catacacaac gagactacga cgaagaagag ctagacgagc 660

ttttccgcgc cgccgccgaa gacgatttgt aagtaggaga tggcgccggc cttacaggcg 720

caggaggaga cgcgggcgac gcagacgcag acgcagacgc agacataagc ccaccctaat 780

actcagacag tggcaacctg actgtatcag acactgtaaa ataacaggat ggatgcccct 840

cattatctgt ggaaaggggt ccacccagtt caactacatc acccacgcgg acgatatcac 900

ccccagggga gcctcctacg gaggcaattt cacaaacatg actttctccc tggaggccat 960

atatgaacag ttcctatacc acagaaacag gtggtcggcc tctaaccacg acctagaact 1020

gtgcagatac aaggggacca ccttaaaact ctacagacac ccagaagtag actacatagt 1080

tacctacagc agaacaggac cctttgaaat cagccacatg acctacctca gcactcaccc 1140

catgctaatg ctgctaaaca agcaccacat tgtggtgccc agcttaaaga ctaagcccag 1200

aggcagaaag gccataaaag tcaggataag gcccccaaaa ctcatgaaca acaagtggta 1260

cttcaccaga gacttctgta acataggcct cttccagctc tgggccacag gcttagaact 1320

cagaaacccc tggctcagaa tgagcaccct gagcccctgc ataggcttta atgtcctcaa 1380

aaacagcatt tacacaaacc tcagcaacct gccacaatac aaaaacgaaa gactaaacat 1440

cattaacaac atacttcacc cacaagaaat tacaggtaca aacaacaaaa agtggcagta 1500

cacatacacc aaactcatgg cccctattta ctattcagca aacagggcca gcacctatga 1560

ctgggaaaat tacagcaaag aaacaaacta caataataca tatgttaaat ttacccagaa 1620

aagacaggaa aaactaacta aaattagaaa agagtggcag atgctttatc cacaacaacc 1680

cacagcactg ccagactcct atgacctcct acaagagtat ggcctctaca gtccatacta 1740

cctaaacccc acaagaataa acctagactg gatgacccca tacacacacg tcagatacaa 1800

tcccctagta gacaagggct ttggaaacag aatatacatc cagtggtgct cagaagcaga 1860

tgttagctac aacaggacaa aatccaagtg tctgctacaa gacatgcccc tgtttttcat 1920

gtgctatggc tacatagact gggcaataaa aaacactgga gtgtcatctc tagtgaagga 1980

cgccagaatc tgcatcaggt gtccctacac agagccacaa ctagttggct ccacagaaga 2040

cataggcttt gtacccatct cagaaacctt catgaggggc gacatgccgg tacttgcacc 2100

atacataccg ttaagctggt tttgcaagtg gtatcccaac atagctcacc aaaaggaagt 2160

ccttgagtca atcatttcct gcagcccctt catgccccgt gaccaagaca tgaacggttg 2220

ggatatcaca atcggttaca aaatggactt cttatggggc ggttcccctc tcccctcaca 2280

gccaatcgac gacccctgcc agcagggaac ccacccgatt cccgaccccg ataaacaccc 2340

tcgcctccta caagtctcga acccgaaact actcggaccg aggacagtgt tccacaagtg 2400

ggacatcaga cgtgggcagt ttagcaaaag aagtattaag agagtgtcag aatactcaag 2460

cgatgatgaa tctcttgcgc caggtctccc atcaaagcga aacaagctcg actcggcgtt 2520

ccgaggagaa aatcgagagc aaaaagaatg ctattctctc ctcaaagcgc tcgaggaaga 2580

agagacccca gaagaagaag aaccagcacc ccaagaaaaa gcccagaaag aggagctact 2640

ccaccagctc cagctccaga gacgccacca gcgagtcctc agacgagggc tcaagctcgt 2700

ctttacagac atcctccgac tccgccaggg agtccactgg aacccggagc tcacatagcg 2760

```
cccccacctt acataccaga cctgcttttt cccaatactg gtaaaaaaaa aaaattctct     2820

cccttcgatt gggagacaga ggcgcaaata gcggggtgga tgcggcggcc catgcgcttc     2880

tatccctcag acacccctca ctacccgtgg ctacccccgg agcgagatat cccgaaaata     2940

tgtaacataa acttcaaaat aaagcttcaa gagtgagtga ttcgaggccc tcctctgttc     3000

acttagcggt gtctacctct taaggtcact aagcactccg agcgtaagcg aggagtgcga     3060

ccctctacca aggggcaact tcctcggggt ccggcgctac gcgcttcgcg ctgcgccgga     3120

catctcggac ccctcgaccc gaatcgcttg cgcgattcgg acctgcggcc tcgggggggt     3180

cgggggcttt actaaacaga ctccgaggtg ccattggaca ctgtaggggg tgaacagcaa     3240

cgaaagtgag tggggccaga cttcgccata aggcctttat cttcttgcca ttggatagtg     3300

acttccgggt ccgcctgggg gccgccattt tagcttcggc cgccatttta ggccctcgcg     3360

ggcctccgta ggcgcgcttt agtgacgtca cggcagccat tttgtcgtga cgtttgagac     3420

acgtgatggg ggcgtgccta aacccggaag catccctggt cacgtgactc tgacgtcacg     3480

gcggccatct tgtgctgtcc gccatcttgt aacttccttc cgctttttca aaaaaaaaga     3540

ggaagtgtga cgtagcggcg ggggggcggc gcgcttcgcg cgccgcccac cagggggcgc     3600

tgcgcgcccc ccgcgcatgc gcaggggcct ctcgaggggc tccgcccccc ccccgtgcta     3660

aatttaccgc gcatgcgcga ccacgccccc gccgcc                               3696
```

<210> 895
<211> 130
<212> PRT
<213> Alphatorquevirus sp.

<400> 895

```
Met Pro Trp Arg Pro Pro Val His Ser Val Gln Gly Arg Glu Asp Gln
1               5                  10                  15


Trp Phe Ala Ser Phe Phe His Gly His Asp Ser Phe Cys Gly Cys Gly
            20                  25                  30


Asp Pro Leu Gly His Ile Asn Ser Ile Ala His Arg Phe Pro Arg Ala
        35                  40                  45
```

```
Gly Pro Pro Arg Pro Pro Pro Gly Leu Asp Gln Pro Asn Pro Arg Glu
    50                  55                  60

Gln Gly Pro Ala Gly Pro Gly Gly Pro Pro Ala Ile Leu Ala Leu Pro
65                  70                  75                  80

Ala Pro Pro Ala Glu Pro Asp Asp Pro Gln Pro Arg Arg Gly Gly Gly
                85                  90                  95

Asp Gly Gly Ala Ala Ala Gly Ala Ala Asp Asp His Thr Gln Arg Asp
            100                 105                 110

Tyr Asp Glu Glu Glu Leu Asp Glu Leu Phe Arg Ala Ala Ala Glu Asp
        115                 120                 125

Asp Leu
    130
```

<210> 896
<211> 303
<212> PRT
<213> Alphatorquevirus sp.

<400> 896

```
Met Pro Trp Arg Pro Pro Val His Ser Val Gln Gly Arg Glu Asp Gln
1               5                   10                  15

Trp Phe Ala Ser Phe Phe His Gly His Asp Ser Phe Cys Gly Cys Gly
            20                  25                  30

Asp Pro Leu Gly His Ile Asn Ser Ile Ala His Arg Phe Pro Arg Ala
        35                  40                  45

Gly Pro Pro Arg Pro Pro Pro Gly Leu Asp Gln Pro Asn Pro Arg Glu
    50                  55                  60

Gln Gly Pro Ala Gly Pro Gly Gly Pro Pro Ala Ile Leu Ala Leu Pro
65                  70                  75                  80
```

```
Ala Pro Pro Ala Glu Pro Asp Asp Pro Gln Pro Arg Arg Gly Gly Gly
                85                  90                  95

Asp Gly Gly Ala Ala Ala Gly Ala Ala Asp Asp His Thr Gln Arg Asp
            100                 105                 110

Tyr Asp Glu Glu Glu Leu Asp Glu Leu Phe Arg Ala Ala Ala Glu Asp
        115                 120                 125

Asp Phe Gln Ser Thr Thr Pro Ala Ser Arg Glu Pro Thr Arg Phe Pro
    130                 135                 140

Thr Pro Ile Asn Thr Leu Ala Ser Tyr Lys Ser Arg Thr Arg Asn Tyr
145                 150                 155                 160

Ser Asp Arg Gly Gln Cys Ser Thr Ser Gly Thr Ser Asp Val Gly Ser
                165                 170                 175

Leu Ala Lys Glu Val Leu Arg Glu Cys Gln Asn Thr Gln Ala Met Met
            180                 185                 190

Asn Leu Leu Arg Gln Val Ser His Gln Ser Glu Thr Ser Ser Thr Arg
        195                 200                 205

Arg Ser Glu Glu Lys Ile Glu Ser Lys Lys Asn Ala Ile Leu Ser Ser
    210                 215                 220

Lys Arg Ser Arg Lys Lys Arg Pro Gln Lys Lys Lys Asn Gln His Pro
225                 230                 235                 240

Lys Lys Lys Pro Arg Lys Arg Ser Tyr Ser Thr Ser Ser Ser Ser Arg
                245                 250                 255

Asp Ala Thr Ser Glu Ser Ser Asp Glu Gly Ser Ser Ser Ser Leu Gln
            260                 265                 270

Thr Ser Ser Asp Ser Ala Arg Glu Ser Thr Gly Thr Arg Ser Ser His
```

275 280 285

Ser Ala Pro Thr Leu His Thr Arg Pro Ala Phe Ser Gln Tyr Trp
290 295 300

<210> 897
<211> 292
<212> PRT
<213> Alphatorquevirus sp.

<400> 897
Met Pro Trp Arg Pro Pro Val His Ser Val Gln Gly Arg Glu Asp Gln
1 5 10 15

Trp Phe Ala Ser Phe Phe His Gly His Asp Ser Phe Cys Gly Cys Gly
20 25 30

Asp Pro Leu Gly His Ile Asn Ser Ile Ala His Arg Phe Pro Arg Ala
35 40 45

Gly Pro Pro Arg Pro Pro Pro Gly Leu Asp Gln Pro Asn Pro Arg Glu
50 55 60

Gln Gly Pro Ala Gly Pro Gly Gly Pro Pro Ala Ile Leu Ala Leu Pro
65 70 75 80

Ala Pro Pro Ala Glu Pro Asp Asp Pro Gln Pro Arg Arg Gly Gly Gly
85 90 95

Asp Gly Gly Ala Ala Ala Gly Ala Ala Asp Asp His Thr Gln Arg Asp
100 105 110

Tyr Asp Glu Glu Glu Leu Asp Glu Leu Phe Arg Ala Ala Ala Glu Asp
115 120 125

Asp Leu Ser Pro Ile Lys Ala Lys Gln Ala Arg Leu Gly Val Pro Arg
130 135 140

Arg Lys Ser Arg Ala Lys Arg Met Leu Phe Ser Pro Gln Ser Ala Arg

```
145                 150                 155                 160

Gly Arg Arg Asp Pro Arg Arg Arg Arg Thr Ser Thr Pro Arg Lys Ser
                165                 170                 175

Pro Glu Arg Gly Ala Thr Pro Pro Ala Pro Ala Pro Glu Thr Pro Pro
            180                 185                 190

Ala Ser Pro Gln Thr Arg Ala Gln Ala Arg Leu Tyr Arg His Pro Pro
        195                 200                 205

Thr Pro Pro Gly Ser Pro Leu Glu Pro Gly Ala His Ile Ala Pro Pro
    210                 215                 220

Pro Tyr Ile Pro Asp Leu Leu Phe Pro Asn Thr Gly Lys Lys Lys Lys
225                 230                 235                 240

Phe Ser Pro Phe Asp Trp Glu Thr Glu Ala Gln Ile Ala Gly Trp Met
                245                 250                 255

Arg Arg Pro Met Arg Phe Tyr Pro Ser Asp Thr Pro His Tyr Pro Trp
            260                 265                 270

Leu Pro Pro Glu Arg Asp Ile Pro Lys Ile Cys Asn Ile Asn Phe Lys
        275                 280                 285

Ile Lys Leu Gln
    290
```

<210> 898
<211> 180
<212> PRT
<213> Alphatorquevirus sp.

<400> 898

```
Met Pro Trp Arg Pro Pro Val His Ser Val Gln Gly Arg Glu Asp Gln
1               5                   10                  15

Trp Ser Pro Ile Lys Ala Lys Gln Ala Arg Leu Gly Val Pro Arg Arg
```

```
            20                  25                  30

Lys Ser Arg Ala Lys Arg Met Leu Phe Ser Pro Gln Ser Ala Arg Gly
        35                  40                  45

Arg Arg Asp Pro Arg Arg Arg Arg Thr Ser Thr Pro Arg Lys Ser Pro
    50                  55                  60

Glu Arg Gly Ala Thr Pro Pro Ala Pro Ala Pro Glu Thr Pro Pro Ala
65                  70                  75                  80

Ser Pro Gln Thr Arg Ala Gln Ala Arg Leu Tyr Arg His Pro Pro Thr
                85                  90                  95

Pro Pro Gly Ser Pro Leu Glu Pro Gly Ala His Ile Ala Pro Pro Pro
            100                 105                 110

Tyr Ile Pro Asp Leu Leu Phe Pro Asn Thr Gly Lys Lys Lys Lys Phe
        115                 120                 125

Ser Pro Phe Asp Trp Glu Thr Glu Ala Gln Ile Ala Gly Trp Met Arg
    130                 135                 140

Arg Pro Met Arg Phe Tyr Pro Ser Asp Thr Pro His Tyr Pro Trp Leu
145                 150                 155                 160

Pro Pro Glu Arg Asp Ile Pro Lys Ile Cys Asn Ile Asn Phe Lys Ile
                165                 170                 175

Lys Leu Gln Glu
            180
```

<210> 899
<211> 49
<212> PRT
<213> Alphatorquevirus sp.

<400> 899
Ile Val Ser Arg Gly Glu Arg Ile Ser Gly Ser Gln Ala Phe Phe Thr

1 5 10 15

Ala Thr Ile Arg Phe Ala Ala Ala Val Thr Leu Leu Ala Ile Leu Ile
20 25 30

Ala Leu Leu Ile Ala Phe Leu Ala Pro Val His Gln Gly Pro Leu Arg
35 40 45

Gly

<210> 900
<211> 728
<212> PRT
<213> Alphatorquevirus sp.

<400> 900
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1 5 10 15

Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro
20 25 30

Arg Arg Arg Arg Arg Arg Phe Val Ser Arg Arg Trp Arg Arg Pro Tyr
35 40 45

Arg Arg Arg Arg Arg Arg Gly Arg Arg Arg Arg Arg Arg Arg Arg Arg
50 55 60

His Lys Pro Thr Leu Ile Leu Arg Gln Trp Gln Pro Asp Cys Ile Arg
65 70 75 80

His Cys Lys Ile Thr Gly Trp Met Pro Leu Ile Ile Cys Gly Lys Gly
85 90 95

Ser Thr Gln Phe Asn Tyr Ile Thr His Ala Asp Asp Ile Thr Pro Arg
100 105 110

Gly Ala Ser Tyr Gly Gly Asn Phe Thr Asn Met Thr Phe Ser Leu Glu

```
        115                 120                 125

Ala Ile Tyr Glu Gln Phe Leu Tyr His Arg Asn Arg Trp Ser Ala Ser
    130                 135                 140

Asn His Asp Leu Glu Leu Cys Arg Tyr Lys Gly Thr Thr Leu Lys Leu
145                 150                 155                 160

Tyr Arg His Pro Glu Val Asp Tyr Ile Val Thr Tyr Ser Arg Thr Gly
                165                 170                 175

Pro Phe Glu Ile Ser His Met Thr Tyr Leu Ser Thr His Pro Met Leu
            180                 185                 190

Met Leu Leu Asn Lys His His Ile Val Val Pro Ser Leu Lys Thr Lys
        195                 200                 205

Pro Arg Gly Arg Lys Ala Ile Lys Val Arg Ile Arg Pro Pro Lys Leu
    210                 215                 220

Met Asn Asn Lys Trp Tyr Phe Thr Arg Asp Phe Cys Asn Ile Gly Leu
225                 230                 235                 240

Phe Gln Leu Trp Ala Thr Gly Leu Glu Leu Arg Asn Pro Trp Leu Arg
                245                 250                 255

Met Ser Thr Leu Ser Pro Cys Ile Gly Phe Asn Val Leu Lys Asn Ser
            260                 265                 270

Ile Tyr Thr Asn Leu Ser Asn Leu Pro Gln Tyr Lys Asn Glu Arg Leu
        275                 280                 285

Asn Ile Ile Asn Asn Ile Leu His Pro Gln Glu Ile Thr Gly Thr Asn
    290                 295                 300

Asn Lys Lys Trp Gln Tyr Thr Tyr Thr Lys Leu Met Ala Pro Ile Tyr
305                 310                 315                 320
```

```
Tyr Ser Ala Asn Arg Ala Ser Thr Tyr Asp Trp Glu Asn Tyr Ser Lys
                325                 330                 335

Glu Thr Asn Tyr Asn Asn Thr Tyr Val Lys Phe Thr Gln Lys Arg Gln
            340                 345                 350

Glu Lys Leu Thr Lys Ile Arg Lys Glu Trp Gln Met Leu Tyr Pro Gln
        355                 360                 365

Gln Pro Thr Ala Leu Pro Asp Ser Tyr Asp Leu Leu Gln Glu Tyr Gly
    370                 375                 380

Leu Tyr Ser Pro Tyr Tyr Leu Asn Pro Thr Arg Ile Asn Leu Asp Trp
385                 390                 395                 400

Met Thr Pro Tyr Thr His Val Arg Tyr Asn Pro Leu Val Asp Lys Gly
                405                 410                 415

Phe Gly Asn Arg Ile Tyr Ile Gln Trp Cys Ser Glu Ala Asp Val Ser
            420                 425                 430

Tyr Asn Arg Thr Lys Ser Lys Cys Leu Leu Gln Asp Met Pro Leu Phe
        435                 440                 445

Phe Met Cys Tyr Gly Tyr Ile Asp Trp Ala Ile Lys Asn Thr Gly Val
    450                 455                 460

Ser Ser Leu Val Lys Asp Ala Arg Ile Cys Ile Arg Cys Pro Tyr Thr
465                 470                 475                 480

Glu Pro Gln Leu Val Gly Ser Thr Glu Asp Ile Gly Phe Val Pro Ile
                485                 490                 495

Ser Glu Thr Phe Met Arg Gly Asp Met Pro Val Leu Ala Pro Tyr Ile
            500                 505                 510

Pro Leu Ser Trp Phe Cys Lys Trp Tyr Pro Asn Ile Ala His Gln Lys
```

```
        515                 520                 525

Glu Val Leu Glu Ser Ile Ile Ser Cys Ser Pro Phe Met Pro Arg Asp
    530                 535                 540

Gln Asp Met Asn Gly Trp Asp Ile Thr Ile Gly Tyr Lys Met Asp Phe
545                 550                 555                 560

Leu Trp Gly Gly Ser Pro Leu Pro Ser Gln Pro Ile Asp Asp Pro Cys
                565                 570                 575

Gln Gln Gly Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu
            580                 585                 590

Leu Gln Val Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His
        595                 600                 605

Lys Trp Asp Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg
    610                 615                 620

Val Ser Glu Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro
625                 630                 635                 640

Ser Lys Arg Asn Lys Leu Asp Ser Ala Phe Arg Gly Glu Asn Arg Glu
                645                 650                 655

Gln Lys Glu Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr
            660                 665                 670

Pro Glu Glu Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu
        675                 680                 685

Leu Leu His Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg
    690                 695                 700

Arg Gly Leu Lys Leu Val Phe Thr Asp Ile Leu Arg Leu Arg Gln Gly
705                 710                 715                 720
```

Val His Trp Asn Pro Glu Leu Thr
725

<210> 901
<211> 197
<212> PRT
<213> Alphatorquevirus sp.

