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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 June 2020 (18.06.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/123795 A3

(51) International Patent Classification:
C12N 15/86 (2006.01)

(21) International Application Number:
PCT/US2019/065963

(22) International Filing Date:
12 December 2019 (12.12.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/778,866 12 December 2018 (12.12.2018) US
62/778,861 12 December 2018 (12.12.2018) US

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

(72) Inventors: **WEINSTEIN, Erica Gabrielle**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **KAHVEJIAN, Avak**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **DELAGRAVE, Simon**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **YOZWIAK, Nathan, Lawrence**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, GA 02142 (US). **LEBO, Kevin, James**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **DIAZ, Fernando, Martin**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **NAWANDAR, Dhananjay Maniklal**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **TEDSTONE, Ryan D.**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US). **PITTS, Jared David**; 55 Cambridge Parkway, 8th Floor, Suite 800E, Cambridge, MA 02142 (US).

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

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

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

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(88) Date of publication of the international search report:
27 August 2020 (27.08.2020)

(54) Title: ANELLOSOMES FOR DELIVERING PROTEIN REPLACEMENT THERAPEUTIC MODALITIES

(57) Abstract: This invention relates generally to anellosomes and compositions and uses thereof.

WO 2020/123795 A3

ANELLOSOMES FOR DELIVERING PROTEIN REPLACEMENT THERAPEUTIC MODALITIES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Nos. 62/778,861, filed December 12, 2018, and 62/778,866, filed December 12, 2018. The contents of the aforementioned applications are hereby incorporated by reference in their entirety.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on December 9, 2019, is named V2057-7003WO_SL.txt and is 825,844 bytes in size.

BACKGROUND

There is an ongoing need to develop suitable vectors to deliver therapeutic genetic material to patients.

SUMMARY

The present disclosure provides an anellosome, e.g., a synthetic anellosome, that can be used as a delivery vehicle, e.g., for delivering genetic material, for delivering an effector, e.g., a payloador for delivering a therapeutic agent or therapeutic effector (e.g., a protein replacement therapeutic) to a eukaryotic cell (e.g., a human cell or a human tissue). Exemplary protein replacement therapeutics that can be delivered include, e.g., enzymes, or structural proteins. The anellosomes can be used to deliver a protein to a subject having disease caused by a mutation, e.g., delivering a functional version of the mutant protein.

In some embodiments, an anellosome (e.g., particle, e.g., a viral particle, e.g., an *Anellovirus* particle) comprises a genetic element (e.g., a genetic element comprising a therapeutic DNA sequence) encapsulated in a proteinaceous exterior (e.g., a proteinaceous exterior comprising an *Anellovirus* capsid protein, e.g., an Anellovirus ORF1 protein or a polypeptide encoded by an *Anellovirus* ORF1 nucleic acid, e.g., as described herein), which is capable of introducing the genetic element into a cell (e.g., a mammalian cell, e.g., a human cell). In some embodiments, the anellosome is a particle comprising a proteinaceous exterior comprising a polypeptide encoded by an *Anellovirus* ORF1 nucleic acid (e.g., an ORF1 nucleic acid of *Alphatorquevirus, Betatorquevirus,* or *Gammatorquevirus*, e.g., an ORF1 of

Alphatorquevirus clade 1, *Alphatorquevirus* clade 2, *Alphatorquevirus* clade 3, *Alphatorquevirus* clade 4, *Alphatorquevirus* clade 5, *Alphatorquevirus* clade 6, or *Alphatorquevirus* clade 7, e.g., as described herein). The genetic element of an anellosome of the present disclosure is typically a circular and/or single-stranded DNA molecule (e.g., circular and single stranded), and generally 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 instances, the genetic element is circular or linear. In some instances, the genetic element 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 instances, 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 anellosome 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 anellosome). In other instances, the effector is an exogenous effector. The effector can, in some instances, modulate a function of the cell or modulate an activity or level of a target molecule in the cell. For example, the effector can decrease levels of a target protein in the cell (e.g., as described in Examples 3 and 4). In another example, the anellosome can deliver and express an effector, e.g., an exogenous protein, *in vivo* (e.g., as described in Examples 19 and 28). Anellosomes can be used, for example, to deliver genetic material to a target cell, tissue or subject; to deliver an effector to a target cell, tissue or subject; or for treatment of diseases and disorders, e.g., by delivering an effector that can operate as a therapeutic agent to a desired cell, tissue, or subject.

The invention further provides synthetic anellosomes. A synthetic anellosome 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. Generally, synthetic anellosomes 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 anellosome 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 anellosome may be substantially non-immunogenic to the target cell, tissue or subject.

In an aspect, the invention features an anellosome comprising: (a) a proteinaceous exterior; (b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an exogenous effector, and a protein binding sequence (e.g., an exterior protein binding sequence); wherein the exogenous effector comprises a polypeptide that, when mutated, causes a human disease, or a functional variant of said polypeptide; wherein the genetic element is enclosed within the proteinaceous exterior; and wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell. Optionally, the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, modification (e.g., enzymatic modification), and/or deletion, e.g., a deletion of a domain or portion thereof (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

In an aspect, the invention features an anellosome 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 anellosome 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 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 anellosome 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 listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17). 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, e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17). 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 an aspect, the invention features an infectious (to a human cell) particle comprising an Anellovirus capsid (e.g., a capsid comprising an Anellovirus ORF, e.g., ORF1, polypeptide) encapsulating a genetic element comprising a protein binding sequence that binds to the capsid and a heterologous (to the Anellovirus) sequence encoding a therapeutic effector. In embodiments, the particle 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 6%, 5.5%, 5%, 4.5%, 4%, 3.5%, 3%, 2.5%, 2%, 1.5%, or less) identity to a wild type Anellovirus. 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. 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. In some embodiments, integration frequency is determined as described in Wang et al. (2004, *Gene Therapy* 11: 711-721, incorporated herein by reference in its entirety).