<400> 901
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1 5 10 15

Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro
20 25 30

Arg Arg Arg Arg Arg Arg Phe Pro Ile Asp Asp Pro Cys Gln Gln Gly
35 40 45

Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu Leu Gln Val
50 55 60

Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His Lys Trp Asp
65 70 75 80

Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg Val Ser Glu
85 90 95

Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro Ser Lys Arg
100 105 110

Asn Lys Leu Asp Ser Ala Phe Arg Gly Glu Asn Arg Glu Gln Lys Glu
115 120 125

Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr Pro Glu Glu
130 135 140

Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu Leu Leu His
145 150 155 160

```
Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg Arg Gly Leu
                165                 170                 175


Lys Leu Val Phe Thr Asp Ile Leu Arg Leu Arg Gln Gly Val His Trp
            180                 185                 190


Asn Pro Glu Leu Thr
        195


<210> 902
<211> 145
<212> PRT
<213> Alphatorquevirus sp.

<400> 902
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1               5                   10                  15


Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro
            20                  25                  30


Arg Arg Arg Arg Arg Arg Phe Val Ser His Gln Ser Glu Thr Ser Ser
        35                  40                  45


Thr Arg Arg Ser Glu Glu Lys Ile Glu Ser Lys Lys Asn Ala Ile Leu
    50                  55                  60


Ser Ser Lys Arg Ser Arg Lys Lys Arg Pro Gln Lys Lys Lys Asn Gln
65                  70                  75                  80


His Pro Lys Lys Lys Pro Arg Lys Arg Ser Tyr Ser Thr Ser Ser Ser
                85                  90                  95


Ser Arg Asp Ala Thr Ser Glu Ser Ser Asp Glu Gly Ser Ser Ser Ser
            100                 105                 110


Leu Gln Thr Ser Ser Asp Ser Ala Arg Glu Ser Thr Gly Thr Arg Ser
        115                 120                 125
```

Ser His Ser Ala Pro Thr Leu His Thr Arg Pro Ala Phe Ser Gln Tyr
130 135 140

Trp
145

<210> 903

<400> 903
000

<210> 904

<400> 904
000

<210> 905

<400> 905
000

<210> 906

<400> 906
000

<210> 907

<400> 907
000

<210> 908

<400> 908
000

<210> 909

<400> 909
000

<210> 910

<400> 910
000

<210> 911

<400> 911
000

<210> 912

<400> 912
000

<210> 913

<400> 913
000

<210> 914

<400> 914
000

<210> 915

<400> 915
000

<210> 916

<400> 916
000

<210> 917

<400> 917
000

<210> 918

<400> 918
000

<210> 919
<211> 677
<212> PRT
<213> Gammatorquevirus sp.

<400> 919
Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg
1   5   10   15

Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg
            20   25   30

Ile Pro Arg Arg Arg Asn Arg Arg Phe Thr Lys Thr Arg Arg Arg Arg
        35   40   45

Lys Arg Lys Lys Val Arg Arg Lys Leu Lys Lys Ile Thr Ile Lys Gln
    50   55   60

Trp Gln Pro Asp Ser Val Lys Lys Cys Lys Ile Lys Gly Tyr Ser Thr
65   70   75   80

Leu Val Met Gly Ala Gln Gly Lys Gln Tyr Asn Cys Tyr Thr Asn Gln
                85   90   95

Ala Ser Asp Tyr Val Gln Pro Lys Ala Pro Gln Gly Gly Gly Phe Gly
            100   105   110

Cys Glu Val Phe Asn Leu Lys Trp Leu Tyr Gln Glu Tyr Thr Ala His
        115   120   125

Arg Asn Ile Trp Thr Lys Thr Asn Glu Tyr Thr Asp Leu Cys Arg Tyr
    130   135   140

Thr Gly Ala Gln Ile Ile Leu Tyr Arg His Pro Asp Val Asp Phe Ile
145   150   155   160

```
Val Ser Trp Asp Asn Gln Pro Pro Phe Leu Leu Asn Lys Tyr Thr Tyr
                165                 170                 175

Pro Glu Leu Gln Pro Gln Asn Leu Leu Leu Ala Arg Arg Lys Arg Ile
            180                 185                 190

Ile Leu Ser Gln Lys Ser Asn Pro Lys Gly Lys Leu Arg Ile Lys Leu
        195                 200                 205

Arg Ile Pro Pro Pro Lys Gln Met Ile Thr Lys Trp Phe Phe Gln Arg
    210                 215                 220

Asp Phe Cys Asp Val Asn Leu Phe Lys Leu Cys Ala Ser Ala Ala Ser
225                 230                 235                 240

Phe Arg Tyr Pro Gly Ile Ser His Gly Ala Gln Ser Thr Ile Phe Ser
                245                 250                 255

Ala Tyr Ala Leu Asn Thr Asp Phe Tyr Gln Cys Ser Asp Trp Cys Gln
            260                 265                 270

Thr Asn Thr Glu Thr Gly Tyr Leu Asn Ile Lys Thr Gln Gln Met Pro
        275                 280                 285

Leu Trp Phe His Tyr Arg Glu Gly Gly Lys Glu Lys Trp Tyr Lys Tyr
    290                 295                 300

Thr Asn Lys Glu His Arg Pro Tyr Thr Asn Thr Tyr Leu Lys Ser Ile
305                 310                 315                 320

Ser Tyr Asn Asp Gly Leu Phe Ser Pro Lys Ala Met Phe Ala Phe Glu
                325                 330                 335

Val Lys Ala Gly Gly Glu Gly Thr Thr Glu Pro Pro Gln Gly Ala Gln
            340                 345                 350

Leu Ile Ala Asn Leu Pro Leu Ile Ala Leu Arg Tyr Asn Pro His Glu
```

```
        355                 360                 365

Asp Thr Gly His Gly Asn Glu Ile Tyr Leu Thr Ser Thr Phe Lys Gly
    370                 375                 380

Thr Tyr Asp Lys Pro Lys Val Thr Asp Ala Leu Tyr Phe Asn Asn Val
385                 390                 395                 400

Pro Leu Trp Met Gly Phe Tyr Gly Tyr Trp Asp Phe Ile Leu Gln Glu
                405                 410                 415

Thr Lys Asn Lys Gly Val Phe Asp Gln His Met Phe Val Val Lys Cys
            420                 425                 430

Pro Ala Leu Arg Pro Ile Ser Gln Val Thr Lys Gln Val Tyr Tyr Pro
        435                 440                 445

Leu Val Asp Met Asp Phe Cys Ser Gly Arg Leu Pro Phe Asp Glu Tyr
    450                 455                 460

Leu Ser Lys Asp Ile Lys Ser His Trp Tyr Pro Thr Ala Glu Arg Gln
465                 470                 475                 480

Thr Val Thr Ile Asn Asn Phe Val Thr Ala Gly Pro Tyr Met Pro Lys
                485                 490                 495

Phe Glu Pro Thr Asp Lys Asp Ser Thr Trp Gln Leu Asn Tyr His Tyr
            500                 505                 510

Lys Phe Phe Phe Lys Trp Gly Gly Pro Gln Val Thr Asp Pro Thr Val
        515                 520                 525

Glu Asp Pro Cys Ser Arg Asn Lys Tyr Pro Val Pro Asp Thr Met Gln
    530                 535                 540

Gln Thr Ile Gln Ile Lys Asn Pro Glu Lys Leu His Pro Ala Thr Leu
545                 550                 555                 560
```

```
Phe His Asp Trp Asp Leu Arg Arg Gly Phe Ile Thr Gln Ala Ala Ile
                565                 570                 575


Lys Arg Met Ser Glu Asn Leu Gln Ile Asp Ser Ser Phe Glu Ser Asp
            580                 585                 590


Gly Thr Glu Ser Pro Lys Lys Lys Lys Arg Cys Thr Lys Glu Ile Pro
        595                 600                 605


Thr Gln Asn Gln Lys Gln Glu Glu Ile Gln Glu Cys Leu Leu Ser Leu
    610                 615                 620


Cys Glu Glu Pro Thr Cys Gln Glu Glu Thr Glu Asp Leu Gln Leu Phe
625                 630                 635                 640


Ile Gln Gln Gln Gln Gln Gln Gln Tyr Lys Leu Arg Lys Asn Leu Phe
                645                 650                 655


Lys Leu Leu Thr His Leu Lys Lys Gly Gln Arg Ile Ser Gln Leu Gln
            660                 665                 670


Thr Gly Leu Leu Glu
        675


<210> 920
<211> 59
<212> PRT
<213> Gammatorquevirus sp.

<400> 920
Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg
1               5                   10                  15


Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg
            20                  25                  30


Ile Pro Arg Arg Arg Asn Arg Arg Phe Thr Lys Thr Arg Arg Arg Arg
        35                  40                  45
```

Lys Arg Lys Lys Val Arg Arg Lys Leu Lys Lys
50 55

<210> 921
<211> 201
<212> PRT
<213> Gammatorquevirus sp.

<400> 921
Ile Thr Ile Lys Gln Trp Gln Pro Asp Ser Val Lys Lys Cys Lys Ile
1 5 10 15

Lys Gly Tyr Ser Thr Leu Val Met Gly Ala Gln Gly Lys Gln Tyr Asn
20 25 30

Cys Tyr Thr Asn Gln Ala Ser Asp Tyr Val Gln Pro Lys Ala Pro Gln
35 40 45

Gly Gly Gly Phe Gly Cys Glu Val Phe Asn Leu Lys Trp Leu Tyr Gln
50 55 60

Glu Tyr Thr Ala His Arg Asn Ile Trp Thr Lys Thr Asn Glu Tyr Thr
65 70 75 80

Asp Leu Cys Arg Tyr Thr Gly Ala Gln Ile Ile Leu Tyr Arg His Pro
85 90 95

Asp Val Asp Phe Ile Val Ser Trp Asp Asn Gln Pro Pro Phe Leu Leu
100 105 110

Asn Lys Tyr Thr Tyr Pro Glu Leu Gln Pro Gln Asn Leu Leu Leu Ala
115 120 125

Arg Arg Lys Arg Ile Ile Leu Ser Gln Lys Ser Asn Pro Lys Gly Lys
130 135 140

Leu Arg Ile Lys Leu Arg Ile Pro Pro Pro Lys Gln Met Ile Thr Lys
145 150 155 160

```
Trp Phe Phe Gln Arg Asp Phe Cys Asp Val Asn Leu Phe Lys Leu Cys
                165                 170                 175


Ala Ser Ala Ala Ser Phe Arg Tyr Pro Gly Ile Ser His Gly Ala Gln
            180                 185                 190


Ser Thr Ile Phe Ser Ala Tyr Ala Leu
        195                 200


<210> 922
<211> 96
<212> PRT
<213> Gammatorquevirus sp.

<400> 922
Asn Thr Asp Phe Tyr Gln Cys Ser Asp Trp Cys Gln Thr Asn Thr Glu
1               5                   10                  15


Thr Gly Tyr Leu Asn Ile Lys Thr Gln Gln Met Pro Leu Trp Phe His
            20                  25                  30


Tyr Arg Glu Gly Gly Lys Glu Lys Trp Tyr Lys Tyr Thr Asn Lys Glu
        35                  40                  45


His Arg Pro Tyr Thr Asn Thr Tyr Leu Lys Ser Ile Ser Tyr Asn Asp
    50                  55                  60


Gly Leu Phe Ser Pro Lys Ala Met Phe Ala Phe Glu Val Lys Ala Gly
65                  70                  75                  80


Gly Glu Gly Thr Thr Glu Pro Pro Gln Gly Ala Gln Leu Ile Ala Asn
                85                  90                  95


<210> 923
<211> 161
<212> PRT
<213> Gammatorquevirus sp.

<400> 923
Leu Pro Leu Ile Ala Leu Arg Tyr Asn Pro His Glu Asp Thr Gly His
```

1 5 10 15

Gly Asn Glu Ile Tyr Leu Thr Ser Thr Phe Lys Gly Thr Tyr Asp Lys
20 25 30

Pro Lys Val Thr Asp Ala Leu Tyr Phe Asn Asn Val Pro Leu Trp Met
35 40 45

Gly Phe Tyr Gly Tyr Trp Asp Phe Ile Leu Gln Glu Thr Lys Asn Lys
50 55 60

Gly Val Phe Asp Gln His Met Phe Val Val Lys Cys Pro Ala Leu Arg
65 70 75 80

Pro Ile Ser Gln Val Thr Lys Gln Val Tyr Tyr Pro Leu Val Asp Met
85 90 95

Asp Phe Cys Ser Gly Arg Leu Pro Phe Asp Glu Tyr Leu Ser Lys Asp
100 105 110

Ile Lys Ser His Trp Tyr Pro Thr Ala Glu Arg Gln Thr Val Thr Ile
115 120 125

Asn Asn Phe Val Thr Ala Gly Pro Tyr Met Pro Lys Phe Glu Pro Thr
130 135 140

Asp Lys Asp Ser Thr Trp Gln Leu Asn Tyr His Tyr Lys Phe Phe Phe
145 150 155 160

Lys

<210> 924
<211> 160
<212> PRT
<213> Gammatorquevirus sp.

<400> 924
Trp Gly Gly Pro Gln Val Thr Asp Pro Thr Val Glu Asp Pro Cys Ser

```
1               5                   10                  15

Arg Asn Lys Tyr Pro Val Pro Asp Thr Met Gln Gln Thr Ile Gln Ile
            20                  25                  30

Lys Asn Pro Glu Lys Leu His Pro Ala Thr Leu Phe His Asp Trp Asp
        35                  40                  45

Leu Arg Arg Gly Phe Ile Thr Gln Ala Ala Ile Lys Arg Met Ser Glu
    50                  55                  60

Asn Leu Gln Ile Asp Ser Ser Phe Glu Ser Asp Gly Thr Glu Ser Pro
65                  70                  75                  80

Lys Lys Lys Lys Arg Cys Thr Lys Glu Ile Pro Thr Gln Asn Gln Lys
                85                  90                  95

Gln Glu Glu Ile Gln Glu Cys Leu Leu Ser Leu Cys Glu Glu Pro Thr
            100                 105                 110

Cys Gln Glu Glu Thr Glu Asp Leu Gln Leu Phe Ile Gln Gln Gln Gln
        115                 120                 125

Gln Gln Gln Tyr Lys Leu Arg Lys Asn Leu Phe Lys Leu Leu Thr His
    130                 135                 140

Leu Lys Lys Gly Gln Arg Ile Ser Gln Leu Gln Thr Gly Leu Leu Glu
145                 150                 155                 160
```

<210> 925
<211> 662
<212> PRT
<213> Gammatorquevirus sp.