Also described herein are viral vectors and viral particles based on Anelloviruses, which can be used to deliver an agent (e.g., an exogenous effector or an endogenous effector, e.g., a therapeutic effector) to a cell (e.g., a cell in a subject to be treated therapeutically). In some embodiments, Anelloviruses 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 an aspect, the invention features 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., amino acid sequence having 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),

(ii) a second region comprising a jelly-roll domain, e.g., an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to a jelly-roll region sequence described herein or a sequence comprising at least 6 beta strands,

(iii) a third region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an N22 domain sequence described herein,

(iv) a fourth region comprising an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to 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 described herein.

In some embodiments, the polypeptide comprises at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100% sequence identity to an Anellovirus ORF1 molecule 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, 18, 20-37, or D1-D10). In some embodiments, the polypeptide comprises at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100% sequence identity to a subsequence (e.g., an arginine (Arg)-rich domain, a jelly-roll domain, a hypervariable region (HVR), an N22 domain, or a C-terminal domain (CTD)) of an Anellovirus ORF1 molecule 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, 18, 20-37, or D1-D10). In one embodiment, the amino acid sequences of the (i), (ii), (iii), and (iv) region have at least 90% sequence identity to their respective references and 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 described herein.

In an aspect, the invention features a complex comprising a polypeptide as described herein (e.g., an Anellovirus ORF1 molecule as described herein) and a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence.

The present disclosure further provides nucleic acid molecules (e.g., a nucleic acid molecule that includes a genetic element as described herein, or a nucleic acid molecule that includes a sequence encoding a proteinaceous exterior protein as described herein). A nucleic acid molecule of the invention may include one or both of (a) a genetic element as described herein, and (b) a nucleic acid sequence encoding a proteinaceous exterior protein as described herein.

In an aspect, the invention features an isolated nucleic acid molecule comprising a genetic element comprising a promoter element operably 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, 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 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, the genetic element comprises an anellovector, e.g., as described herein. 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 or helper cell comprising: (a) a nucleic acid 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), wherein the nucleic acid is a plasmid, is a viral nucleic acid, or is integrated into a helper cell 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 polypeptide of (a), wherein optionally the genetic element does not encode an ORF1 polypeptide (e.g., an ORF1 protein). For example, the host cell or helper 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 genetic element of (b) is circular, single-stranded DNA. In some embodiments, the host cell is a manufacturing cell line. In some embodiments, the host cell or helper cell is adherent or in suspension, or both. In some embodiments, the host cell or helper cell is grown in a microcarrier. In some mbodiments, 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 anellosomes by the host cell or helper cell.

In an aspect, the invention features a pharmaceutical composition comprising an anellosome (e.g., a synthetic anellosome) as 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} genome equivalents of the anellosome 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 anellosome 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 anellosome, e.g., a synthetic anellosome, 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 anellosome, e.g., a synthetic anellosome, e.g., as described herein, wherein the anellosome 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 anellosome to a cell, comprising contacting the anellosome, e.g., a synthetic anellosome, 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 anellosome, e.g., a synthetic anellosome. 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 anellosome, e.g., a synthetic anellosome, as described herein, and

(ii) the first nucleic acid or 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 listed in any of Table 16, 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 to make the anellosome.

In some embodiments, the method further includes, prior to step (a), 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 a helper (e.g., a helper plasmid or the genome of a helper virus).In another aspect, the invention features a method of manufacturing an anellosome composition, comprising:

a) providing a host cell comprising, e.g., expressing one or more components (e.g., all of the components) of an anellosome, e.g., a synthetic anellosome, e.g., as described herein. For example, the host cell 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 is integrated into a helper cell 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 (i) a protein binding sequence (e.g, packaging sequence) that binds the polypeptide of (a), wherein the host cell or helper 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;

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

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

In some embodiments, the components of the anellosome 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 anellosome (e.g., wherein one or more nucleic acids encoding the components of the anellosome are introduced into the host cell, or a progenitor thereof, e.g., by stable transfection).

In some embodiments, the method further comprises one or more purification steps (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 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.

In an aspect, the invention features a method of manufacturing an anellosome composition, comprising: a) providing a plurality of anellosomes described herein, or a preparation of anellosomes described herein; and b) formulating the anellosomes 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), e.g., a population of first host cells, comprising an anellosome, the method comprising introducing a genetic element, e.g., as described herein, to a host cell and culturing the host cell under conditions suitable for production of the anellosome. 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 anellosome.

In an aspect, the invention features a method of making an anellosome, comprising providing a host cell, e.g., a first host cell or producer cell (e.g., as shown in Figure 12), comprising an anellosome, e.g., as described herein, and purifying the anellosome from the host cell. In some embodiments, the method further comprises, prior to the providing step, contacting the host cell with an anellosome, e.g., as described herein, and incubating the host cell under conditions suitable for production of the anellosome. 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 anellosome from the host cell comprises lysing the host cell.

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

comprises lysing the third host cell. In embodiments, at least about 2-100-fold more of the anellosome 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 anellosomes by the host cell. In some embodiments, anellosomes 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, anellosomes 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 anellosome preparation. The method comprises (a) making an anellosome preparation as described herein, (b) evaluating the preparation (e.g., a pharmaceutical anellosome preparation, anellosome seed population or the anellosome stock population) for one or more pharmaceutical quality control parameters, e.g., identity, purity, titer, potency (e.g., in genomic equivalents per anellosome particle), and/or the nucleic acid sequence, e.g., from the genetic element comprised by the anellosome, 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 anellosome, 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 anellosomes (e.g., an anellosome other than the desired anellosome, e.g., a synthetic anellosome 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) anellosomes in the preparation (e.g., as evaluated by HPLC). In some embodiments, evaluating potency comprises evaluating the level of anellosome 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 anellosomes can be produced in a single batch. In embodiments, the levels of the anellosomes 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 the nucleic acid sequence of a genetic element of an anellosome 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 as listed in any of Table 16, 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 anellosome described herein and a helper virus, wherein the helper virus comprises a polynucleotide, e.g., 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 anellosome (e.g., a synthetic anellosome) 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 anellosome (e.g., a synthetic anellosome) is purified, e.g., from a solution (e.g., a supernatant). In some embodiments, an anellosome is enriched in a solution relative to other constituents in the solution.

In some embodiments of any of the aforesaid anellosomes, anellovectors, compositions or methods, the genetic element comprises an anellosome genome, e.g., as identified according to the method described in Example 9. In embodiments, the anellosome genome comprises a TTV-tth8 nucleic acid sequence, e.g., a TTV-tth8 nucleic acid sequence shown in Table 5, having deletions of at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or 100% of nucleotides 3436-3707 of the TTV-tth8 nucleic acid sequence. In embodiments, the anellosome genome comprises a TTMV-LY2 nucleic acid sequence, e.g., a TTMV-LY2 nucleic acid sequence shown in Table 15, having deletions of at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or 100% of nucleotides 574-1371, 1432-2210, 574-2210, and/or 2610-2809 of the TTMV-LY2 nucleic acid sequence. In embodiments, the anellosome genome is an anellosome genome capable of self-replication and/or self-amplification. In embodiments, the anellosome genome is not capable of self-replication and/or self-amplification. In embodiments, the anellosome genome 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 anellosomes, 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. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an exogenous effector, and a protein binding sequence (e.g., an exterior protein binding sequence);

wherein the exogenous effector comprises a polypeptide that, when mutated, causes a human disease, or a functional variant of said polypeptide;

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

2. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element operably linked to a heterologous nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector);

wherein the exogenous effector comprises a polypeptide that, when mutated, causes a human disease, or a functional variant of said polypeptide;

wherein the genetic element is enclosed within the proteinaceous exterior;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

3. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an exogenous effector, and a protein binding sequence (e.g., an exterior protein binding sequence);

wherein the exogenous effector comprises an enzyme or a structural protein;

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

4. The anellosome of any of the preceding embodiments, wherein the effector comprises an enzyme that acts on a small molecule substrate that is endogenous to a human cell.

5. The anellosome of any of the preceding embodiments, wherein the effector comprises a polypeptide that is endogenous to neurons or muscle cells, or a functional variant of said polypeptide.

6. The anellosome of any of the preceding embodiments, wherein the effector comprises an enzyme, a structural protein, or a complement protein.

7. The anellosome of any of the preceding embodiments, wherein the effector is an enzyme of Table A or a functional variant thereof, or wherein the effector is a polypeptide of Table B, or a functional variant thereof.

8. The anellosome of any of the preceding embodiments, wherein the effector comprises a wild-type human protein, e.g., wild-type SMN, dystrophin, UGT1A1, OCA1, or C1.

9. The anellosome of any of the preceding embodiments, wherein the human disease is a disease that leads to elevated levels of a substrate compared to level of the substrate in a healthy subject.

10. The anellosome of any of the preceding embodiments, wherein the human disease is a disease caused by lowered levels of a structural protein compared to level of the structural protein in a healthy subject.

11. A method of treating a disease or disorder in a subject, the method comprising administering an effective amount of an anellosome composition to the subject, wherein the anellosome composition comprises a plurality of anellosomes that comprise:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence);

wherein the effector comprises a polypeptide that, when mutated, causes a human disease, or a functional variant of said polypeptide;

wherein the genetic element is enclosed within the proteinaceous exterior; and

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

12. The method of embodiment 11, wherein the disease or disorder is a genetic disease or disorder, e.g., a genetic disease other than an interferonopathy.

13. The method of embodiment 11, wherein the disease or disorder is a neuromuscular disease (e.g., spinal muscular atrophy), a muscular dystrophy (e.g., Duchenne muscular dystrophy or Becker muscular dystrophy), an enzymatic insufficiency (e.g., Crigler–Najjar syndrome), or albinism (e.g., ocular albinism), or C1 deficiency.