<400> 925

```
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1               5                   10                  15

Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
```

```
            20                  25                  30

Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35                  40                  45

Lys Leu Arg Lys Val Lys Arg Lys Lys Lys Ser Leu Ile Val Arg Gln
    50                  55                  60

Trp Gln Pro Asp Ser Ile Arg Thr Cys Lys Ile Ile Gly Gln Ser Ala
65                  70                  75                  80

Ile Val Val Gly Ala Glu Gly Lys Gln Met Tyr Cys Tyr Thr Val Asn
                85                  90                  95

Lys Leu Ile Asn Val Pro Pro Lys Thr Pro Tyr Gly Gly Gly Phe Gly
            100                 105                 110

Val Asp Gln Tyr Thr Leu Lys Tyr Leu Tyr Glu Glu Tyr Arg Phe Ala
        115                 120                 125

Gln Asn Ile Trp Thr Gln Ser Asn Val Leu Lys Asp Leu Cys Arg Tyr
    130                 135                 140

Ile Asn Val Lys Leu Ile Phe Tyr Arg Asp Asn Lys Thr Asp Phe Val
145                 150                 155                 160

Leu Ser Tyr Asp Arg Asn Pro Pro Phe Gln Leu Thr Lys Phe Thr Tyr
                165                 170                 175

Pro Gly Ala His Pro Gln Gln Ile Met Leu Gln Lys His His Lys Phe
            180                 185                 190

Ile Leu Ser Gln Met Thr Lys Pro Asn Gly Arg Leu Thr Lys Lys Leu
        195                 200                 205

Lys Ile Lys Pro Pro Lys Gln Met Leu Ser Lys Trp Phe Phe Ser Lys
    210                 215                 220
```

```
Gln Phe Cys Lys Tyr Pro Leu Leu Ser Leu Lys Ala Ser Ala Leu Asp
225                 230                 235                 240

Leu Arg His Ser Tyr Leu Gly Cys Cys Asn Glu Asn Pro Gln Val Phe
                245                 250                 255

Phe Tyr Tyr Leu Asn His Gly Tyr Tyr Thr Ile Thr Asn Trp Gly Ala
            260                 265                 270

Gln Ser Ser Thr Ala Tyr Arg Pro Asn Ser Lys Val Thr Asp Thr Thr
        275                 280                 285

Tyr Tyr Arg Tyr Lys Asn Asp Arg Lys Asn Ile Asn Ile Lys Ser His
    290                 295                 300

Glu Tyr Glu Lys Ser Ile Ser Tyr Glu Asn Gly Tyr Phe Gln Ser Ser
305                 310                 315                 320

Phe Leu Gln Thr Gln Cys Ile Tyr Thr Ser Glu Arg Gly Glu Ala Cys
                325                 330                 335

Ile Ala Glu Lys Pro Leu Gly Ile Ala Ile Tyr Asn Pro Val Lys Asp
            340                 345                 350

Asn Gly Asp Gly Asn Met Ile Tyr Leu Val Ser Thr Leu Ala Asn Thr
        355                 360                 365

Trp Asp Gln Pro Pro Lys Asp Ser Ala Ile Leu Ile Gln Gly Val Pro
    370                 375                 380

Ile Trp Leu Gly Leu Phe Gly Tyr Leu Asp Tyr Cys Arg Gln Ile Lys
385                 390                 395                 400

Ala Asp Lys Thr Trp Leu Asp Ser His Val Leu Val Ile Gln Ser Pro
                405                 410                 415

Ala Ile Phe Thr Tyr Pro Asn Pro Gly Ala Gly Lys Trp Tyr Cys Pro
```

```
            420                 425                 430

Leu Ser Gln Ser Phe Ile Asn Gly Asn Gly Pro Phe Asn Gln Pro Pro
        435                 440                 445

Thr Leu Leu Gln Lys Ala Lys Trp Phe Pro Gln Ile Gln Tyr Gln Gln
    450                 455                 460

Glu Ile Ile Asn Ser Phe Val Glu Ser Gly Pro Phe Val Pro Lys Tyr
465                 470                 475                 480

Ala Asn Gln Thr Glu Ser Asn Trp Glu Leu Lys Tyr Lys Tyr Val Phe
                485                 490                 495

Thr Phe Lys Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala Asp
            500                 505                 510

Pro Ser Lys Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln Thr
        515                 520                 525

Ile Gln Ile Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile His
    530                 535                 540

Asp Trp Asp Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys Arg
545                 550                 555                 560

Met Ser Thr Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser Glu
                565                 570                 575

Glu Asp Ile Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr Val
            580                 585                 590

Pro Gln Gln Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu Cys
        595                 600                 605

Lys Lys Asp Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln Gln
    610                 615                 620
```

```
Leu Ile Lys Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn Ile
625                 630                 635                 640


Leu Gln Leu Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln Leu
                645                 650                 655


His Thr Gly Met Leu Pro
            660
```

<210> 926
<211> 58
<212> PRT
<213> Gammatorquevirus sp.

<400> 926

```
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1               5                   10                  15


Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
            20                  25                  30


Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35                  40                  45


Lys Leu Arg Lys Val Lys Arg Lys Lys Lys
    50                  55
```

<210> 927
<211> 202
<212> PRT
<213> Gammatorquevirus sp.

<400> 927

```
Ser Leu Ile Val Arg Gln Trp Gln Pro Asp Ser Ile Arg Thr Cys Lys
1               5                   10                  15


Ile Ile Gly Gln Ser Ala Ile Val Val Gly Ala Glu Gly Lys Gln Met
            20                  25                  30


Tyr Cys Tyr Thr Val Asn Lys Leu Ile Asn Val Pro Pro Lys Thr Pro
```

```
        35                  40                  45

Tyr Gly Gly Gly Phe Gly Val Asp Gln Tyr Thr Leu Lys Tyr Leu Tyr
    50                  55                  60

Glu Glu Tyr Arg Phe Ala Gln Asn Ile Trp Thr Gln Ser Asn Val Leu
65                  70                  75                  80

Lys Asp Leu Cys Arg Tyr Ile Asn Val Lys Leu Ile Phe Tyr Arg Asp
                85                  90                  95

Asn Lys Thr Asp Phe Val Leu Ser Tyr Asp Arg Asn Pro Pro Phe Gln
            100                 105                 110

Leu Thr Lys Phe Thr Tyr Pro Gly Ala His Pro Gln Gln Ile Met Leu
        115                 120                 125

Gln Lys His His Lys Phe Ile Leu Ser Gln Met Thr Lys Pro Asn Gly
    130                 135                 140

Arg Leu Thr Lys Lys Leu Lys Ile Lys Pro Pro Lys Gln Met Leu Ser
145                 150                 155                 160

Lys Trp Phe Phe Ser Lys Gln Phe Cys Lys Tyr Pro Leu Leu Ser Leu
                165                 170                 175

Lys Ala Ser Ala Leu Asp Leu Arg His Ser Tyr Leu Gly Cys Cys Asn
            180                 185                 190

Glu Asn Pro Gln Val Phe Phe Tyr Tyr Leu
        195                 200
```

<210> 928
<211> 79
<212> PRT
<213> Gammatorquevirus sp.

<400> 928
Asn His Gly Tyr Tyr Thr Ile Thr Asn Trp Gly Ala Gln Ser Ser Thr

1 5 10 15

Ala Tyr Arg Pro Asn Ser Lys Val Thr Asp Thr Thr Tyr Tyr Arg Tyr
20 25 30

Lys Asn Asp Arg Lys Asn Ile Asn Ile Lys Ser His Glu Tyr Glu Lys
35 40 45

Ser Ile Ser Tyr Glu Asn Gly Tyr Phe Gln Ser Ser Phe Leu Gln Thr
50 55 60

Gln Cys Ile Tyr Thr Ser Glu Arg Gly Glu Ala Cys Ile Ala Glu
65 70 75

<210> 929
<211> 160
<212> PRT
<213> Gammatorquevirus sp.

<400> 929
Lys Pro Leu Gly Ile Ala Ile Tyr Asn Pro Val Lys Asp Asn Gly Asp
1 5 10 15

Gly Asn Met Ile Tyr Leu Val Ser Thr Leu Ala Asn Thr Trp Asp Gln
20 25 30

Pro Pro Lys Asp Ser Ala Ile Leu Ile Gln Gly Val Pro Ile Trp Leu
35 40 45

Gly Leu Phe Gly Tyr Leu Asp Tyr Cys Arg Gln Ile Lys Ala Asp Lys
50 55 60

Thr Trp Leu Asp Ser His Val Leu Val Ile Gln Ser Pro Ala Ile Phe
65 70 75 80

Thr Tyr Pro Asn Pro Gly Ala Gly Lys Trp Tyr Cys Pro Leu Ser Gln
85 90 95

Ser Phe Ile Asn Gly Asn Gly Pro Phe Asn Gln Pro Pro Thr Leu Leu

100 105 110

Gln Lys Ala Lys Trp Phe Pro Gln Ile Gln Tyr Gln Gln Glu Ile Ile
115 120 125

Asn Ser Phe Val Glu Ser Gly Pro Phe Val Pro Lys Tyr Ala Asn Gln
130 135 140

Thr Glu Ser Asn Trp Glu Leu Lys Tyr Lys Tyr Val Phe Thr Phe Lys
145 150 155 160

<210> 930
<211> 163
<212> PRT
<213> Gammatorquevirus sp.

<400> 930
Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala Asp Pro Ser Lys
1 5 10 15

Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln Thr Ile Gln Ile
20 25 30

Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile His Asp Trp Asp
35 40 45

Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys Arg Met Ser Thr
50 55 60

Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser Glu Glu Asp Ile
65 70 75 80

Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr Val Pro Gln Gln
85 90 95

Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu Cys Lys Lys Asp
100 105 110

Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln Gln Leu Ile Lys

```
        115                 120                 125

Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn Ile Leu Gln Leu
    130                 135                 140

Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln Leu His Thr Gly
145                 150                 155                 160

Met Leu Pro
```

<210> 931
<211> 728
<212> PRT
<213> Alphatorquevirus sp.

<400> 931

```
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1               5                   10                  15

Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro
            20                  25                  30

Arg Arg Arg Arg Arg Arg Phe Val Ser Arg Arg Trp Arg Arg Pro Tyr
        35                  40                  45

Arg Arg Arg Arg Arg Arg Gly Arg Arg Arg Arg Arg Arg Arg Arg Arg
    50                  55                  60

His Lys Pro Thr Leu Ile Leu Arg Gln Trp Gln Pro Asp Cys Ile Arg
65                  70                  75                  80

His Cys Lys Ile Thr Gly Trp Met Pro Leu Ile Ile Cys Gly Lys Gly
                85                  90                  95

Ser Thr Gln Phe Asn Tyr Ile Thr His Ala Asp Asp Ile Thr Pro Arg
            100                 105                 110

Gly Ala Ser Tyr Gly Gly Asn Phe Thr Asn Met Thr Phe Ser Leu Glu
```

```
        115                 120                 125

Ala Ile Tyr Glu Gln Phe Leu Tyr His Arg Asn Arg Trp Ser Ala Ser
    130                 135                 140

Asn His Asp Leu Glu Leu Cys Arg Tyr Lys Gly Thr Thr Leu Lys Leu
145                 150                 155                 160

Tyr Arg His Pro Glu Val Asp Tyr Ile Val Thr Tyr Ser Arg Thr Gly
                165                 170                 175

Pro Phe Glu Ile Ser His Met Thr Tyr Leu Ser Thr His Pro Met Leu
            180                 185                 190

Met Leu Leu Asn Lys His His Ile Val Val Pro Ser Leu Lys Thr Lys
        195                 200                 205

Pro Arg Gly Arg Lys Ala Ile Lys Val Arg Ile Arg Pro Pro Lys Leu
    210                 215                 220

Met Asn Asn Lys Trp Tyr Phe Thr Arg Asp Phe Cys Asn Ile Gly Leu
225                 230                 235                 240

Phe Gln Leu Trp Ala Thr Gly Leu Glu Leu Arg Asn Pro Trp Leu Arg
                245                 250                 255

Met Ser Thr Leu Ser Pro Cys Ile Gly Phe Asn Val Leu Lys Asn Ser
            260                 265                 270

Ile Tyr Thr Asn Leu Ser Asn Leu Pro Gln Tyr Lys Asn Glu Arg Leu
        275                 280                 285

Asn Ile Ile Asn Asn Ile Leu His Pro Gln Glu Ile Thr Gly Thr Asn
    290                 295                 300

Asn Lys Lys Trp Gln Tyr Thr Tyr Thr Lys Leu Met Ala Pro Ile Tyr
305                 310                 315                 320
```

```
Tyr Ser Ala Asn Arg Ala Ser Thr Tyr Asp Trp Glu Asn Tyr Ser Lys
                325                 330                 335

Glu Thr Asn Tyr Asn Asn Thr Tyr Val Lys Phe Thr Gln Lys Arg Gln
            340                 345                 350

Glu Lys Leu Thr Lys Ile Arg Lys Glu Trp Gln Met Leu Tyr Pro Gln
        355                 360                 365

Gln Pro Thr Ala Leu Pro Asp Ser Tyr Asp Leu Leu Gln Glu Tyr Gly
    370                 375                 380

Leu Tyr Ser Pro Tyr Tyr Leu Asn Pro Thr Arg Ile Asn Leu Asp Trp
385                 390                 395                 400

Met Thr Pro Tyr Thr His Val Arg Tyr Asn Pro Leu Val Asp Lys Gly
                405                 410                 415

Phe Gly Asn Arg Ile Tyr Ile Gln Trp Cys Ser Glu Ala Asp Val Ser
            420                 425                 430

Tyr Asn Arg Thr Lys Ser Lys Cys Leu Leu Gln Asp Met Pro Leu Phe
        435                 440                 445

Phe Met Cys Tyr Gly Tyr Ile Asp Trp Ala Ile Lys Asn Thr Gly Val
    450                 455                 460

Ser Ser Leu Val Lys Asp Ala Arg Ile Cys Ile Arg Cys Pro Tyr Thr
465                 470                 475                 480

Glu Pro Gln Leu Val Gly Ser Thr Glu Asp Ile Gly Phe Val Pro Ile
                485                 490                 495

Ser Glu Thr Phe Met Arg Gly Asp Met Pro Val Leu Ala Pro Tyr Ile
            500                 505                 510

Pro Leu Ser Trp Phe Cys Lys Trp Tyr Pro Asn Ile Ala His Gln Lys
```

```
        515                 520                 525

Glu Val Leu Glu Ser Ile Ile Ser Cys Ser Pro Phe Met Pro Arg Asp
    530                 535                 540

Gln Asp Met Asn Gly Trp Asp Ile Thr Ile Gly Tyr Lys Met Asp Phe
545                 550                 555                 560

Leu Trp Gly Gly Ser Pro Leu Pro Ser Gln Pro Ile Asp Asp Pro Cys
                565                 570                 575

Gln Gln Gly Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu
            580                 585                 590

Leu Gln Val Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His
        595                 600                 605

Lys Trp Asp Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg
    610                 615                 620

Val Ser Glu Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro
625                 630                 635                 640

Ser Lys Arg Asn Lys Leu Asp Ser Ala Phe Arg Gly Glu Asn Arg Glu
                645                 650                 655

Gln Lys Glu Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr
            660                 665                 670

Pro Glu Glu Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu
        675                 680                 685

Leu Leu His Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg
    690                 695                 700

Arg Gly Leu Lys Leu Val Phe Thr Asp Ile Leu Arg Leu Arg Gln Gly
705                 710                 715                 720
```

Val His Trp Asn Pro Glu Leu Thr
725

<210> 932
<211> 66
<212> PRT
<213> Alphatorquevirus sp.

<400> 932
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1 5 10 15

Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro
20 25 30

Arg Arg Arg Arg Arg Arg Phe Val Ser Arg Arg Trp Arg Arg Pro Tyr
35 40 45

Arg Arg Arg Arg Arg Arg Gly Arg Arg Arg Arg Arg Arg Arg Arg Arg
50 55 60

His Lys
65

<210> 933
<211> 211
<212> PRT
<213> Alphatorquevirus sp.

<400> 933
Pro Thr Leu Ile Leu Arg Gln Trp Gln Pro Asp Cys Ile Arg His Cys
1 5 10 15

Lys Ile Thr Gly Trp Met Pro Leu Ile Ile Cys Gly Lys Gly Ser Thr
20 25 30

Gln Phe Asn Tyr Ile Thr His Ala Asp Asp Ile Thr Pro Arg Gly Ala
35 40 45

Ser Tyr Gly Gly Asn Phe Thr Asn Met Thr Phe Ser Leu Glu Ala Ile

```
    50                  55                  60

Tyr Glu Gln Phe Leu Tyr His Arg Asn Arg Trp Ser Ala Ser Asn His
65                  70                  75                  80

Asp Leu Glu Leu Cys Arg Tyr Lys Gly Thr Thr Leu Lys Leu Tyr Arg
                85                  90                  95

His Pro Glu Val Asp Tyr Ile Val Thr Tyr Ser Arg Thr Gly Pro Phe
            100                 105                 110

Glu Ile Ser His Met Thr Tyr Leu Ser Thr His Pro Met Leu Met Leu
        115                 120                 125

Leu Asn Lys His His Ile Val Val Pro Ser Leu Lys Thr Lys Pro Arg
    130                 135                 140

Gly Arg Lys Ala Ile Lys Val Arg Ile Arg Pro Pro Lys Leu Met Asn
145                 150                 155                 160

Asn Lys Trp Tyr Phe Thr Arg Asp Phe Cys Asn Ile Gly Leu Phe Gln
                165                 170                 175

Leu Trp Ala Thr Gly Leu Glu Leu Arg Asn Pro Trp Leu Arg Met Ser
            180                 185                 190

Thr Leu Ser Pro Cys Ile Gly Phe Asn Val Leu Lys Asn Ser Ile Tyr
        195                 200                 205

Thr Asn Leu
    210


<210> 934
<211> 118
<212> PRT
<213> Alphatorquevirus sp.

<400> 934
Ser Asn Leu Pro Gln Tyr Lys Asn Glu Arg Leu Asn Ile Ile Asn Asn
```

```
1               5                   10                  15

Ile Leu His Pro Gln Glu Ile Thr Gly Thr Asn Asn Lys Lys Trp Gln
            20                  25                  30

Tyr Thr Tyr Thr Lys Leu Met Ala Pro Ile Tyr Tyr Ser Ala Asn Arg
        35                  40                  45

Ala Ser Thr Tyr Asp Trp Glu Asn Tyr Ser Lys Glu Thr Asn Tyr Asn
    50                  55                  60

Asn Thr Tyr Val Lys Phe Thr Gln Lys Arg Gln Glu Lys Leu Thr Lys
65                  70                  75                  80

Ile Arg Lys Glu Trp Gln Met Leu Tyr Pro Gln Gln Pro Thr Ala Leu
                85                  90                  95

Pro Asp Ser Tyr Asp Leu Leu Gln Glu Tyr Gly Leu Tyr Ser Pro Tyr
            100                 105                 110

Tyr Leu Asn Pro Thr Arg
        115
```

<210> 935
<211> 166
<212> PRT
<213> Alphatorquevirus sp.

<400> 935

```
Ile Asn Leu Asp Trp Met Thr Pro Tyr Thr His Val Arg Tyr Asn Pro
1               5                   10                  15

Leu Val Asp Lys Gly Phe Gly Asn Arg Ile Tyr Ile Gln Trp Cys Ser
            20                  25                  30

Glu Ala Asp Val Ser Tyr Asn Arg Thr Lys Ser Lys Cys Leu Leu Gln
        35                  40                  45

Asp Met Pro Leu Phe Phe Met Cys Tyr Gly Tyr Ile Asp Trp Ala Ile
```

50 55 60

Lys Asn Thr Gly Val Ser Ser Leu Val Lys Asp Ala Arg Ile Cys Ile
65 70 75 80

Arg Cys Pro Tyr Thr Glu Pro Gln Leu Val Gly Ser Thr Glu Asp Ile
85 90 95

Gly Phe Val Pro Ile Ser Glu Thr Phe Met Arg Gly Asp Met Pro Val
100 105 110

Leu Ala Pro Tyr Ile Pro Leu Ser Trp Phe Cys Lys Trp Tyr Pro Asn
115 120 125

Ile Ala His Gln Lys Glu Val Leu Glu Ser Ile Ile Ser Cys Ser Pro
130 135 140

Phe Met Pro Arg Asp Gln Asp Met Asn Gly Trp Asp Ile Thr Ile Gly
145 150 155 160

Tyr Lys Met Asp Phe Leu
165

<210> 936
<211> 167
<212> PRT
<213> Alphatorquevirus sp.

<400> 936
Trp Gly Gly Ser Pro Leu Pro Ser Gln Pro Ile Asp Asp Pro Cys Gln
1 5 10 15

Gln Gly Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu Leu
20 25 30

Gln Val Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His Lys
35 40 45

Trp Asp Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg Val

50 55 60

Ser Glu Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro Ser
65 70 75 80

Lys Arg Asn Lys Leu Asp Ser Ala Phe Arg Gly Glu Asn Arg Glu Gln
85 90 95

Lys Glu Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr Pro
100 105 110

Glu Glu Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu Leu
115 120 125

Leu His Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg Arg
130 135 140

Gly Leu Lys Leu Val Phe Thr Asp Ile Leu Arg Leu Arg Gln Gly Val
145 150 155 160

His Trp Asn Pro Glu Leu Thr
165

<210> 937

<400> 937
000

<210> 938

<400> 938
000

<210> 939

<400> 939
000

<210> 940

<400> 940
000

<210> 941

<400> 941
000

<210> 942

<400> 942
000

<210> 943

<400> 943
000

<210> 944

<400> 944
000

<210> 945

<400> 945
000

<210> 946

<400> 946
000

<210> 947

<400> 947
000

<210> 948

<400> 948

000

<210> 949
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
peptide

<220>
<221> MOD_RES
<222> (1)..(1)
<223> W or F

<220>
<221> MOD_RES
<222> (2)..(8)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (10)..(12)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (14)..(14)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (16)..(20)
<223> Any amino acid

<400> 949
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Cys Xaa Cys Xaa
1 5 10 15

Xaa Xaa Xaa Xaa His
20

<210> 950

<400> 950

000

<210> 951

<400> 951
000

<210> 952

<400> 952
000

<210> 953

<400> 953
000

<210> 954

<400> 954
000

<210> 955

<400> 955
000

<210> 956

<400> 956
000

<210> 957

<400> 957
000

<210> 958

<400> 958
000

<210> 959

<400> 959
000

<210> 960

<400> 960
000

<210> 961

<400> 961
000

<210> 962

<400> 962
000

<210> 963

<400> 963
000

<210> 964

<400> 964
000

<210> 965

<400> 965
000

<210> 966

<400> 966
000

<210> 967

<400> 967
000

<210> 968

<400> 968
000

<210> 969

<400> 969
000

<210> 970

<400> 970
000

<210> 971

<400> 971
000

<210> 972

<400> 972
000

<210> 973

<400> 973
000

<210> 974

<400> 974
000

<210> 975

<400> 975

000

<210> 976

<400> 976
000

<210> 977

<400> 977
000

<210> 978

<400> 978
000

<210> 979

<400> 979
000

<210> 980

<400> 980
000

<210> 981

<400> 981
000

<210> 982

<400> 982
000

<210> 983

<400> 983
000

<210> 984

<400> 984
000

<210> 985

<400> 985
000

<210> 986

<400> 986
000

<210> 987

<400> 987
000

<210> 988

<400> 988
000

<210> 989

<400> 989
000

<210> 990

<400> 990
000

<210> 991

<400> 991
000

<210> 992

<400> 992
000

<210> 993

<400> 993
000

<210> 994

<400> 994
000

<210> 995

<400> 995
000

<210> 996

<400> 996
000

<210> 997

<400> 997
000

<210> 998

<400> 998
000

<210> 999

<400> 999
000

<210> 1000

<400> 1000

000

<210> 1001
<211> 2845
<212> DNA
<213> Betatorquevirus sp.

<400> 1001
ttattaatat tcaacaggaa aaccacctaa tttaaattgc cgaccacaaa ccgtcactaa 60

cttccttatt taacattact tccctttaa ccaatgaata ttcatacaac acatcacact 120

tcctgggagg agacataaaa ctatataact aactacacag acgaatggct gagtttatgc 180

cgctagacgg aggacgcaca gctactgctg cgacctgaac ttgggcgggt gccgaaggtg 240

agtgtaacca ccgtagtcaa ggggcaattc gggctagttc agtctagcgg aacgggcaag 300

attattaata caaacttatt tttacagatg agcaaacaac taaaaccaac tttatacaaa 360

gacaaatcat tggaattaca atggctaaac aacattttta gctctcacga cctgtgctgc 420

ggctgcaacg atccagtttt acatttactg attttaatta acaaaaccgg agaagcacct 480

aaaccagaag aagacattaa aaatataaaa tgcctcctta ctggcgccaa aaatactacc 540

gaagaagata tagacctttc tcctggagaa ctagaagaat tattcaaaga agaaaaagat 600

ggagataccg caaaccaaga aaaacatact ggagaagaaa actgcgggta agaaaacgtt 660

tttataaaag aaagttaaaa aaaattgtac ttaaacagtt tcaaccaaaa attattagaa 720

gatgtacaat atttggaaca atctgcctat ttcaaggctc tccagaaaga gccaacaata 780

attatattca aacaatctac tcctacgtac cagataaaga accaggagga gggggatgga 840

ctttaataac tgaaagctta agtagtttat gggaagactg ggaacattta aaaaatgtat 900

ggactcaaag taacgctggt ttaccacttg taagatacgg gggagtaaca ttatactttt 960

atcaatctgc ctatactgac tatattgctc aagttttcaa ctgttatcct atgacagaca 1020

caaaatacac acatgcagac tcagcaccaa acagaatgtt attaaaaaaa catgtaataa 1080

gagtacctag cagagaaaca cgcaaaaaaa gaaagccata caaaagagtt agagtaggac 1140

ctccttctca aatgcaaaac aaatggtact ttcaaagaga catatgtgaa ataccattaa 1200

taatgattgc agccacagcc gttgacttta gatatccctt ttgtgcaagc gactgtgcta 1260

gtaacaactt aactctaaca tgtttaaacc cactattgtt tcaaaaccaa gactttgacc 1320

acccatccga tacacaaggc tactttccaa aacctggagt atatctatac tcaacacaaa 1380

gaagtaacaa gccaagttct tcagactgta tatacttagg aaacacaaaa gacaatcaag 1440

aaggtaaatc tgcaagtagt ctaatgactc taaaaacaca aaaaataaca gattggggaa 1500

atccattttg gcattattat atagacggtt ctaaaaaaat attttcttac tttaaacccc 1560

catcacaatt agacagcagc gactttgaac acatgacaga attagcagaa ccaatgttta 1620

tacaagttag atacaaccca gaaagagaca caggacaagg aaacttaata tacgtaacag 1680

aaaactttag aggacaacac tgggaccctc catctagtga caacctaaaa ttagatggat 1740

ttcccttata tgacatgtgc tggggtttca tagactggat agaaaaagtt catgaaacag 1800

aaaacttact taccaactac tgcttctgta ttagaagcag cgctttcaat gaaaaaaaaa 1860

cagtttttat acctgtagat cattcatttt taacaggttt tagcccatat gaaactccag 1920

ttaaatcatc agaccaagct cactggcacc cacaaataag atttcaaaca aaatcaataa 1980

atgacatttg tttaacaggc cccggttgtg ctaggtcccc atatggcaat tacatgcagg 2040

caaaaatgag ttataaattt catgtaaaat ggggaggatg tccaaaaact tatgaaaaac 2100

catatgatcc ttgttcacag cccaattgga ctattcccca taacctcaat gaaacaatac 2160

aaatccagaa tccaaacaca tgcccacaaa cagaactcca agaatgggac tggcgacgtg 2220

atattgttac aaaaaaagct atcgaaagaa ttagacaaca cacggaacct catgaaactt 2280

tgcaaatctc tacaggttcc aaacacaacc caccagtaca cagacaaaca tcaccgtgga 2340

cggactcaga aacggactcg gaagaggaaa aagaccaaac acaagagatc cagatccagc 2400

tcaacaagct cagaaagcat caacagcatc tcaagcagca gctcaagcag tacctgaaac 2460

cccaaaatat agaatagttg caagcaacat aaaagttgaa ctttttccta ctaaaaaacc 2520

ttttaaaaac agacgcttta ctccttctga aagagaaaca gaaagacaat gtgctaaagc 2580

tttttgtaga ccagaaagac atttctttta tgatcctcct ttttaccctt actgtgtacc 2640

tgaacctatt gtaaactttg ctttgggata taaaatttaa ggccaacaaa tttcacttag 2700

tggtgtctgt ttattaaagt ttaaccttaa taagcatact ccgcctccct acattaaggc 2760

gccaaaaggg ggctccgccc ccttaaaccc caaggggggct ccgccccctt aaaccccccaa 2820

gggggctccg ccccccttaca ccccc 2845

<210> 1002
<211> 107
<212> PRT
<213> Betatorquevirus sp.

<400> 1002
Met Ser Lys Gln Leu Lys Pro Thr Leu Tyr Lys Asp Lys Ser Leu Glu
1 5 10 15

Leu Gln Trp Leu Asn Asn Ile Phe Ser Ser His Asp Leu Cys Cys Gly
20 25 30

Cys Asn Asp Pro Val Leu His Leu Leu Ile Leu Ile Asn Lys Thr Gly
35 40 45

Glu Ala Pro Lys Pro Glu Glu Asp Ile Lys Asn Ile Lys Cys Leu Leu
50 55 60

Thr Gly Ala Lys Asn Thr Thr Glu Glu Asp Ile Asp Leu Ser Pro Gly
65 70 75 80

Glu Leu Glu Glu Leu Phe Lys Glu Glu Lys Asp Gly Asp Thr Ala Asn
85 90 95

Gln Glu Lys His Thr Gly Glu Glu Asn Cys Gly
100 105

<210> 1003
<211> 218
<212> PRT
<213> Betatorquevirus sp.

<400> 1003
Met Ser Lys Gln Leu Lys Pro Thr Leu Tyr Lys Asp Lys Ser Leu Glu
1 5 10 15

```
Leu Gln Trp Leu Asn Asn Ile Phe Ser Ser His Asp Leu Cys Cys Gly
            20                  25                  30

Cys Asn Asp Pro Val Leu His Leu Leu Ile Leu Ile Asn Lys Thr Gly
        35                  40                  45

Glu Ala Pro Lys Pro Glu Glu Asp Ile Lys Asn Ile Lys Cys Leu Leu
    50                  55                  60

Thr Gly Ala Lys Asn Thr Thr Glu Glu Asp Ile Asp Leu Ser Pro Gly
65                  70                  75                  80

Glu Leu Glu Glu Leu Phe Lys Glu Glu Lys Asp Gly Asp Thr Ala Asn
                85                  90                  95

Gln Glu Lys His Thr Gly Glu Glu Asn Cys Gly Pro Ile Gly Leu Phe
            100                 105                 110

Pro Ile Thr Ser Met Lys Gln Tyr Lys Ser Arg Ile Gln Thr His Ala
        115                 120                 125

His Lys Gln Asn Ser Lys Asn Gly Thr Gly Asp Val Ile Leu Leu Gln
    130                 135                 140

Lys Lys Leu Ser Lys Glu Leu Asp Asn Thr Arg Asn Leu Met Lys Leu
145                 150                 155                 160

Cys Lys Ser Leu Gln Val Pro Asn Thr Thr His Gln Tyr Thr Asp Lys
                165                 170                 175

His His Arg Gly Arg Thr Gln Lys Arg Thr Arg Lys Arg Lys Lys Thr
            180                 185                 190

Lys His Lys Arg Ser Arg Ser Ser Ser Thr Ser Ser Glu Ser Ile Asn
        195                 200                 205

Ser Ile Ser Ser Ser Ser Ser Ser Ser Thr
    210                 215
```

<210> 1004
<211> 234
<212> PRT
<213> Betatorquevirus sp.

<400> 1004
Met Ser Lys Gln Leu Lys Pro Thr Leu Tyr Lys Asp Lys Ser Leu Glu
1   5   10   15

Leu Gln Trp Leu Asn Asn Ile Phe Ser Ser His Asp Leu Cys Cys Gly
            20   25   30

Cys Asn Asp Pro Val Leu His Leu Leu Ile Leu Ile Asn Lys Thr Gly
        35   40   45

Glu Ala Pro Lys Pro Glu Glu Asp Ile Lys Asn Ile Lys Cys Leu Leu
    50   55   60

Thr Gly Ala Lys Asn Thr Thr Glu Glu Asp Ile Asp Leu Ser Pro Gly
65   70   75   80

Glu Leu Glu Glu Leu Phe Lys Glu Glu Lys Asp Gly Asp Thr Ala Asn
                85   90   95

Gln Glu Lys His Thr Gly Glu Glu Asn Cys Gly Phe Gln Thr Gln Pro
            100   105   110

Thr Ser Thr Gln Thr Asn Ile Thr Val Asp Gly Leu Arg Asn Gly Leu
        115   120   125

Gly Arg Gly Lys Arg Pro Asn Thr Arg Asp Pro Asp Pro Ala Gln Gln
    130   135   140

Ala Gln Lys Ala Ser Thr Ala Ser Gln Ala Ala Ala Gln Ala Val Pro
145   150   155   160

Glu Thr Pro Lys Tyr Arg Ile Val Ala Ser Asn Ile Lys Val Glu Leu
                165   170   175

```
Phe Pro Thr Lys Lys Pro Phe Lys Asn Arg Arg Phe Thr Pro Ser Glu
            180                 185                 190


Arg Glu Thr Glu Arg Gln Cys Ala Lys Ala Phe Cys Arg Pro Glu Arg
        195                 200                 205


His Phe Phe Tyr Asp Pro Pro Phe Tyr Pro Tyr Cys Val Pro Glu Pro
    210                 215                 220


Ile Val Asn Phe Ala Leu Gly Tyr Lys Ile
225                 230


<210> 1005
<211> 655
<212> PRT
<213> Betatorquevirus sp.

<400> 1005
Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg Pro
1               5                   10                  15


Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp Arg
            20                  25                  30


Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Val Arg
        35                  40                  45


Lys Arg Phe Tyr Lys Arg Lys Leu Lys Lys Ile Val Leu Lys Gln Phe
    50                  55                  60


Gln Pro Lys Ile Ile Arg Arg Cys Thr Ile Phe Gly Thr Ile Cys Leu
65                  70                  75                  80


Phe Gln Gly Ser Pro Glu Arg Ala Asn Asn Asn Tyr Ile Gln Thr Ile
                85                  90                  95


Tyr Ser Tyr Val Pro Asp Lys Glu Pro Gly Gly Gly Gly Trp Thr Leu
            100                 105                 110
```

```
Ile Thr Glu Ser Leu Ser Ser Leu Trp Glu Asp Trp Glu His Leu Lys
        115                 120                 125

Asn Val Trp Thr Gln Ser Asn Ala Gly Leu Pro Leu Val Arg Tyr Gly
    130                 135                 140

Gly Val Thr Leu Tyr Phe Tyr Gln Ser Ala Tyr Thr Asp Tyr Ile Ala
145                 150                 155                 160

Gln Val Phe Asn Cys Tyr Pro Met Thr Asp Thr Lys Tyr Thr His Ala
                165                 170                 175

Asp Ser Ala Pro Asn Arg Met Leu Leu Lys Lys His Val Ile Arg Val
            180                 185                 190

Pro Ser Arg Glu Thr Arg Lys Lys Arg Lys Pro Tyr Lys Arg Val Arg
        195                 200                 205

Val Gly Pro Pro Ser Gln Met Gln Asn Lys Trp Tyr Phe Gln Arg Asp
    210                 215                 220

Ile Cys Glu Ile Pro Leu Ile Met Ile Ala Ala Thr Ala Val Asp Phe
225                 230                 235                 240

Arg Tyr Pro Phe Cys Ala Ser Asp Cys Ala Ser Asn Asn Leu Thr Leu
                245                 250                 255

Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp His Pro
            260                 265                 270

Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu Tyr Ser
        275                 280                 285

Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr Leu Gly
    290                 295                 300
```

```
Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu Met Thr
305                 310                 315                 320

Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp His Tyr
                325                 330                 335

Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro Pro Ser
            340                 345                 350

Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala Glu Pro
        355                 360                 365

Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly Gln Gly
    370                 375                 380

Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp Asp Pro
385                 390                 395                 400

Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr Asp Met
                405                 410                 415

Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr Glu Asn
            420                 425                 430

Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe Asn Glu
        435                 440                 445

Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr Gly Phe
    450                 455                 460

Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His Trp His
465                 470                 475                 480

Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys Leu Thr
                485                 490                 495

Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala Lys
            500                 505                 510
```

```
Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys Thr Tyr
        515                 520                 525

Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile Pro His
    530                 535                 540

Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro Gln
545                 550                 555                 560

Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys Lys
                565                 570                 575

Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu Gln
            580                 585                 590

Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr Ser
        595                 600                 605

Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln Thr
    610                 615                 620

Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln His
625                 630                 635                 640

Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu
                645                 650                 655
```

<210> 1006
<211> 164
<212> PRT
<213> Betatorquevirus sp.

<400> 1006

```
Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg Pro
1               5                   10                  15

Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp Arg
            20                  25                  30
```

```
Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Pro Asn
        35                  40                  45

Trp Thr Ile Pro His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro
    50                  55                  60

Asn Thr Cys Pro Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp
65                  70                  75                  80

Ile Val Thr Lys Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro
                85                  90                  95

His Glu Thr Leu Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val
            100                 105                 110

His Arg Gln Thr Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu
        115                 120                 125

Glu Lys Asp Gln Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg
    130                 135                 140

Lys His Gln Gln His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro
145                 150                 155                 160

Gln Asn Ile Glu
```

<210> 1007
<211> 99
<212> PRT
<213> Betatorquevirus sp.