14. A method of treating a genetic disease or disorder in a subject, the method comprising administering an effective amount of an anellosome composition to the subject, wherein the anellosome composition comprises a plurality of anellosomes that comprise:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence);

wherein the genetic element is enclosed within the proteinaceous exterior; and

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

15. The method of embodiment 14, wherein the effector comprises an enzyme, a structural protein, or a complement protein.

16. A method of delivering an enzymatic activity to a subject in need of the enzymatic activity, the method comprising administering an effective amount of an anellosome composition to the subject,

wherein the anellosome composition comprises a plurality of anellosomes that comprise:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence);

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

thereby delivering the enzymatic activity to the subject.

17. A method of restoring (at least partially) a function in a subject, e.g., to a wild-type phenotype, the method comprising administering an effective amount of an anellosome composition to the subject,

wherein the anellosome composition comprises a plurality of anellosomes that comprise:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence);

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g.,

as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

thereby restoring the function in the subject.

18. The method of any of embodiments 11-17, wherein the effector covalently modifies a substrate that is present at an elevated level in the subject compared to level of the substrate in a healthy subject.

19. The method of any of embodiments 11-18, wherein the effector is a functional variant of the protein that is mutant in the subject, e.g., is a wild-type protein corresponding to the protein that is mutant in the subject.

20. The method of any of embodiments 11-19, wherein the effector is a functional homolog (e.g., ortholog or paralog) of the protein that is mutant in the subject.

21. The method of any of embodiments 11-20, wherein the effector covalently modifies a substrate, wherein the substrate is also modified by a wild-type protein corresponding to the protein that is mutant in the subject.

22. The method of any of embodiments 11-21, wherein the effector catalyzes the same chemical reaction as the wild-type protein corresponding to the protein that is mutant in the subject.

23. The method of any of embodiments 11-22, wherein the effector comprises a structural protein, an enzyme, or a complement protein.

24. A method of delivering an effector to a subject, the method comprising administering an effective amount of an anellosome composition to the subject,

wherein the anellosome composition comprises a plurality of anellosomes that comprise:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence);

wherein the effector comprises a polypeptide that, when mutated, causes a human disease, or a functional variant of said polypeptide;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

thereby delivering the effector to the subject.

25. A method of modulating, e.g., inhibiting or enhancing, a biological function in a subject, e.g., a subject having a disease or disorder treatable by modulating the biological function in the subject, the method comprising administering an effective amount of an anellosome composition to the subject,

wherein the anellosome composition comprises a plurality of anellosomes that comprise:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence);

wherein the effector comprises a polypeptide that, when mutated, causes a human disease, or a functional variant of said polypeptide;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

thereby modulating, e.g., inhibiting or enhancing, the biological function in the subject.

26. The anellosome or method of any of the preceding embodiments, wherein the effector comprises SMN, dystrophin, UGT1A1, OCA1, C1, or a functional variant of any of the foregoing.

27. The anellosome or method of any of the preceding embodiments, wherein the effector is a secreted polypeptide (e.g., the effector comprises a signal sequence, e.g., a signal sequence that is endogenous to the effector, or a heterologous signal sequence), or the effector is an intracellular polypeptide.

28. A method of treating a disease or disorder in a subject, the method comprising administering an effective amount of an anellosome composition or an isolated nucleic acid molecule (e.g., an expression vector) to the subject,

wherein the anellosome composition or isolated nucleic acid molecule comprises a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector)) (e.g., each as described herein), and wherein

(i) the disease or disorder is or comprises a neuromuscular disease, e.g., spinal muscular atrophy and the effector comprises a SMN or a functional variant thereof;

(ii) the disease or disorder is a muscular dystrophy (e.g., Duchenne muscular dystrophy or Becker muscular dystrophy), and the effector comprises dystrophin or a functional variant thereof, e.g., a micro-dystrophin;

(iii) the disease or disorder is an enzymatic insufficiency, e.g., Crigler–Najjar syndrome, and the effector comprises a uridine diphosphate glucuronyl-transferase, e.g., UGT1A1 or a functional variant thereof;

(iv) the disease or disorder is albinism, e.g., ocular albinism, and the effector comprises a tyrosinase, e.g., OCA1 or a functional variant thereof;

(v) the disease or disorder is C1 complement deficiency, and the effector comprises C1 or a functional variant thereof; or

(vi) the disease or disorder is a disease or disorder of Table A, and the effector comprises an enzyme of Table A, or a functional variant thereof;

(vii) the disease or disorder is a disease or disorder of Table B, and the effector comprises a polypeptide of Table B, or a functional variant thereof;

thereby treating the disease or disorder in the subject.

29. A method of delivering an effector to a subject having a disease or disorder, the method comprising administering an effective amount of an anellosome composition or an isolated nucleic acid molecule (e.g., an expression vector) to the subject,

wherein the anellosome composition or isolated nucleic acid molecule comprises a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector); and

wherein:

(i) the disease or disorder comprises a neuromuscular disease, e.g., spinal muscular atrophy and the effector comprises SMN or a functional variant thereof;

(ii) the disease or disorder is a muscular dystrophy (e.g., Duchenne muscular dystrophy or Becker muscular dystrophy), and the effector comprises dystrophin or a functional variant thereof, e.g., a micro-dystrophin;

(iii) the disease or disorder is an enzymatic insufficiency, e.g., Crigler–Najjar syndrome, and the effector comprises a uridine diphosphate glucuronyl-transferase, e.g., UGT1A1 or a functional variant thereof;

(iv) the disease or disorder is albinism, e.g., ocular albinism, and the effector comprises a tyrosinase, e.g., OCA1 or a functional variant thereof;

(v) the disease or disorder is a C1 complement deficiency, and the effector comprises C1 or a functional variant thereof; or

(vi) the disease or disorder is a disease or disorder of Table A, and the effector comprises an enzyme of Table A, or a functional variant thereof;

(vii) the disease or disorder is a disease or disorder of Table B, and the effector comprises a polypeptide of Table B, or a functional variant thereof;

thereby delivering the effector to the subject.