<400> 1007

```
Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg Pro
1               5                   10                  15

Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp Arg
            20                  25                  30
```

Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Val Pro
35 40 45

Asn Thr Thr His Gln Tyr Thr Asp Lys His His Arg Gly Arg Thr Gln
50 55 60

Lys Arg Thr Arg Lys Arg Lys Lys Thr Lys His Lys Arg Ser Arg Ser
65 70 75 80

Ser Ser Thr Ser Ser Glu Ser Ile Asn Ser Ile Ser Ser Ser Ser Ser
85 90 95

Ser Ser Thr

<210> 1008
<211> 2912
<212> DNA
<213> Betatorquevirus sp.

<400> 1008
taataaatat tcaacaggaa aaccacctaa tttaaattgc cgaccacaaa ccgtcactta 60

gttcctcttt ttccacaact tcctctttta ctaatgaata ttcatgtaat taattaataa 120

tcaccgtaat tccggggagg agcctttaaa ctataaaact aactacacat tcgaatggct 180

gagtttatgc cgccagacgg agacgggatc acttcagtga ctccaggctg atcaagggcg 240

ggtgccgaag gtgagtgaaa ccaccgtagt caaggggcaa ttcgggctag atcagtctgg 300

cggaacgggc aagaaactta aaatgtactt tattttacag aaatgttcaa atctccaaca 360

tacttaacaa ctaaaggcaa aaacaatgcc ttaatcaact gcttcgttgg agaccacgat 420

cttctgtgca gctgtaacaa tcctgcctac cattgcctcc aaatacttgc aactacctta 480

gcacctcaac taaaacaaga agaaaaacaa caaataatac aatgccttgg tggtacagac 540

gccgtagcta caacccgtgg agacgaagaa attggtttag aagacctaga aaaactattt 600

acagaagata cagaagaaga cgccgctggg taagaagaaa acctttttac aaacgtaaaa 660

ttaagagact aaatatagta gaatggcaac ctaaatcaat tagaaaatgt agaataaaag 720

gaatgctatg cttgtttcaa acgacagaag acagactgtc atataacttt gatatgtatg 780

aagagtctat tataccagaa aaactgccgg gagggggggg atttagcatt aagaatataa 840

gcttatatgc cttataccaa gaacacatac atgcacacaa catatttaca cacacaaaca 900

cagacagacc actagcaaga tacacaggct gttctttaaa attctaccaa agcaaagaca 960

tagactacgt agtaacatat tctacatcac tcccactaag aagctcaatg ggaatgtaca 1020

actccatgca accatccata catctaatgc aacaaaacaa actaattgta ccaagcaaac 1080

aaacacaaaa aagaagaaaa ccatatatta aaaaacatat atcaccacca acacaaatga 1140

aatctcaatg gtactttcaa cataacattg caaacatacc gctactaatg ataagaacca 1200

cagcattaac attagataat tactatatag gaagcagaca attaagtaca aatgtcacta 1260

tacatacact taacacaaca tacatccaaa acagagactg gggagacaga aataaaactt 1320

actactgcca aacattagga acacaaagat acttcctata tggaacacat tcaactgcac 1380

aaaatattaa tgacataaag ctacaagaac taataccttt aacaaacaca caagactatg 1440

tacaaggctt tgattggaca gaaaaagaca aacataacat aacaacctac aaagaattct 1500

taactaaagg agcaggaaat ccatttcacg cagaatggat aacagcacaa aacccagtaa 1560

tacacacagc aaacagtcct acacaaatag aacaaatata caccgcttca acaacaacat 1620

tccaaaacaa aaaactaaca gacctaccaa cgccaggata tatatttata actccaacag 1680

taagcttaag atacaaccca tacaaagacc tagcagaaag aaacaaatgc tactttgtaa 1740

gaagcaaaat aaatgcacac gggtgggacc cagaacaaca ccaagaatta ataaacagtg 1800

acctaccaca atggttacta ttatttggct acccagacta cataaaaaga acacaaaact 1860

ttgcattagt agacacaaat tacatactag tagaccactg cccatacaca aatccagaaa 1920

aaacaccatt tataccttta agcacatcat ttatagaagg tagaagccca tacagtcctt 1980

cagacacaca tgaaccagat gaagaagacc aaaacaggtg gtacccatgc taccaatatc 2040

aacaagaatc aataaattca atatgtctta gcggtccagg cacaccaaaa ataccaaaag 2100

gaataacagc agaagcaaaa gtaaaatatt cctttaattt taagtggggt ggtgacctac 2160

caccaatgtc tacaattaca aacccgacag accagccaac atatgttgtt cccaataact 2220

tcaatgaaac aacttcgtta cagaatccaa ccaccagacc agagcacttc ttgtactcct 2280

ttgacgaaag gaggggacaa cttacagaaa aagctacaaa acgcttgctt aaagactggg 2340

aaactaaaga aacttcttta ttgtctacag aatacagatt cgcggagcca acacaaacac 2400

aagccccaca agaggacccg tcctcggaag aagaagaaga gagcaacctc ttcgagcgac 2460

tcctccgaca gcgaaccaag cagctccagc tcaagcgcag aataatacaa acattgaaag 2520

acctacaaaa attagaataa ctaacagcaa aaacaccgtt tacctatttc cacctgaaca 2580

aaagaacaga agactaacac catgggaaat acaagaagac aaagaaatag ccaatttatt 2640

tggcagacca catagatact ttttaaaaga cattcctttc tattgggata tacccccaga 2700

gcctaaagta aactttgatt taaattttca ataaagaaat aaagggcaag gccccattaa 2760

ctcaaagtcg gtgtctacct ctttaagttt aactttacta aacggactcc gcctccctaa 2820

atttgggcgc caaaaggggg ctccgccccc ttaaacccca gggggctccg ccccctaaaa 2880

cccccaaggg ggctacgccc ccttacaccc cc 2912

<210> 1009
<211> 96
<212> PRT
<213> Betatorquevirus sp.

<400> 1009
Met Phe Lys Ser Pro Thr Tyr Leu Thr Thr Lys Gly Lys Asn Asn Ala
1 5 10 15

Leu Ile Asn Cys Phe Val Gly Asp His Asp Leu Leu Cys Ser Cys Asn
20 25 30

Asn Pro Ala Tyr His Cys Leu Gln Ile Leu Ala Thr Thr Leu Ala Pro
35 40 45

Gln Leu Lys Gln Glu Glu Lys Gln Gln Ile Ile Gln Cys Leu Gly Gly
50 55 60

Thr Asp Ala Val Ala Thr Thr Arg Gly Asp Glu Glu Ile Gly Leu Glu

65 70 75 80

Asp Leu Glu Lys Leu Phe Thr Glu Asp Thr Glu Glu Asp Ala Ala Gly
85 90 95

<210> 1010
<211> 198
<212> PRT
<213> Betatorquevirus sp.

<400> 1010
Met Phe Lys Ser Pro Thr Tyr Leu Thr Thr Lys Gly Lys Asn Asn Ala
1 5 10 15

Leu Ile Asn Cys Phe Val Gly Asp His Asp Leu Leu Cys Ser Cys Asn
20 25 30

Asn Pro Ala Tyr His Cys Leu Gln Ile Leu Ala Thr Thr Leu Ala Pro
35 40 45

Gln Leu Lys Gln Glu Glu Lys Gln Gln Ile Ile Gln Cys Leu Gly Gly
50 55 60

Thr Asp Ala Val Ala Thr Thr Arg Gly Asp Glu Glu Ile Gly Leu Glu
65 70 75 80

Asp Leu Glu Lys Leu Phe Thr Glu Asp Thr Glu Glu Asp Ala Ala Gly
85 90 95

Gln His Met Leu Phe Pro Ile Thr Ser Met Lys Gln Leu Arg Tyr Arg
100 105 110

Ile Gln Pro Pro Asp Gln Ser Thr Ser Cys Thr Pro Leu Thr Lys Gly
115 120 125

Gly Asp Asn Leu Gln Lys Lys Leu Gln Asn Ala Cys Leu Lys Thr Gly
130 135 140

Lys Leu Lys Lys Leu Leu Tyr Cys Leu Gln Asn Thr Asp Ser Arg Ser

145 150 155 160

Gln His Lys His Lys Pro His Lys Arg Thr Arg Pro Arg Lys Lys Lys
165 170 175

Lys Arg Ala Thr Ser Ser Ser Asp Ser Ser Asp Ser Glu Pro Ser Ser
180 185 190

Ser Ser Ser Ser Ala Glu
195

<210> 1011
<211> 216
<212> PRT
<213> Betatorquevirus sp.

<400> 1011
Met Phe Lys Ser Pro Thr Tyr Leu Thr Thr Lys Gly Lys Asn Asn Ala
1 5 10 15

Leu Ile Asn Cys Phe Val Gly Asp His Asp Leu Leu Cys Ser Cys Asn
20 25 30

Asn Pro Ala Tyr His Cys Leu Gln Ile Leu Ala Thr Thr Leu Ala Pro
35 40 45

Gln Leu Lys Gln Glu Glu Lys Gln Gln Ile Ile Gln Cys Leu Gly Gly
50 55 60

Thr Asp Ala Val Ala Thr Thr Arg Gly Asp Glu Glu Ile Gly Leu Glu
65 70 75 80

Asp Leu Glu Lys Leu Phe Thr Glu Asp Thr Glu Glu Asp Ala Ala Gly
85 90 95

Ile Gln Ile Arg Gly Ala Asn Thr Asn Thr Ser Pro Thr Arg Gly Pro
100 105 110

Val Leu Gly Arg Arg Arg Arg Glu Gln Pro Leu Arg Ala Thr Pro Pro

```
        115                 120                 125

Thr Ala Asn Gln Ala Ala Pro Ala Gln Ala Gln Asn Asn Thr Asn Ile
    130                 135                 140

Glu Arg Pro Thr Lys Ile Arg Ile Thr Asn Ser Lys Asn Thr Val Tyr
145                 150                 155                 160

Leu Phe Pro Pro Glu Gln Lys Asn Arg Arg Leu Thr Pro Trp Glu Ile
                165                 170                 175

Gln Glu Asp Lys Glu Ile Ala Asn Leu Phe Gly Arg Pro His Arg Tyr
            180                 185                 190

Phe Leu Lys Asp Ile Pro Phe Tyr Trp Asp Ile Pro Pro Glu Pro Lys
        195                 200                 205

Val Asn Phe Asp Leu Asn Phe Gln
    210                 215
```

<210> 1012
<211> 672
<212> PRT
<213> Betatorquevirus sp.

<400> 1012

```
Met Pro Trp Trp Tyr Arg Arg Arg Ser Tyr Asn Pro Trp Arg Arg Arg
1               5                   10                  15

Asn Trp Phe Arg Arg Pro Arg Lys Thr Ile Tyr Arg Arg Tyr Arg Arg
            20                  25                  30

Arg Arg Arg Trp Val Arg Arg Lys Pro Phe Tyr Lys Arg Lys Ile Lys
        35                  40                  45

Arg Leu Asn Ile Val Glu Trp Gln Pro Lys Ser Ile Arg Lys Cys Arg
    50                  55                  60

Ile Lys Gly Met Leu Cys Leu Phe Gln Thr Thr Glu Asp Arg Leu Ser
```

```
65                  70                  75                  80

Tyr Asn Phe Asp Met Tyr Glu Glu Ser Ile Ile Pro Glu Lys Leu Pro
                85                  90                  95

Gly Gly Gly Gly Phe Ser Ile Lys Asn Ile Ser Leu Tyr Ala Leu Tyr
            100                 105                 110

Gln Glu His Ile His Ala His Asn Ile Phe Thr His Thr Asn Thr Asp
        115                 120                 125

Arg Pro Leu Ala Arg Tyr Thr Gly Cys Ser Leu Lys Phe Tyr Gln Ser
    130                 135                 140

Lys Asp Ile Asp Tyr Val Val Thr Tyr Ser Thr Ser Leu Pro Leu Arg
145                 150                 155                 160

Ser Ser Met Gly Met Tyr Asn Ser Met Gln Pro Ser Ile His Leu Met
                165                 170                 175

Gln Gln Asn Lys Leu Ile Val Pro Ser Lys Gln Thr Gln Lys Arg Arg
            180                 185                 190

Lys Pro Tyr Ile Lys Lys His Ile Ser Pro Pro Thr Gln Met Lys Ser
        195                 200                 205

Gln Trp Tyr Phe Gln His Asn Ile Ala Asn Ile Pro Leu Leu Met Ile
    210                 215                 220

Arg Thr Thr Ala Leu Thr Leu Asp Asn Tyr Tyr Ile Gly Ser Arg Gln
225                 230                 235                 240

Leu Ser Thr Asn Val Thr Ile His Thr Leu Asn Thr Thr Tyr Ile Gln
                245                 250                 255

Asn Arg Asp Trp Gly Asp Arg Asn Lys Thr Tyr Tyr Cys Gln Thr Leu
            260                 265                 270
```

```
Gly Thr Gln Arg Tyr Phe Leu Tyr Gly Thr His Ser Thr Ala Gln Asn
        275                 280                 285

Ile Asn Asp Ile Lys Leu Gln Glu Leu Ile Pro Leu Thr Asn Thr Gln
    290                 295                 300

Asp Tyr Val Gln Gly Phe Asp Trp Thr Glu Lys Asp Lys His Asn Ile
305                 310                 315                 320

Thr Thr Tyr Lys Glu Phe Leu Thr Lys Gly Ala Gly Asn Pro Phe His
                325                 330                 335

Ala Glu Trp Ile Thr Ala Gln Asn Pro Val Ile His Thr Ala Asn Ser
            340                 345                 350

Pro Thr Gln Ile Glu Gln Ile Tyr Thr Ala Ser Thr Thr Thr Phe Gln
        355                 360                 365

Asn Lys Lys Leu Thr Asp Leu Pro Thr Pro Gly Tyr Ile Phe Ile Thr
    370                 375                 380

Pro Thr Val Ser Leu Arg Tyr Asn Pro Tyr Lys Asp Leu Ala Glu Arg
385                 390                 395                 400

Asn Lys Cys Tyr Phe Val Arg Ser Lys Ile Asn Ala His Gly Trp Asp
                405                 410                 415

Pro Glu Gln His Gln Glu Leu Ile Asn Ser Asp Leu Pro Gln Trp Leu
            420                 425                 430

Leu Leu Phe Gly Tyr Pro Asp Tyr Ile Lys Arg Thr Gln Asn Phe Ala
        435                 440                 445

Leu Val Asp Thr Asn Tyr Ile Leu Val Asp His Cys Pro Tyr Thr Asn
    450                 455                 460

Pro Glu Lys Thr Pro Phe Ile Pro Leu Ser Thr Ser Phe Ile Glu Gly
```

```
465                 470                 475                 480

Arg Ser Pro Tyr Ser Pro Ser Asp Thr His Glu Pro Asp Glu Glu Asp
                485                 490                 495

Gln Asn Arg Trp Tyr Pro Cys Tyr Gln Tyr Gln Gln Glu Ser Ile Asn
            500                 505                 510

Ser Ile Cys Leu Ser Gly Pro Gly Thr Pro Lys Ile Pro Lys Gly Ile
        515                 520                 525

Thr Ala Glu Ala Lys Val Lys Tyr Ser Phe Asn Phe Lys Trp Gly Gly
    530                 535                 540

Asp Leu Pro Pro Met Ser Thr Ile Thr Asn Pro Thr Asp Gln Pro Thr
545                 550                 555                 560

Tyr Val Val Pro Asn Asn Phe Asn Glu Thr Thr Ser Leu Gln Asn Pro
                565                 570                 575

Thr Thr Arg Pro Glu His Phe Leu Tyr Ser Phe Asp Glu Arg Arg Gly
            580                 585                 590

Gln Leu Thr Glu Lys Ala Thr Lys Arg Leu Leu Lys Asp Trp Glu Thr
        595                 600                 605

Lys Glu Thr Ser Leu Leu Ser Thr Glu Tyr Arg Phe Ala Glu Pro Thr
    610                 615                 620

Gln Thr Gln Ala Pro Gln Glu Asp Pro Ser Ser Glu Glu Glu Glu Glu
625                 630                 635                 640

Ser Asn Leu Phe Glu Arg Leu Leu Arg Gln Arg Thr Lys Gln Leu Gln
                645                 650                 655

Leu Lys Arg Arg Ile Ile Gln Thr Leu Lys Asp Leu Gln Lys Leu Glu
            660                 665                 670
```

<210> 1013
<211> 80
<212> PRT
<213> Betatorquevirus sp.

<400> 1013
Met Pro Trp Trp Tyr Arg Arg Arg Ser Tyr Asn Pro Trp Arg Arg Arg
1   5   10   15

Asn Trp Phe Arg Arg Pro Arg Lys Thr Ile Tyr Arg Arg Tyr Arg Arg
            20   25   30

Arg Arg Arg Trp Asn Thr Asp Ser Arg Ser Gln His Lys His Lys Pro
        35   40   45

His Lys Arg Thr Arg Pro Arg Lys Lys Lys Lys Arg Ala Thr Ser Ser
    50   55   60

Ser Asp Ser Ser Asp Ser Glu Pro Ser Ser Ser Ser Ser Ser Ala Glu
65   70   75   80

<210> 1014
<211> 3853
<212> DNA
<213> Alphatorquevirus sp.

<400> 1014
ggcttagtgc gtcaccaccc acgtgacccg cctccgccaa ttaacaggta cttcgtacac   60

ttcctgggcg ggcttataag actaatataa gtagctgcac ttccgaatgg ctgagttttc   120

cacgcccgtc cgcagcggtg aagccacgga gggagctcag cgcgtcccga gggcgggtgc   180

cggaggtgag tttacacacc gcagtcaagg ggcaattcgg gctcgggact ggccgggctt   240

tgggcaaggc tcttaaaaaa gctatgttta ttggcaggca ctaccgaaag aaaagggcgc   300

tgctactgct atctgtgcat tctacaaaga caaaagggaa acttctaata gctatgtgga   360

ctcccccacg caatgatcaa caatacctta actggcaatg gtacacttct gtacttagct   420

cccactctgc tatgtgcggg tgttccgacg ctatcgctca tcttaatcat cttgctaatc   480

tgcttcgtgc cccgcaaaat ccgcccccgc ctgataatcc aagaccccta cccgtgcgag   540

cactgcctgc tcccccggct gcccacgagg cagccggtga tcgagcacca tggcctatgg 600

gtggtggagg agacgccgga ggcgctggcg caggtggaga cgccgaccat ggaggcgccg 660

ctggaggacc cgcagacgca gacctgctag acgccgtggc cgccgcagaa acgtaaggag 720

acggcgcaga gggaggtgga gaaggaggta caggaggtgg aaaagaaagg gcagacgtag 780

aagaaaagca aaaataataa taagacagtg gcagccaaac tacagaagaa gatgtaatat 840

agtgggctac ctccctatac ttatctgtgg tggaaatact gtttctagaa actatgccac 900

acactcagac gatactaact atccaggacc ctttgggggа ggcatgacca cagacaaatt 960

cagccttaga atactatatg atgaatacaa aagatttatg aactactgga cagcctcaaa 1020

tgaggaccta gatctctgta gatatctagg atgcactttt tacttcttta gacaccctga 1080

agtagacttt attataaaaa taaacaccat gcccccattc ttagatacaa ccataacagc 1140

acctagcata cacccaggcc tcatggccct agacaaaaga gccagatgga ttccttctct 1200

taaaaataga ccaggtaaaa aacactatat aaaaattaga gtaggggctc ctaaaatgtt 1260

cacagataaa tggtaccctc aaacagacct ctgtgacatg acactgctaa ctatctatgc 1320

aaccgcagcg gatatgcaat atccgttcgg ctcaccacta actgacactg tggttgttaa 1380

ctcccaagtt ctgcaatcca tgtatgatga aacaattagc atattacctg atgaaaaaac 1440

taaaagaaat agccttctta cttctataag aagctacata ccttttttata atactacaca 1500

aacaatagct caattaaaac catttgtaga tgcaggagga cacacaacag gctcaacaac 1560

aactacatgg ggacaactat taaacacaac taaatttacc actaccacaa caaccacata 1620

cacataccct ggcaccacaa atacagcagt aacatttata acagccaatg atacctggta 1680

caggggaaca gcatataaag ataacattaa agatgtacca caaaaagcag cacaattata 1740

ctttcaaaca acacaaaaac tactaggaaa cacattccat ggctcagatg aaacacttga 1800

ataccatgca ggcctataca gctctatctg gctatcacca ggtagatcct actttgaaac 1860

accaggtgca tacacagaca ttaaatataa ccctttttaca gacagaggag aaggcaacat 1920

gctgtggata gactggctaa gtaaaaaaaa catgaaatat gacaaagtgc aaagtaagtg 1980

cctagtagca gacctaccac tgtgggcagc agcatatggt tatgtagaat tctgctctaa 2040

aagcacagga gacacaaaca tacacatgaa tgccagacta ctaataagaa gtccttttac 2100

agacccccag ctaatagtac acacagaccc cactaaaggc tttgtaccct attctttaaa 2160

ctttggaaat ggtaaaatgc caggaggtag cagcaatgtt cccataagaa tgagagctaa 2220

gtggtacccc actttatccc accaacaaga agttctagag gccttagcac agtcaggacc 2280

ctttgcttat cactcagaca ttaaaaaagt atctctaggc ataaaatacc gttttaagtg 2340

gatctggggt ggaaaccccg ttcgccaaca ggttgttaga aatccctgca aggaacccca 2400

ctcctcgggc aatagagtcc ctagaagcat acaaatcgtt gacccgagat acaactcacc 2460

ggaacttacc atccatgcct gggacttcag acgtggcttc tttggcccga aagctattca 2520

aagaatgcaa caacaaccaa ctgctactga atttttttca gcaggccgca agagacccag 2580

aagggacaca gaagtgtatc agtccgacca agaaaaggag caaaaagaaa gctcgctttt 2640

ccccccagtc aagctcctcc gaagagtccc cccgtgggag gactcggaac aggagcaaag 2700

cgggtcgcaa agctcagagg aagagacggc gaccctctcc cagcagctca aacagcagct 2760

gcagcagcag cgagtcttgg gagtcaaact cagactcctg ttcaaccaag tccaaaaaat 2820

ccaacaaaat caagatatca accctacctt gttaccaagg gggggggatc tagtatcctt 2880

ctttcaggct gtaccataaa tatgtttcca gaccctaaac cttactgccc ctccagcaat 2940

gactggaaag aagagtatga ggcctgtaaa tattgggata gacctcccag acacaacctt 3000

agagaccccc ccttttaccc ctgggccccc aaaaacaatc cttgcaatgt aagctttaaa 3060

cttggcttca aataaactag gccgtgggag tttcacttgt cggtgtctac ctctataagt 3120

cactaagcac tccgagcgca gcgaggagtg cgacccttcc ccctggtgca acgccctcgg 3180

cggccgcgcg ctacgccttc ggctgcgcgc ggcacctcgg acccccgctc gtgctgacac 3240

gcttgcgcgt gtcagaccac ttcgggctcg cgggggtcgg gaaatttgct aaacagactc 3300

cgagttgcca ttggacactg tagctatgaa tcagtaacga aagtgagtgg ggccagactt 3360

cgccataagg cctttatctt cttgccattt gtcagtattg ggggtcgcca taaactttgg 3420

gctccatttt aggccttccg gactacaaaa atcgccatat ttgtgacgtc agagccgcca 3480

ttttaagtca gctctgggga ggcgtgactt ccagttcaaa ggtcatcctc accataactg 3540

gcacaaaatg gccgccaact tcttccgggt caaaggtcac tgctacgtca taggtgacgt 3600

gggggggggac ctacttaaac acggaagtag gccccgacac gtcactgtca cgtgacagta 3660

cgtcacagcc gccattttgt tttacaaaat agccgacttc cttcctcttt tttaaaaaaa 3720

ggcgccaaaa aaccgtcggc gggggggccg cgcgctgcgc gcgcggcccc cgggggaggc 3780

acagcctccc ccccccgcgc gcatgcgcgc gggtcccccc ccctccgggg ggctccgccc 3840

cccggccccc ccc 3853

<210> 1015
<211> 120
<212> PRT
<213> Alphatorquevirus sp.

<400> 1015
Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1 5 10 15

Tyr Thr Ser Val Leu Ser Ser His Ser Ala Met Cys Gly Cys Ser Asp
20 25 30

Ala Ile Ala His Leu Asn His Leu Ala Asn Leu Leu Arg Ala Pro Gln
35 40 45

Asn Pro Pro Pro Pro Asp Asn Pro Arg Pro Leu Pro Val Arg Ala Leu
50 55 60

Pro Ala Pro Pro Ala Ala His Glu Ala Ala Gly Asp Arg Ala Pro Trp
65 70 75 80

Pro Met Gly Gly Gly Gly Asp Ala Gly Gly Ala Gly Ala Gly Gly Asp
85 90 95

Ala Asp His Gly Gly Ala Ala Gly Gly Pro Ala Asp Ala Asp Leu Leu
100 105 110

Asp Ala Val Ala Ala Ala Glu Thr
115 120

<210> 1016
<211> 286
<212> PRT
<213> Alphatorquevirus sp.

<400> 1016
Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1   5   10   15

Tyr Thr Ser Val Leu Ser Ser His Ser Ala Met Cys Gly Cys Ser Asp
            20   25   30

Ala Ile Ala His Leu Asn His Leu Ala Asn Leu Leu Arg Ala Pro Gln
        35   40   45

Asn Pro Pro Pro Pro Asp Asn Pro Arg Pro Leu Pro Val Arg Ala Leu
    50   55   60

Pro Ala Pro Pro Ala Ala His Glu Ala Ala Gly Asp Arg Ala Pro Trp
65   70   75   80

Pro Met Gly Gly Gly Gly Asp Ala Gly Gly Ala Gly Ala Gly Gly Asp
                85   90   95

Ala Asp His Gly Gly Ala Ala Gly Gly Pro Ala Asp Ala Asp Leu Leu
            100   105   110

Asp Ala Val Ala Ala Ala Glu Thr Leu Leu Glu Ile Pro Ala Arg Asn
        115   120   125

Pro Thr Pro Arg Ala Ile Glu Ser Leu Glu Ala Tyr Lys Ser Leu Thr
    130   135   140

Arg Asp Thr Thr His Arg Asn Leu Pro Ser Met Pro Gly Thr Ser Asp
145   150   155   160

Val Ala Ser Leu Ala Arg Lys Leu Phe Lys Glu Cys Asn Asn Asn Gln
                165   170   175

```
Leu Leu Leu Asn Phe Phe Gln Gln Ala Ala Arg Asp Pro Glu Gly Thr
            180                 185                 190

Gln Lys Cys Ile Ser Pro Thr Lys Lys Arg Ser Lys Lys Lys Ala Arg
        195                 200                 205

Phe Ser Pro Gln Ser Ser Ser Ser Glu Glu Ser Pro Arg Gly Arg Thr
    210                 215                 220

Arg Asn Arg Ser Lys Ala Gly Arg Lys Ala Gln Arg Lys Arg Arg Arg
225                 230                 235                 240

Pro Ser Pro Ser Ser Ser Asn Ser Ser Cys Ser Ser Ser Glu Ser Trp
                245                 250                 255

Glu Ser Asn Ser Asp Ser Cys Ser Thr Lys Ser Lys Lys Ser Asn Lys
            260                 265                 270

Ile Lys Ile Ser Thr Leu Pro Cys Tyr Gln Gly Gly Gly Ile
        275                 280                 285
```

<210> 1017
<211> 289
<212> PRT
<213> Alphatorquevirus sp.

<400> 1017

```
Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1               5                   10                  15

Tyr Thr Ser Val Leu Ser Ser His Ser Ala Met Cys Gly Cys Ser Asp
            20                  25                  30

Ala Ile Ala His Leu Asn His Leu Ala Asn Leu Leu Arg Ala Pro Gln
        35                  40                  45

Asn Pro Pro Pro Pro Asp Asn Pro Arg Pro Leu Pro Val Arg Ala Leu
    50                  55                  60
```

```
Pro Ala Pro Pro Ala Ala His Glu Ala Ala Gly Asp Arg Ala Pro Trp
65                  70                  75                  80

Pro Met Gly Gly Gly Gly Asp Ala Gly Gly Ala Gly Ala Gly Gly Asp
                85                  90                  95

Ala Asp His Gly Gly Ala Ala Gly Gly Pro Ala Asp Ala Asp Leu Leu
            100                 105                 110

Asp Ala Val Ala Ala Ala Glu Thr Pro Gln Glu Thr Gln Lys Gly His
        115                 120                 125

Arg Ser Val Ser Val Arg Pro Arg Lys Gly Ala Lys Arg Lys Leu Ala
    130                 135                 140

Phe Pro Pro Ser Gln Ala Pro Pro Lys Ser Pro Pro Val Gly Gly Leu
145                 150                 155                 160

Gly Thr Gly Ala Lys Arg Val Ala Lys Leu Arg Gly Arg Asp Gly Asp
                165                 170                 175

Pro Leu Pro Ala Ala Gln Thr Ala Ala Ala Ala Ala Ala Ser Leu Gly
            180                 185                 190

Ser Gln Thr Gln Thr Pro Val Gln Pro Ser Pro Lys Asn Pro Thr Lys
        195                 200                 205

Ser Arg Tyr Gln Pro Tyr Leu Val Thr Lys Gly Gly Gly Ser Ser Ile
    210                 215                 220

Leu Leu Ser Gly Cys Thr Ile Asn Met Phe Pro Asp Pro Lys Pro Tyr
225                 230                 235                 240

Cys Pro Ser Ser Asn Asp Trp Lys Glu Glu Tyr Glu Ala Cys Lys Tyr
                245                 250                 255
```

```
Trp Asp Arg Pro Pro Arg His Asn Leu Arg Asp Pro Pro Phe Tyr Pro
            260                 265                 270

Trp Ala Pro Lys Asn Asn Pro Cys Asn Val Ser Phe Lys Leu Gly Phe
        275                 280                 285

Lys
```

<210> 1018
<211> 185
<212> PRT
<213> Alphatorquevirus sp.

<400> 1018

```
Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1               5                   10                  15

Pro Gln Glu Thr Gln Lys Gly His Arg Ser Val Ser Val Arg Pro Arg
            20                  25                  30

Lys Gly Ala Lys Arg Lys Leu Ala Phe Pro Pro Ser Gln Ala Pro Pro
        35                  40                  45

Lys Ser Pro Pro Val Gly Gly Leu Gly Thr Gly Ala Lys Arg Val Ala
    50                  55                  60

Lys Leu Arg Gly Arg Asp Gly Asp Pro Leu Pro Ala Ala Gln Thr Ala
65                  70                  75                  80

Ala Ala Ala Ala Ala Ser Leu Gly Ser Gln Thr Gln Thr Pro Val Gln
                85                  90                  95

Pro Ser Pro Lys Asn Pro Thr Lys Ser Arg Tyr Gln Pro Tyr Leu Val
            100                 105                 110

Thr Lys Gly Gly Gly Ser Ser Ile Leu Leu Ser Gly Cys Thr Ile Asn
        115                 120                 125
```

```
Met Phe Pro Asp Pro Lys Pro Tyr Cys Pro Ser Ser Asn Asp Trp Lys
    130                 135                 140

Glu Glu Tyr Glu Ala Cys Lys Tyr Trp Asp Arg Pro Pro Arg His Asn
145                 150                 155                 160

Leu Arg Asp Pro Pro Phe Tyr Pro Trp Ala Pro Lys Asn Asn Pro Cys
                165                 170                 175

Asn Val Ser Phe Lys Leu Gly Phe Lys
            180                 185
```

<210> 1019
<211> 105
<212> PRT
<213> Alphatorquevirus sp.

<400> 1019

```
Met Ile Asn Asn Thr Leu Thr Gly Asn Gly Thr Leu Leu Tyr Leu Ala
1               5                   10                  15

Pro Thr Leu Leu Cys Ala Gly Val Pro Thr Leu Ser Leu Ile Leu Ile
            20                  25                  30

Ile Leu Leu Ile Cys Phe Val Pro Arg Lys Ile Arg Pro Arg Leu Ile
        35                  40                  45

Ile Gln Asp Pro Tyr Pro Cys Glu His Cys Leu Leu Pro Arg Leu Pro
    50                  55                  60

Thr Arg Gln Pro Val Ile Glu His His Gly Leu Trp Val Val Glu Glu
65                  70                  75                  80

Thr Pro Glu Ala Leu Ala Gln Val Glu Thr Pro Thr Met Glu Ala Pro
                85                  90                  95

Leu Glu Asp Pro Gln Thr Gln Thr Cys
            100                 105
```

<210> 1020
<211> 769
<212> PRT
<213> Alphatorquevirus sp.

<400> 1020
Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Trp Arg Arg Trp
1   5   10   15

Arg Arg Arg Pro Trp Arg Arg Arg Trp Arg Thr Arg Arg Arg Arg Pro
            20   25   30

Ala Arg Arg Arg Gly Arg Arg Arg Asn Val Arg Arg Arg Arg Arg Gly
        35   40   45

Arg Trp Arg Arg Arg Tyr Arg Arg Trp Lys Arg Lys Gly Arg Arg Arg
    50   55   60

Arg Lys Ala Lys Ile Ile Ile Arg Gln Trp Gln Pro Asn Tyr Arg Arg
65   70   75   80

Arg Cys Asn Ile Val Gly Tyr Leu Pro Ile Leu Ile Cys Gly Gly Asn
                85   90   95

Thr Val Ser Arg Asn Tyr Ala Thr His Ser Asp Asp Thr Asn Tyr Pro
            100   105   110

Gly Pro Phe Gly Gly Gly Met Thr Thr Asp Lys Phe Ser Leu Arg Ile
        115   120   125

Leu Tyr Asp Glu Tyr Lys Arg Phe Met Asn Tyr Trp Thr Ala Ser Asn
    130   135   140

Glu Asp Leu Asp Leu Cys Arg Tyr Leu Gly Cys Thr Phe Tyr Phe Phe
145   150   155   160