30. A method of manufacturing an anellosome composition, the method comprising:

a) providing a host cell comprising one or more nucleic acid molecules encoding components of an anellosome of any of the preceding embodiments;

b) maintaining (e.g., culturing) the host cell under conditions that allow the cell to produce one or more anellosomes, thereby making anellosomes; and

c) formulating the anellosomes, e.g., as a pharmaceutical composition suitable for administration to a subject.

31. A method of manufacturing an anellosome composition, the method comprising:

a) providing a plurality of anellosomes 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); and

c) formulating the plurality of anellosomes, 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.

32. The method of embodiment 31, wherein the anellosome 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} anellosomes.

33. The method of embodiment 31 or 32, wherein the anellosome composition comprises at least 10 ml, 20 ml, 50 ml, 100 ml, 200 ml, 500 ml, 1 L, 2 L, 5 L, 10 L, 20 L, or 50 L.

34. The anellosome or method of any of the preceding embodiments, wherein the effector comprises a polypeptide that, when mutated, causes a human disease, (e.g., SMN, dystrophin, UGT1A1, OCA1, or C1) or a functional variant of said polypeptide.

35. The anellosome or method of any of the preceding embodiments, wherein the effector comprises an enzyme that covalently modifies a small molecule substrate that is endogenous to a human cell (e.g., a uridine diphosphate glucuronyl-transferase, e.g., UGT1A1; or a tyrosinase, e.g., OCA1).

36. The anellosome or method of any of the preceding embodiments, wherein the effector comprises a polypeptide that is endogenous to neurons (e.g., SMN) or muscle cells (e.g., dystrophin), or a functional variant of said polypeptide.

37. The anellosome or method of any of the preceding embodiments, wherein the effector comprises a complement protein (e.g., C1).

38. The anellosome or method of any of the preceding embodiments, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

39. The anellosome or method of any of the preceding embodiments, wherein the genetic element comprises a region comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

(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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (SEQ ID NO: 168);

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

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

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

(x) GCGCTACGCGCGCGCGCCGGGGGGCTCTGCCCCCCC (SEQ ID NO: 172);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto.

40. The anellosome or method of any of the preceding embodiments, wherein the genetic element comprises a sequence comprising at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%.

41. The anellosome or method of embodiment 40, wherein the genetic element comprises at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 80%.

42. The anellosome or method of embodiment 40, wherein the genetic element comprises at least 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%.

43. The anellosome or method of embodiment 40, wherein the genetic element comprises at least 36 consecutive nucleotides having a GC content of at least 80%.

1000. A polypeptide, e.g., an ORF1 molecule, comprising one or more of:

(a) a first region comprising an amino acid sequence having 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 (e.g., MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVR (SEQ ID NO: 216) or MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRRWRRGRRKTRTYRRRR RFRRRGRK (SEQ ID NO: 186), or as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10) 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),

(b) a second region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to a jelly-roll region sequence described herein (e.g., PTYTTIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLD ALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLM MMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACNLQNPFVKP DKLSNNVTLWSLNT (SEQ ID NO: 217), or as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10) or a sequence comprising at least 6 (e.g., at least 6, 7, 8, 9, 10, 11, or 12) beta strands;

(c) a third region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an N22 domain sequence described herein (e.g., TMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFADFQ KNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQL EAQNKLLQTGPFTPNIQGQLSDNISMFYKFYFK (SEQ ID NO: 219), or as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10); and

(d) a fourth region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an Anellovirus ORF1 C-terminal domain (CTD) sequence described herein (e.g., WGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRISQDWA LKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 220), or as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10);

wherein the 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), e.g., an insertion, substitution, chemical or enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of an arginine-rich region, jelly-roll domain, HVR, N22, or CTD, e.g., as described herein).

1000A. The polypeptide of embodiment 1000, wherein the amino acid sequences of the region of (a), (b), (c), and (d) have at least 90% sequence identity to their respective references.

1001. The polypeptide of embodiment 1000, wherein the polypeptide comprises:

(i) the first region and the second region;

(ii) the first region and the third region;

(iii) the first region and the fourth region;

(iv) the second region and the third region;

(v) the second region and the fourth region;

(vi) the third region and the fourth region;

(vii) the first region, the second region, and the third region;

(viii) the first region, the second region, and the fourth region;

(ix) the first region, the third region, and the fourth region; or

(x) the second region, the third region, and the fourth region.