Arg His Pro Glu Val Asp Phe Ile Ile Lys Ile Asn Thr Met Pro Pro
                165   170   175

```
Phe Leu Asp Thr Thr Ile Thr Ala Pro Ser Ile His Pro Gly Leu Met
            180                 185                 190

Ala Leu Asp Lys Arg Ala Arg Trp Ile Pro Ser Leu Lys Asn Arg Pro
        195                 200                 205

Gly Lys Lys His Tyr Ile Lys Ile Arg Val Gly Ala Pro Lys Met Phe
    210                 215                 220

Thr Asp Lys Trp Tyr Pro Gln Thr Asp Leu Cys Asp Met Thr Leu Leu
225                 230                 235                 240

Thr Ile Tyr Ala Thr Ala Ala Asp Met Gln Tyr Pro Phe Gly Ser Pro
                245                 250                 255

Leu Thr Asp Thr Val Val Val Asn Ser Gln Val Leu Gln Ser Met Tyr
            260                 265                 270

Asp Glu Thr Ile Ser Ile Leu Pro Asp Glu Lys Thr Lys Arg Asn Ser
        275                 280                 285

Leu Leu Thr Ser Ile Arg Ser Tyr Ile Pro Phe Tyr Asn Thr Thr Gln
    290                 295                 300

Thr Ile Ala Gln Leu Lys Pro Phe Val Asp Ala Gly Gly His Thr Thr
305                 310                 315                 320

Gly Ser Thr Thr Thr Thr Trp Gly Gln Leu Leu Asn Thr Thr Lys Phe
                325                 330                 335

Thr Thr Thr Thr Thr Thr Thr Tyr Thr Tyr Pro Gly Thr Thr Asn Thr
            340                 345                 350

Ala Val Thr Phe Ile Thr Ala Asn Asp Thr Trp Tyr Arg Gly Thr Ala
        355                 360                 365

Tyr Lys Asp Asn Ile Lys Asp Val Pro Gln Lys Ala Ala Gln Leu Tyr
    370                 375                 380
```

```
Phe Gln Thr Thr Gln Lys Leu Leu Gly Asn Thr Phe His Gly Ser Asp
385                 390                 395                 400

Glu Thr Leu Glu Tyr His Ala Gly Leu Tyr Ser Ser Ile Trp Leu Ser
                405                 410                 415

Pro Gly Arg Ser Tyr Phe Glu Thr Pro Gly Ala Tyr Thr Asp Ile Lys
            420                 425                 430

Tyr Asn Pro Phe Thr Asp Arg Gly Glu Gly Asn Met Leu Trp Ile Asp
        435                 440                 445

Trp Leu Ser Lys Lys Asn Met Lys Tyr Asp Lys Val Gln Ser Lys Cys
    450                 455                 460

Leu Val Ala Asp Leu Pro Leu Trp Ala Ala Ala Tyr Gly Tyr Val Glu
465                 470                 475                 480

Phe Cys Ser Lys Ser Thr Gly Asp Thr Asn Ile His Met Asn Ala Arg
                485                 490                 495

Leu Leu Ile Arg Ser Pro Phe Thr Asp Pro Gln Leu Ile Val His Thr
            500                 505                 510

Asp Pro Thr Lys Gly Phe Val Pro Tyr Ser Leu Asn Phe Gly Asn Gly
        515                 520                 525

Lys Met Pro Gly Gly Ser Ser Asn Val Pro Ile Arg Met Arg Ala Lys
    530                 535                 540

Trp Tyr Pro Thr Leu Ser His Gln Gln Glu Val Leu Glu Ala Leu Ala
545                 550                 555                 560

Gln Ser Gly Pro Phe Ala Tyr His Ser Asp Ile Lys Lys Val Ser Leu
                565                 570                 575
```

```
Gly Ile Lys Tyr Arg Phe Lys Trp Ile Trp Gly Gly Asn Pro Val Arg
            580                 585                 590

Gln Gln Val Val Arg Asn Pro Cys Lys Glu Pro His Ser Ser Gly Asn
        595                 600                 605

Arg Val Pro Arg Ser Ile Gln Ile Val Asp Pro Arg Tyr Asn Ser Pro
    610                 615                 620

Glu Leu Thr Ile His Ala Trp Asp Phe Arg Arg Gly Phe Phe Gly Pro
625                 630                 635                 640

Lys Ala Ile Gln Arg Met Gln Gln Gln Pro Thr Ala Thr Glu Phe Phe
                645                 650                 655

Ser Ala Gly Arg Lys Arg Pro Arg Arg Asp Thr Glu Val Tyr Gln Ser
            660                 665                 670

Asp Gln Glu Lys Glu Gln Lys Glu Ser Ser Leu Phe Pro Pro Val Lys
        675                 680                 685

Leu Leu Arg Arg Val Pro Pro Trp Glu Asp Ser Glu Gln Glu Gln Ser
    690                 695                 700

Gly Ser Gln Ser Ser Glu Glu Glu Thr Ala Thr Leu Ser Gln Gln Leu
705                 710                 715                 720

Lys Gln Gln Leu Gln Gln Gln Arg Val Leu Gly Val Lys Leu Arg Leu
                725                 730                 735

Leu Phe Asn Gln Val Gln Lys Ile Gln Gln Asn Gln Asp Ile Asn Pro
            740                 745                 750

Thr Leu Leu Pro Arg Gly Gly Asp Leu Val Ser Phe Phe Gln Ala Val
        755                 760                 765

Pro
```

<210> 1021
<211> 216
<212> PRT
<213> Alphatorquevirus sp.

<400> 1021
Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Trp Arg Arg Trp
1   5   10   15

Arg Arg Arg Pro Trp Arg Arg Arg Trp Arg Thr Arg Arg Arg Arg Pro
            20   25   30

Ala Arg Arg Arg Gly Arg Arg Arg Asn Val Val Arg Asn Pro Cys Lys
        35   40   45

Glu Pro His Ser Ser Gly Asn Arg Val Pro Arg Ser Ile Gln Ile Val
    50   55   60

Asp Pro Arg Tyr Asn Ser Pro Glu Leu Thr Ile His Ala Trp Asp Phe
65   70   75   80

Arg Arg Gly Phe Phe Gly Pro Lys Ala Ile Gln Arg Met Gln Gln Gln
                85   90   95

Pro Thr Ala Thr Glu Phe Phe Ser Ala Gly Arg Lys Arg Pro Arg Arg
            100   105   110

Asp Thr Glu Val Tyr Gln Ser Asp Gln Glu Lys Glu Gln Lys Glu Ser
        115   120   125

Ser Leu Phe Pro Pro Val Lys Leu Leu Arg Arg Val Pro Pro Trp Glu
    130   135   140

Asp Ser Glu Gln Glu Gln Ser Gly Ser Gln Ser Ser Glu Glu Glu Thr
145   150   155   160

Ala Thr Leu Ser Gln Gln Leu Lys Gln Gln Leu Gln Gln Gln Arg Val
                165   170   175

```
Leu Gly Val Lys Leu Arg Leu Leu Phe Asn Gln Val Gln Lys Ile Gln
            180                 185                 190


Gln Asn Gln Asp Ile Asn Pro Thr Leu Leu Pro Arg Gly Gly Asp Leu
        195                 200                 205


Val Ser Phe Phe Gln Ala Val Pro
    210                 215
```

<210> 1022
<211> 143
<212> PRT
<213> Alphatorquevirus sp.

<400> 1022

```
Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Trp Arg Arg Trp
1               5                   10                  15


Arg Arg Arg Pro Trp Arg Arg Arg Trp Arg Thr Arg Arg Arg Arg Pro
            20                  25                  30


Ala Arg Arg Arg Gly Arg Arg Arg Asn Ala Ala Arg Asp Pro Glu Gly
        35                  40                  45


Thr Gln Lys Cys Ile Ser Pro Thr Lys Lys Arg Ser Lys Lys Lys Ala
    50                  55                  60


Arg Phe Ser Pro Gln Ser Ser Ser Ser Glu Glu Ser Pro Arg Gly Arg
65                  70                  75                  80


Thr Arg Asn Arg Ser Lys Ala Gly Arg Lys Ala Gln Arg Lys Arg Arg
                85                  90                  95


Arg Pro Ser Pro Ser Ser Ser Asn Ser Ser Cys Ser Ser Ser Glu Ser
            100                 105                 110


Trp Glu Ser Asn Ser Asp Ser Cys Ser Thr Lys Ser Lys Lys Ser Asn
        115                 120                 125
```

Lys Ile Lys Ile Ser Thr Leu Pro Cys Tyr Gln Gly Gly Gly Ile
130 135 140

<210> 1023

<400> 1023
000

<210> 1024

<400> 1024
000

<210> 1025

<400> 1025
000

<210> 1026

<400> 1026
000

<210> 1027

<400> 1027
000

<210> 1028

<400> 1028
000

<210> 1029

<400> 1029
000

<210> 1030
<211> 1932

<212> DNA
<213> Adeno-associated virus

<400> 1030
atgccggggt tttacgagat tgtgattaag gtccccagcg accttgacga gcatctgccc 60

ggcatttctg acagctttgt gaactgggtg gccgagaagg aatgggagtt gccgccagat 120

tctgacatgg atctgaatct gattgagcag gcacccctga ccgtggccga gaagctgcag 180

cgcgactttc tgacggaatg gcgccgtgtg agtaaggccc cggaggccct tttctttgtg 240

caatttgaga agggagagag ctacttccac atgcacgtgc tcgtggaaac caccggggtg 300

aaatccatgg ttttgggacg tttcctgagt cagattcgcg aaaaactgat tcagagaatt 360

taccgcggga tcgagccgac tttgccaaac tggttcgcgg tcacaaagac cagaaatggc 420

gccggaggcg ggaacaaggt ggtggatgag tgctacatcc ccaattactt gctccccaaa 480

acccagcctg agctccagtg ggcgtggact aatatggaac agtatttaag cgcctgtttg 540

aatctcacgg agcgtaaacg gttggtggcg cagcatctga cgcacgtgtc gcagacgcag 600

gagcagaaca aagagaatca gaatcccaat tctgatgcgc cggtgatcag atcaaaaact 660

tcagccaggt acatggagct ggtcgggtgg ctcgtggaca aggggattac ctcggagaag 720

cagtggatcc aggaggacca ggcctcatac atctccttca atgcggcctc caactcgcgg 780

tcccaaatca aggctgcctt ggacaatgcg ggaaagatta tgagcctgac taaaaccgcc 840

cccgactacc tggtgggcca gcagcccgtg gaggacattt ccagcaatcg gatttataaa 900

attttggaac taaacgggta cgatccccaa tatgcggctt ccgtctttct gggatgggcc 960

acgaaaaagt tcggcaagag gaacaccatc tggctgtttg ggcctgcaac taccgggaag 1020

accaacatcg cggaggccat agcccacact gtgcccttct acgggtgcgt aaactggacc 1080

aatgagaact ttcccttcaa cgactgtgtc gacaagatgg tgatctggtg ggaggagggg 1140

aagatgaccg ccaaggtcgt ggagtcggcc aaagccattc tcggaggaag caaggtgcgc 1200

gtggaccaga aatgcaagtc ctcggcccag atagacccga ctcccgtgat cgtcacctcc 1260

aacaccaaca tgtgcgccgt gattgacggg aactcaacga ccttcgaaca ccagcagccg 1320

ttgcaagacc ggatgttcaa atttgaactc acccgccgtc tggatcatga ctttgggaag 1380

gtcaccaagc aggaagtcaa agacttttc cggtgggcaa aggatcacgt ggttgaggtg 1440

gagcatgaat tctacgtcaa aaagggtgga gccaagaaaa gacccgcccc cagtgacgca 1500

gatataagtg agcccaaacg ggtgcgcgag tcagttgcgc agccatcgac gtcagacgcg 1560

gaagcttcga tcaactacgc agacaggtac caaaacaaat gttctcgtca cgtgggcatg 1620

aatctgatgc tgtttccctg cagacaatgc gagagaatga atcagaattc aaatatctgc 1680

ttcactcacg gacagaaaga ctgtttagag tgctttcccg tgtcagaatc tcaacccgtt 1740

tctgtcgtca aaaaggcgta tcagaaactg tgctacattc atcatatcat gggaaaggtg 1800

ccagacgctt gcactgcctg cgatctggtc aatgtggatt tggatgactg catctttgaa 1860

caataaatga tttaaatcag gtatggctgc cgatggttat cttccagatt ggctcgagga 1920

cactctctct ga 1932

<210> 1031
<211> 621
<212> PRT
<213> Adeno-associated virus

<400> 1031
Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
1 5 10 15

Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
20 25 30

Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
35 40 45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
50 55 60

Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
65 70 75 80

Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu
85 90 95

```
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
            100                 105                 110

Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu
        115                 120                 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
    130                 135                 140

Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
145                 150                 155                 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu
                165                 170                 175

Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His
            180                 185                 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
        195                 200                 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
    210                 215                 220

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
225                 230                 235                 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
                245                 250                 255

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
            260                 265                 270

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
        275                 280                 285

Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
```

```
    290                 295                 300

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
305                 310                 315                 320

Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
                325                 330                 335

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
            340                 345                 350

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
        355                 360                 365

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
    370                 375                 380

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
385                 390                 395                 400

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
                405                 410                 415

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
            420                 425                 430

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
        435                 440                 445

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
    450                 455                 460

Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
465                 470                 475                 480

Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
                485                 490                 495
```

```
Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
            500                 505                 510

Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
        515                 520                 525

Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
    530                 535                 540

Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Ser Asn Ile Cys
545                 550                 555                 560

Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu
                565                 570                 575

Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr
            580                 585                 590

Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp
        595                 600                 605

Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln
    610                 615                 620
```

<210> 1032
<211> 536
<212> PRT
<213> Adeno-associated virus

<400> 1032

```
Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
1               5                   10                  15

Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
            20                  25                  30

Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
        35                  40                  45
```

```
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
    50                  55                  60

Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
65                  70                  75                  80

Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu
                85                  90                  95

Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
            100                 105                 110

Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu
        115                 120                 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
    130                 135                 140

Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
145                 150                 155                 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu
                165                 170                 175

Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His
            180                 185                 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
        195                 200                 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
    210                 215                 220

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
225                 230                 235                 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
```

```
                245                 250                 255

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
            260                 265                 270

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
        275                 280                 285

Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
    290                 295                 300

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
305                 310                 315                 320

Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
                325                 330                 335

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
            340                 345                 350

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
        355                 360                 365

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
    370                 375                 380

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
385                 390                 395                 400

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
                405                 410                 415

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
            420                 425                 430

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
        435                 440                 445
```

```
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
    450                 455                 460

Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
465                 470                 475                 480

Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
                485                 490                 495

Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
            500                 505                 510

Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
        515                 520                 525

Arg Leu Ala Arg Gly His Ser Leu
    530                 535
```

<210> 1033
<211> 397
<212> PRT
<213> Adeno-associated virus

<400> 1033

```
Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
1               5                   10                  15

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
            20                  25                  30

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
        35                  40                  45

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
    50                  55                  60

Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
65                  70                  75                  80
```

```
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
                85                  90                  95

Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
            100                 105                 110

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
        115                 120                 125

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
    130                 135                 140

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
145                 150                 155                 160

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
                165                 170                 175

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
            180                 185                 190

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
        195                 200                 205

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
    210                 215                 220

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
225                 230                 235                 240

Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
                245                 250                 255

Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
            260                 265                 270

Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
```

```
        275                 280                 285

Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
    290                 295                 300

Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
305                 310                 315                 320

Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Ser Asn Ile Cys
                325                 330                 335

Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu
            340                 345                 350

Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr
        355                 360                 365

Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp
    370                 375                 380

Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln
385                 390                 395


<210> 1034
<211> 312
<212> PRT
<213> Adeno-associated virus

<400> 1034
Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
1               5                   10                  15

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
            20                  25                  30

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
        35                  40                  45

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
```

```
    50                  55                  60

Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
65                  70                  75                  80

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
                85                  90                  95

Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
            100                 105                 110

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
        115                 120                 125

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
    130                 135                 140

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
145                 150                 155                 160

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
                165                 170                 175

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
            180                 185                 190

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
        195                 200                 205

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
    210                 215                 220

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
225                 230                 235                 240

Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
                245                 250                 255
```

Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
260 265 270

Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
275 280 285

Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
290 295 300

Arg Leu Ala Arg Gly His Ser Leu
305 310

<210> 1035
<211> 1941
<212> DNA
<213> Adeno-associated virus

<400> 1035
atgccggggt tctacgagat tgtcctgaag gtcccgagtg acctggacga gcgcctgccg 60

ggcatttcta actcgtttgt taactgggtg gccgagaagg aatgggacgt gccgccggat 120

tctgacatgg atccgaatct gattgagcag gcacccctga ccgtggccga aaagcttcag 180

cgcgagttcc tggtggagtg gcgccgcgtg agtaaggccc cggaggccct ctttttgtc 240

cagttcgaaa agggggagac ctacttccac ctgcacgtgc tgattgagac catcggggtc 300

aaatccatgg tggtcggccg ctacgtgagc cagattaaag agaagctggt gacccgcatc 360

taccgcgggg tcgagccgca gcttccgaac tggttcgcgg tgaccaaaac gcgaaatggc 420

gccgggggcg ggaacaaggt ggtggacgac tgctacatcc ccaactacct gctccccaag 480

acccagcccg agctccagtg ggcgtggact aacatggacc agtatttaag cgcctgtttg 540

aatctcgcgg agcgtaaacg gctggtggcg cagcatctga cgcacgtgtc gcagacgcag 600

gagcagaaca aagagaatca gaaccccaat tctgacgcgc cggtcatcag gtcaaaaacc 660

tcagccaggt acatggagct ggtcgggtgg ctggtggacc gcgggatcac gtcagaaaag 720

caatggattc aggaggacca ggcctcgtac atctccttca acgccgcctc caactcgcgg 780

tcccagatca aggccgcgct ggacaatgcc tccaagatca tgagcctgac aaagacggct 840

ccggactacc tggtgggcag caacccgccg gaggacatta ccaaaaatcg gatctaccaa 900

atcctggagc tgaacgggta cgatccgcag tacgcggcct ccgtcttcct gggctgggcg 960

caaaagaagt tcgggaagag gaacaccatc tggctctttg ggccggccac gacgggtaaa 1020

accaacatcg cggaagccat cgcccacgcc gtgcccttct acggctgcgt aaactggacc 1080

aatgagaact ttcccttcaa cgattgcgtc gacaagatgg tgatctggtg ggaggagggc 1140

aagatgacgg ccaaggtcgt ggagagcgcc aaggccattc tgggcggaag caaggtgcgc 1200

gtggaccaaa agtgcaagtc atcggcccag atcgaaccca ctcccgtgat cgtcacctcc 1260

aacaccaaca tgtgcgccgt gattgacggg aacagcacca ccttcgagca tcagcagccg 1320

ctgcaggacc ggatgtttga atttgaactt acccgccgtt tggaccatga ctttgggaag 1380

gtcaccaaac aggaagtaaa ggacttttc cggtgggctt ccgatcacgt gactgacgtg 1440

gctcatgagt tctacgtcag aaagggtgga gctaagaaac gccccgcctc caatgacgcg 1500

gatgtaagcg agccaaaacg ggagtgcacg tcacttgcgc agccgacaac gtcagacgcg 1560

gaagcaccgg cggactacgc ggacaggtac caaaacaaat gttctcgtca cgtgggcatg 1620

aatctgatgc tttttccctg taaaacatgc gagagaatga atcaaatttc caatgtctgt 1680

tttacgcatg gtcaaagaga ctgtggggaa tgcttccctg gaatgtcaga atctcaaccc 1740

gtttctgtcg tcaaaaagaa gacttatcag aaactgtgtc caattcatca tatcctggga 1800

agggcacccg agattgcctg ttcggcctgc gatttggcca atgtggactt ggatgactgt 1860

gtttctgagc aataaatgac ttaaaccagg tatggctgct gacggttatc ttccagattg 1920

gctcgaggac aacctttctg a 1941

<210> 1036
<211> 624
<212> PRT
<213> Adeno-associated virus

<400> 1036
Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
1 5 10 15

```
Glu Arg Leu Pro Gly Ile Ser Asn Ser Phe Val Asn Trp Val Ala Glu
            20                  25                  30

Lys Glu Trp Asp Val Pro Pro Asp Ser Asp Met Asp Pro Asn Leu Ile
        35                  40                  45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
    50                  55                  60

Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
65                  70                  75                  80

Gln Phe Glu Lys Gly Glu Thr Tyr Phe His Leu His Val Leu Ile Glu
                85                  90                  95

Thr Ile Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
            100                 105                 110

Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
        115                 120                 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
    130                 135                 140

Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
145                 150                 155                 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Leu
                165                 170                 175

Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
            180                 185                 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
        195                 200                 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
    210                 215                 220
```

```
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
225                 230                 235                 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
                245                 250                 255

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
            260                 265                 270

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
        275                 280                 285

Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
    290                 295                 300

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
305                 310                 315                 320

Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
                325                 330                 335

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
            340                 345                 350

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
        355                 360                 365

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
    370                 375                 380

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
385                 390                 395                 400

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
                405                 410                 415
```

```
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
            420                 425                 430

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe
        435                 440                 445

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
    450                 455                 460

Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
465                 470                 475                 480

Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
                485                 490                 495

Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
            500                 505                 510

Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
        515                 520                 525

Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
    530                 535                 540

Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Ile Ser Asn Val Cys
545                 550                 555                 560

Phe Thr His Gly Gln Arg Asp Cys Gly Glu Cys Phe Pro Gly Met Ser
                565                 570                 575

Glu Ser Gln Pro Val Ser Val Val Lys Lys Lys Thr Tyr Gln Lys Leu
            580                 585                 590

Cys Pro Ile His His Ile Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser
        595                 600                 605

Ala Cys Asp Leu Ala Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln
    610                 615                 620
```