1002. A polypeptide, e.g., an ORF1 molecule, comprising:

(a) a first region comprising an amino acid sequence having 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 (e.g., MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVR (SEQ ID NO: 216) or MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRWRRGRRKTRTYRRRR RFRRRGRK (SEQ ID NO: 186), or as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6,

8, 10, 12, 14, 16, 18, 20-37, or D1-D10) 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),

(b) a second region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to a jelly-roll region sequence described herein (e.g., PTYTTIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLD ALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLM MMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACNLQNPFVKP DKLSNNVTLWSLNT (SEQ ID NO: 217), or as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10) or a sequence comprising at least 6 beta strands;

(c) a third region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an N22 domain sequence described herein (e.g., TMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFADFQ KNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQL EAQNKLLQTGPFTPNIQGQLSDNISMFYKFYFK (SEQ ID NO: 219), or as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10); and

(d) a fourth region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an Anellovirus ORF1 C-terminal domain (CTD) sequence described herein (e.g., WGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRISQDWA LKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 220), or as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10);

wherein the 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), e.g., an insertion, substitution, chemical or enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of an arginine-rich region, jelly-roll domain, HVR, N22, or CTD, e.g., as described herein).

1002A. The polypeptide according to embodiment 1002, wherein the amino acid sequences of the (a), (b), (c), and (d) region have at least 90% sequence identity to their respective references.

1003. The polypeptide of any of the preceding embodiments, wherein:

the first region comprises an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to amino acids 1-38 of the ORF1 sequence listed in Table 16;

the second region comprises an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to amino acids 39-246 of the ORF1 sequence listed in Table 16;

the third region comprises an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to amino acids 375-537 of the ORF1 sequence listed in Table 16; and/or

the fourth region comprises an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to amino acids 538-666 of the ORF1 sequence listed in Table 16.

1003A. The polypeptide according to embodiment 1003, wherein the amino acid sequences of the first, second, third and fourth region have at least 90% sequence identity to their respective references.

1004. The polypeptide of any of the preceding embodiments, wherein:

the first region comprises an amino acid sequence having 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 as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10;

the second region comprises an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to a jelly-roll region sequence as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10;

the third region comprises an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an N22 domain sequence as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10; and/or

the fourth region comprises an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to a CTD sequence as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1004A. The polypeptide according to embodiment 1004, wherein the amino acid sequences of the first, second, third and fourth region have at least 90% sequence identity to their respective references.

1005. The polypeptide of any of the preceding embodiments, wherein the polypeptide comprises, in N-terminal to C-terminal order, the first region, the second region, the third region, and the fourth region.

1006. The polypeptide of any of the preceding embodiments, wherein the at least one difference comprises at least one difference in the first region relative to the arginine-rich region of a wild-type ORF1 protein.

1007. The polypeptide of any of the preceding embodiments, wherein the first region comprises an arginine-rich region from the ORF1 protein of an Anellovirus other than the wild-type Anellovirus to which the polypeptide, or the portion thereof excluding the first region, has greatest sequence identity.

1008. The polypeptide of any of the preceding embodiments, wherein the first region comprises an amino acid sequence having at least 70% sequence identity to the arginine-rich region from an Anellovirus other than the wild-type Anellovirus to which the polypeptide has greatest sequence identity.

1009. The polypeptide of any of the preceding embodiments, wherein the first region comprises a polypeptide that has less than 15% (e.g., less than 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1%) sequence identity to an wild-type Anellovirus genome (e.g., as described herein), or a portion thereof having the same amino acid length as the first region.

1010. The polypeptide of any of the preceding embodiments, wherein the first region has DNA binding activity and/or nuclear localization activity.

1011. The polypeptide of any of the preceding embodiments, wherein the first region comprises a DNA-binding region and/or a nuclear localization sequence.

1012. The polypeptide of any of the preceding embodiments, wherein the at least one difference comprises at least one difference in the second region relative to the jelly-roll region of a wild-type ORF1 protein.

1013. The polypeptide of any of the preceding embodiments, wherein the second region comprises a jelly-roll region from the ORF1 protein of an Anellovirus other than the wild-type Anellovirus to which the polypeptide, or the portion thereof excluding the second region, has greatest sequence identity.

1014. The polypeptide of any of the preceding embodiments, wherein the second region comprises an amino acid sequence having at least 70% sequence identity to the jelly-roll region from an Anellovirus other than the wild-type Anellovirus to which the polypeptide has greatest sequence identity.

1015. The polypeptide of any of the preceding embodiments, wherein the second region comprises a polypeptide that has less than 15% (e.g., less than 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1%) sequence identity to an wild-type Anellovirus genome (e.g., as described herein), or a portion thereof having the same amino acid length as the second region.

1016. The polypeptide of any of the preceding embodiments, wherein the at least one difference comprises at least one difference in the third region relative to the N22 domain of a wild-type ORF1 protein.

1017. The polypeptide of any of the preceding embodiments, wherein the third region comprises an N22 domain from the ORF1 protein of an Anellovirus other than the wild-type Anellovirus to which the polypeptide, or the portion thereof excluding the third region, has greatest sequence identity.

1018. The polypeptide of any of the preceding embodiments, wherein the third region comprises an amino acid sequence having at least 70% sequence identity to the N22 region from an Anellovirus other than the wild-type Anellovirus to which the polypeptide has greatest sequence identity.

1019. The polypeptide of any of the preceding embodiments, wherein the third region comprises a polypeptide that has less than 15% (e.g., less than 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1%) sequence identity to an wild-type Anellovirus genome (e.g., as described herein), or a portion thereof having the same amino acid length as the third region.

1020. The polypeptide of any of the preceding embodiments, wherein the at least one difference comprises at least one difference in the fourth region relative to the CTD domain of a wild-type ORF1 protein.

1021. The polypeptide of any of the preceding embodiments, wherein the fourth region comprises a CTD domain from the ORF1 protein of an Anellovirus other than the wild-type Anellovirus to which the polypeptide, or the portion thereof excluding the fourth region, has greatest sequence identity.

1022. The polypeptide of any of the preceding embodiments, wherein the fourth region comprises an amino acid sequence having at least 70% sequence identity to the CTD region from an Anellovirus other than the wild-type Anellovirus to which the polypeptide has greatest sequence identity.

1023. The polypeptide of any of the preceding embodiments, wherein the fourth region comprises a polypeptide that has less than 15% (e.g., less than 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1%) sequence identity to an wild-type Anellovirus genome (e.g., as described herein), or a portion thereof having the same amino acid length as the fourth region.

1024. The polypeptide of any of the preceding embodiments, further comprising an amino acid sequence, e.g., a hypervariable region (HVR) sequence (e.g., the HVR sequence of an Anellovirus ORF1 molecule, e.g., as described herein), wherein the amino acid sequence 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).

1025. The polypeptide of embodiment 1024, wherein the HVR sequence is positioned between the second region and the third region.

1026. The polypeptide of embodiment 1024 or 1025, wherein the HVR sequence comprises an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to the HVR from an Anellovirus other than the wild-type Anellovirus to which the ORF1 protein has greatest sequence identity.

1027. The polypeptide of any of embodiments 1024-1026, wherein the HVR sequence is heterologous relative to one or more of the first region, second region, third region, and/or fourth region.

1028. The polypeptide of any of embodiments 1024-1027, wherein the at least one difference comprises at least one difference in the HVR sequence relative to the sequence of an HVR of a wild-type ORF1 protein (e.g., from a wild-type Anellovirus genome, e.g., as described herein).

1029. The polypeptide of any of embodiments 1024-1028, wherein the HVR sequence comprises an HVR from the ORF1 protein of an Anellovirus other than the wild-type Anellovirus to which the polypeptide, or the portion thereof excluding the HVR sequence, has greatest sequence identity.

1030. The polypeptide of any of embodiments 1024-1029, wherein the HVR sequence comprises an amino acid sequence having at least 70% sequence identity to the HVR from an Anellovirus other than the wild-type Anellovirus to which the polypeptide has greatest sequence identity.

1031. The polypeptide of any of embodiments 1024-1030, wherein the HVR comprises an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to HVR sequence as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1032. The polypeptide of any of embodiments 1024-1031, wherein the HVR sequence comprises at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to amino acids 247-374 of the ORF1 sequence listed in Table 16.

1033. The polypeptide of any of the preceding embodiments, further comprising a heterologous polypeptide, e.g., a polypeptide that is heterologous relative to one or more of the first region, second region, third region, and/or fourth region, and/or is exogenous relative to an anellosome comprising the polypeptide.

1034. The polypeptide of embodiment 1033, wherein the polypeptide lacks an Anellovirus HVR sequence.

1035. The polypeptide of embodiment 1033, wherein the heterologous polypeptide is present on the exterior of the anellosome.

1036. The polypeptide of embodiment 1033, wherein the heterologous polypeptide is present on the interior of the anellosome.

1037. The polypeptide of any of embodiments 1033-1036, wherein the heterologous polypeptide has a functionality that is exogenous to the anellosome or a wild-type Anellovirus.

1038. The polypeptide of any of embodiments 1033-1037, wherein the heterologous polypeptide consists of about 140 or fewer amino acids (e.g., 100, 110, 120, 125, 130, 135, 136, 137, 138, 139, 140, 145, 150, 155, or 160 or fewer amino acids).

1039. The polypeptide of any of embodiments 1033-1038, wherein the size of the heterologous polypeptide is between 50-150% relative to a wild-type HVR region of an Anellovirus, e.g., as described herein.

1039A. The polypeptide of any of embodiments 1033-1039, wherein the heterologous polypeptide is positioned between the second region and the third region.

1040. The polypeptide of any of the preceding embodiments, further comprising one or more amino acids between the first region and the second region, one or more amino acids between the second region and the third region, and/or one or more amino acids between the third region and the fourth region.

1041. The polypeptide of any of the preceding embodiments, further comprising one or more amino acids positioned N-terminal relative to the first region.

1042. The polypeptide of any of the preceding embodiments, further comprising one or more amino acids positioned C-terminal relative to the fourth region.

1043. The polypeptide of any of the preceding embodiments, comprising a plurality of subsequences of at least four (e.g., 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30) contiguous amino acids having 100% sequence identity to the corresponding subsequences of a wild-type Anellovirus ORF1 amino acid sequence, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1044. The polypeptide of any of the preceding embodiments, comprising a plurality of subsequences of at least ten (e.g., 10, 15, 20, 25, 30, 40, or 50) contiguous amino acids having at least 80% sequence identity to the corresponding subsequences of a wild-type Anellovirus ORF1 amino acid sequence, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1045. The polypeptide of any of the preceding embodiments, comprising a plurality of subsequences of at least twenty (e.g., 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100) contiguous amino acids having at least 60% sequence identity to the corresponding subsequences of a wild-type Anellovirus ORF1 amino acid sequence, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1046. The polypeptide of any of embodiments 1043-1045, wherein the plurality of subsequences are positioned within the first region, second region, third region, and/or fourth region.