<210> 1037
<211> 536
<212> PRT
<213> Adeno-associated virus

<400> 1037
Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
1   5   10   15

Glu Arg Leu Pro Gly Ile Ser Asn Ser Phe Val Asn Trp Val Ala Glu
            20   25   30

Lys Glu Trp Asp Val Pro Pro Asp Ser Asp Met Asp Pro Asn Leu Ile
        35   40   45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
    50   55   60

Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
65   70   75   80

Gln Phe Glu Lys Gly Glu Thr Tyr Phe His Leu His Val Leu Ile Glu
                85   90   95

Thr Ile Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
            100   105   110

Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
        115   120   125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
    130   135   140

Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
145   150   155   160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Leu
                165   170   175

```
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
            180                 185                 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
        195                 200                 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
    210                 215                 220

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
225                 230                 235                 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
                245                 250                 255

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
            260                 265                 270

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
        275                 280                 285

Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
    290                 295                 300

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
305                 310                 315                 320

Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
                325                 330                 335

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
            340                 345                 350

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
        355                 360                 365
```

```
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
    370                 375                 380

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
385                 390                 395                 400

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
                405                 410                 415

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
            420                 425                 430

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe
        435                 440                 445

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
    450                 455                 460

Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
465                 470                 475                 480

Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
                485                 490                 495

Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
            500                 505                 510

Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
        515                 520                 525

Arg Leu Ala Arg Gly Gln Pro Phe
    530                 535
```

<210> 1038
<211> 400
<212> PRT
<213> Adeno-associated virus

<400> 1038

```
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
1               5                  10                  15

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
            20                  25                  30

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
        35                  40                  45

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
    50                  55                  60

Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
65                  70                  75                  80

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
                85                  90                  95

Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
            100                 105                 110

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
        115                 120                 125

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
    130                 135                 140

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
145                 150                 155                 160

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
                165                 170                 175

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
            180                 185                 190

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
        195                 200                 205
```

```
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe
    210                 215                 220

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
225                 230                 235                 240

Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
                245                 250                 255

Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
            260                 265                 270

Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
        275                 280                 285

Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
    290                 295                 300

Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
305                 310                 315                 320

Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Ile Ser Asn Val Cys
                325                 330                 335

Phe Thr His Gly Gln Arg Asp Cys Gly Glu Cys Phe Pro Gly Met Ser
            340                 345                 350

Glu Ser Gln Pro Val Ser Val Val Lys Lys Lys Thr Tyr Gln Lys Leu
        355                 360                 365

Cys Pro Ile His His Ile Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser
    370                 375                 380

Ala Cys Asp Leu Ala Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln
385                 390                 395                 400
```

<210> 1039
<211> 312
<212> PRT
<213> Adeno-associated virus

<400> 1039
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
1   5   10   15

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
            20   25   30

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
        35   40   45

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
    50   55   60

Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
65   70   75   80

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
                85   90   95

Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
            100   105   110

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
        115   120   125

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
    130   135   140

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
145   150   155   160

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
                165   170   175

```
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
            180                 185                 190

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
        195                 200                 205

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe
    210                 215                 220

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
225                 230                 235                 240

Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
                245                 250                 255

Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
            260                 265                 270

Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
        275                 280                 285

Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
    290                 295                 300

Arg Leu Ala Arg Gly Gln Pro Phe
305                 310
```

<210> 1040
<211> 610
<212> PRT
<213> Adeno-associated virus

<400> 1040

```
Met Ala Thr Phe Tyr Glu Val Ile Val Arg Val Pro Phe Asp Val Glu
1               5                   10                  15

Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asp Trp Val Thr Gly
            20                  25                  30
```

```
Gln Ile Trp Glu Leu Pro Pro Glu Ser Asp Leu Asn Leu Thr Leu Val
        35                  40                  45

Glu Gln Pro Gln Leu Thr Val Ala Asp Arg Ile Arg Arg Val Phe Leu
    50                  55                  60

Tyr Glu Trp Asn Lys Phe Ser Lys Gln Glu Ser Lys Phe Phe Val Gln
65                  70                  75                  80

Phe Glu Lys Gly Ser Glu Tyr Phe His Leu His Thr Leu Val Glu Thr
                85                  90                  95

Ser Gly Ile Ser Ser Met Val Leu Gly Arg Tyr Val Ser Gln Ile Arg
            100                 105                 110

Ala Gln Leu Val Lys Val Val Phe Gln Gly Ile Glu Pro Gln Ile Asn
        115                 120                 125

Asp Trp Val Ala Ile Thr Lys Val Lys Lys Gly Gly Ala Asn Lys Val
    130                 135                 140

Val Asp Ser Gly Tyr Ile Pro Ala Tyr Leu Leu Pro Lys Val Gln Pro
145                 150                 155                 160

Glu Leu Gln Trp Ala Trp Thr Asn Leu Asp Glu Tyr Lys Leu Ala Ala
                165                 170                 175

Leu Asn Leu Glu Glu Arg Lys Arg Leu Val Ala Gln Phe Leu Ala Glu
            180                 185                 190

Ser Ser Gln Arg Ser Gln Glu Ala Ala Ser Gln Arg Glu Phe Ser Ala
        195                 200                 205

Asp Pro Val Ile Lys Ser Lys Thr Ser Gln Lys Tyr Met Ala Leu Val
    210                 215                 220

Asn Trp Leu Val Glu His Gly Ile Thr Ser Glu Lys Gln Trp Ile Gln
225                 230                 235                 240
```

```
Glu Asn Gln Glu Ser Tyr Leu Ser Phe Asn Ser Thr Gly Asn Ser Arg
                245                 250                 255

Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Thr Lys Ile Met Ser Leu
            260                 265                 270

Thr Lys Ser Ala Val Asp Tyr Leu Val Gly Ser Ser Val Pro Glu Asp
        275                 280                 285

Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met Asn Gly Tyr Asp
    290                 295                 300

Pro Ala Tyr Ala Gly Ser Ile Leu Tyr Gly Trp Cys Gln Arg Ser Phe
305                 310                 315                 320

Asn Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala Thr Thr Gly Lys
                325                 330                 335

Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro Phe Tyr Gly Cys
            340                 345                 350

Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp Cys Val Asp Lys
        355                 360                 365

Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Asn Lys Val Val Glu
    370                 375                 380

Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg Val Asp Gln Lys
385                 390                 395                 400

Cys Lys Ser Ser Val Gln Ile Asp Ser Thr Pro Val Ile Val Thr Ser
                405                 410                 415

Asn Thr Asn Met Cys Val Val Val Asp Gly Asn Ser Thr Thr Phe Glu
            420                 425                 430
```

```
His Gln Gln Pro Leu Glu Asp Arg Met Phe Lys Phe Glu Leu Thr Lys
        435                 440                 445

Arg Leu Pro Pro Asp Phe Gly Lys Ile Thr Lys Gln Glu Val Lys Asp
    450                 455                 460

Phe Phe Ala Trp Ala Lys Val Asn Gln Val Pro Val Thr His Glu Phe
465                 470                 475                 480

Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala Glu Lys Ser Leu
                485                 490                 495

Lys Arg Pro Leu Gly Asp Val Thr Asn Thr Ser Tyr Lys Ser Leu Glu
            500                 505                 510

Lys Arg Ala Arg Leu Ser Phe Val Pro Glu Thr Pro Arg Ser Ser Asp
        515                 520                 525

Val Thr Val Asp Pro Ala Pro Leu Arg Pro Leu Asn Trp Asn Ser Arg
    530                 535                 540

Tyr Asp Cys Lys Cys Asp Tyr His Ala Gln Phe Asp Asn Ile Ser Asn
545                 550                 555                 560

Lys Cys Asp Glu Cys Glu Tyr Leu Asn Arg Gly Lys Asn Gly Cys Ile
                565                 570                 575

Cys His Asn Val Thr His Cys Gln Ile Cys His Gly Ile Pro Pro Trp
            580                 585                 590

Glu Lys Glu Asn Leu Ser Asp Phe Gly Asp Phe Asp Asp Ala Asn Lys
        595                 600                 605

Glu Gln
    610
```

<210> 1041
<211> 390

<212> PRT
<213> Adeno-associated virus

<400> 1041
Met Ala Leu Val Asn Trp Leu Val Glu His Gly Ile Thr Ser Glu Lys
1   5   10   15

Gln Trp Ile Gln Glu Asn Gln Glu Ser Tyr Leu Ser Phe Asn Ser Thr
            20   25   30

Gly Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Thr Lys
        35   40   45

Ile Met Ser Leu Thr Lys Ser Ala Val Asp Tyr Leu Val Gly Ser Ser
    50   55   60

Val Pro Glu Asp Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met
65   70   75   80

Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Ile Leu Tyr Gly Trp Cys
                85   90   95

Gln Arg Ser Phe Asn Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala
            100   105   110

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
        115   120   125

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
    130   135   140

Cys Val Asp Lys Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Asn
145   150   155   160

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
                165   170   175

Val Asp Gln Lys Cys Lys Ser Ser Val Gln Ile Asp Ser Thr Pro Val
            180   185   190

```
Ile Val Thr Ser Asn Thr Asn Met Cys Val Val Val Asp Gly Asn Ser
        195                 200                 205

Thr Thr Phe Glu His Gln Gln Pro Leu Glu Asp Arg Met Phe Lys Phe
    210                 215                 220

Glu Leu Thr Lys Arg Leu Pro Pro Asp Phe Gly Lys Ile Thr Lys Gln
225                 230                 235                 240

Glu Val Lys Asp Phe Phe Ala Trp Ala Lys Val Asn Gln Val Pro Val
                245                 250                 255

Thr His Glu Phe Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala
            260                 265                 270

Glu Lys Ser Leu Lys Arg Pro Leu Gly Asp Val Thr Asn Thr Ser Tyr
        275                 280                 285

Lys Ser Leu Glu Lys Arg Ala Arg Leu Ser Phe Val Pro Glu Thr Pro
    290                 295                 300

Arg Ser Ser Asp Val Thr Val Asp Pro Ala Pro Leu Arg Pro Leu Asn
305                 310                 315                 320

Trp Asn Ser Arg Tyr Asp Cys Lys Cys Asp Tyr His Ala Gln Phe Asp
                325                 330                 335

Asn Ile Ser Asn Lys Cys Asp Glu Cys Glu Tyr Leu Asn Arg Gly Lys
            340                 345                 350

Asn Gly Cys Ile Cys His Asn Val Thr His Cys Gln Ile Cys His Gly
        355                 360                 365

Ile Pro Pro Trp Glu Lys Glu Asn Leu Ser Asp Phe Gly Asp Phe Asp
    370                 375                 380
```

Asp Ala Asn Lys Glu Gln
385 390

<210> 1042
<211> 341
<212> PRT
<213> Adeno-associated virus

<400> 1042
Met Ser Leu Thr Lys Ser Ala Val Asp Tyr Leu Val Gly Ser Ser Val
1 5 10 15

Pro Glu Asp Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met Asn
20 25 30

Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Ile Leu Tyr Gly Trp Cys Gln
35 40 45

Arg Ser Phe Asn Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala Thr
50 55 60

Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro Phe
65 70 75 80

Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp Cys
85 90 95

Val Asp Lys Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Asn Lys
100 105 110

Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg Val
115 120 125

Asp Gln Lys Cys Lys Ser Ser Val Gln Ile Asp Ser Thr Pro Val Ile
130 135 140

Val Thr Ser Asn Thr Asn Met Cys Val Val Val Asp Gly Asn Ser Thr
145 150 155 160

```
Thr Phe Glu His Gln Gln Pro Leu Glu Asp Arg Met Phe Lys Phe Glu
                165                 170                 175

Leu Thr Lys Arg Leu Pro Pro Asp Phe Gly Lys Ile Thr Lys Gln Glu
            180                 185                 190

Val Lys Asp Phe Phe Ala Trp Ala Lys Val Asn Gln Val Pro Val Thr
        195                 200                 205

His Glu Phe Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala Glu
    210                 215                 220

Lys Ser Leu Lys Arg Pro Leu Gly Asp Val Thr Asn Thr Ser Tyr Lys
225                 230                 235                 240

Ser Leu Glu Lys Arg Ala Arg Leu Ser Phe Val Pro Glu Thr Pro Arg
                245                 250                 255

Ser Ser Asp Val Thr Val Asp Pro Ala Pro Leu Arg Pro Leu Asn Trp
            260                 265                 270

Asn Ser Arg Tyr Asp Cys Lys Cys Asp Tyr His Ala Gln Phe Asp Asn
        275                 280                 285

Ile Ser Asn Lys Cys Asp Glu Cys Glu Tyr Leu Asn Arg Gly Lys Asn
    290                 295                 300

Gly Cys Ile Cys His Asn Val Thr His Cys Gln Ile Cys His Gly Ile
305                 310                 315                 320

Pro Pro Trp Glu Lys Glu Asn Leu Ser Asp Phe Gly Asp Phe Asp Asp
                325                 330                 335

Ala Asn Lys Glu Gln
            340
```

<210> 1043

<400> 1043
000

<210> 1044

<400> 1044
000

<210> 1045

<400> 1045
000

<210> 1046

<400> 1046
000

<210> 1047

<400> 1047
000

<210> 1048

<400> 1048
000

<210> 1049

<400> 1049
000

<210> 1050

<400> 1050
000

<210> 1051
<211> 62
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1051
aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctcactgagg 60

cc 62

<210> 1052
<211> 62
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1052
cgggcgggtg gtggcggcgg ttggggctcg gcgctcgctc gctcgctggg cgggcgggcg 60

gt 62

<210> 1053
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
peptide

<400> 1053
Gly Met Gly Tyr Gly Met Gly Tyr Gly Met Gly Tyr Gly Met Gly Tyr
1 5 10 15

<210> 1054
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
peptide

<400> 1054

Gly Met Gly Tyr Gly Met Gly Tyr Gly Met Gly Tyr Gly Met Gly Tyr
1 5 10 15

Gly Met Gly Tyr
20

<210> 1055
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1055
gcgcgctcgc tcgctc 16

<210> 1056
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1056
gctcgctcgc tcgctg 16

<210> 1057
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<220>
<221> modified_base
<222> (1)..(1)
<223> a, c, t, or g

<400> 1057

ngttgg 6

<210> 1058
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1058
agttgg 6

<210> 1059
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1059
ggttgg 6

<210> 1060
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
peptide

<220>
<221> MOD_RES
<222> (4)..(5)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (7)..(7)
<223> Any amino acid

<220>

<221> MOD_RES
<222> (9)..(10)
<223> Any amino acid

<400> 1060
Tyr Asn Pro Xaa Xaa Asp Xaa Gly Xaa Xaa Asn
1   5   10

<210> 1061
<211> 22
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      peptide

<220>
<221> MOD_RES
<222> (6)..(7)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(9)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (15)..(16)
<223> Any amino acid

<400> 1061
Tyr Asn Cys Ser Pro Xaa Xaa Asp Xaa Gly Ala Ser Lys Arg Xaa Xaa
1   5   10   15

Asn Thr Ser Val Ala Lys
            20

<210> 1062
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220>
<221> modified_base
<222> (45)..(45)
<223> a, c, t, g, unknown or other

<220>
<223> See specification as filed for detailed description of substitutions and preferred embodiments

<400> 1062
aggtgagtga aaccaccgaa gtcaaggggc aattcgggct agggncagtc t 51

<210> 1063
<211> 50
<212> DNA
<213> Alphatorquevirus sp.

<400> 1063
aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 50

<210> 1064
<211> 50
<212> DNA
<213> Betatorquevirus sp.

<400> 1064
aggtgagtga aaccaccgaa gtcaaggggc aattcgggct agatcagtct 50

<210> 1065
<211> 50
<212> DNA
<213> Gammatorquevirus sp.

<400> 1065
aggtgagtga aaccaccgag gtctaggggc aattcgggct agggcagtct 50

<210> 1066
<211> 150
<212> PRT
<213> Betatorquevirus sp.

<400> 1066
Met Pro Trp Trp Tyr Arg Arg Arg Ser Tyr Asn Pro Trp Arg Arg Arg

1 5 10 15

Asn Trp Phe Arg Arg Pro Arg Lys Thr Ile Tyr Arg Arg Tyr Arg Arg
20 25 30

Arg Arg Arg Trp Pro Thr Tyr Val Val Pro Asn Asn Phe Asn Glu Thr
35 40 45

Thr Ser Leu Gln Asn Pro Thr Thr Arg Pro Glu His Phe Leu Tyr Ser
50 55 60

Phe Asp Glu Arg Arg Gly Gln Leu Thr Glu Lys Ala Thr Lys Arg Leu
65 70 75 80

Leu Lys Asp Trp Glu Thr Lys Glu Thr Ser Leu Leu Ser Thr Glu Tyr
85 90 95

Arg Phe Ala Glu Pro Thr Gln Thr Gln Ala Pro Gln Glu Asp Pro Ser
100 105 110

Ser Glu Glu Glu Glu Glu Ser Asn Leu Phe Glu Arg Leu Leu Arg Gln
115 120 125

Arg Thr Lys Gln Leu Gln Leu Lys Arg Arg Ile Ile Gln Thr Leu Lys
130 135 140

Asp Leu Gln Lys Leu Glu
145 150

<210> 1067
<211> 62
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1067
aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctcactgagg 60