1047. The polypeptide of any of the preceding embodiments, wherein the first region comprises at least about 40 amino acids (e.g., at least about 50, 60, 70, 80, 90, or 100 amino acids, e.g., about 40-100, 40-90, 40-80, 40-70, 50-100, 50-70, 60-100, 60-90, 60-80, or 60-70 amino acids).

1048. The polypeptide of any of the preceding embodiments, wherein the first region comprises at least about 70% (e.g., at least about 70%, 75%, 80%, 85%, 90%, 95%, or 100%) basic residues (e.g., arginine, lysine, or a combination thereof).

1049. The polypeptide of any of the preceding embodiments, wherein the first region comprises at least about 70% (e.g., at least about 70%, 75%, 80%, 85%, 90%, 95%, or 100%) arginine residues.

1050. The polypeptide of any of the preceding embodiments, wherein the polypeptide forms homomultimers with additional copies of the polypeptide.

1051. The polypeptide of embodiment 1050, wherein the first region binds to corresponding first regions on additional copies of the polypeptide.

1052. The polypeptide of embodiment 1050, wherein the homomultimers form a capsid, e.g., encapsulating a nucleic acid, e.g., a genetic element or an Anellovirus genome or a portion thereof.

1053. The polypeptide of any of the preceding embodiments, wherein the polypeptide is a capsid protein or can form a portion of a capsid.

1054. The polypeptide of any of the preceding embodiments, wherein the polypeptide has replicase activity.

1055. The polypeptide of any of the preceding embodiments, wherein the polypeptide binds to a nucleic acid (e.g., DNA).

1056. A complex comprising:

(a) the polypeptide of any of the preceding embodiments, and

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence.

1057. A complex comprising:

(a) an ORF1 molecule, and

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence;

wherein the ORF1 molecule is bound to (e.g., non-covalently bound to) the genetic element,

wherein the ORF1 molecule, the genetic element, or both of the ORF1 molecule and the genetic element comprise at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type ORF1 protein, wild-type Anellovirus genome, or both of the wild-type ORF1 protein and wild-type Anellovirus genome, respectively (e.g., as described herein), e.g., an insertion, substitution, chemical or enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of an arginine-rich region, jelly-roll domain, HVR, N22, or CTD, e.g., as described herein) or genomic region (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region, e.g., as described herein).

1058. The complex of embodiment 1056 or 1057, wherein the complex is in vitro, e.g., wherein the complex is in a substantially cell-free composition.

1059. The complex of any of embodiments 1056-1058, wherein the complex is in a cell, e.g., a host cell, e.g., a helper cell, e.g., in the nucleus of the cell.

1060. The complex of any of embodiments 1056-1059, wherein the ORF1 molecule is part of a proteinaceous exterior.

1061. The complex of any of embodiments 1056-1060, wherein the genetic element is undergoing replication.

1062. The complex of any of embodiments 1056-1061, wherein the complex is in an anellosome.

1063. The complex of any of embodiments 1056-1062, wherein the genetic element further comprises a nucleic acid sequence encoding the polypeptide.

1064. The complex of any of embodiments 1056-1063, wherein the genetic element does not comprise a nucleic acid sequence encoding the polypeptide.

1065. The complex of any of embodiments 1056-1064, wherein the genetic element comprises a GC-rich region, e.g., as described herein.

1066. The complex of embodiment 1065, wherein the GC-rich region comprises at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence of any of:

(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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (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);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto.

1067. An anellosome comprising:

(a) a proteinaceous exterior;

(b) the polypeptide or complex of any of the preceding embodiments;

(c) a genetic element comprising a promoter element operably linked to a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector, e.g., as described herein); and

wherein the genetic element is enclosed within the proteinaceous exterior.

1068. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising:

(i) a promoter element operably linked to a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector, e.g., as described herein), and

(ii) a nucleic acid encoding the polypeptide of any of the preceding embodiments; and

wherein the genetic element is enclosed within the proteinaceous exterior.

1069. An anellosome comprising:

(a) a proteinaceous exterior;

(b) an ORF1 molecule or a nucleic acid encoding the ORF1 molecule;

(c) a genetic element comprising a promoter element operably linked to a heterologous nucleic acid sequence (e.g., a DNA sequence) encoding an effector; and

wherein the genetic element is enclosed within the proteinaceous exterior.

1070. An anellosome comprising:

(a) a proteinaceous exterior;

(b) an ORF1 molecule or a nucleic acid encoding the ORF1 molecule;

(c) a genetic element comprising a promoter element, a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a region comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

(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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (SEQ ID NO: 168);

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

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

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

(x) GCGCTACGCGCGCGCGCCGGGGGGCTCTGCCCCCCC (SEQ ID NO: 172);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto; and

wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell; and optionally, wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1071. An anellosome comprising:

(a) a proteinaceous exterior;

(b) an ORF1 molecule or a nucleic acid encoding the ORF1 molecule;

(c)a genetic element comprising a promoter element, a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a sequence comprising at least 20 (e.g., at least 20, 25, 30, 31, 32, 33, 34, 35, or 36) consecutive nucleotides having a GC content of at least 70% (e.g., at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%);

wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1072. An anellosome comprising:

(a) a proteinaceous exterior;

(b) an ORF1 molecule or a nucleic acid encoding the ORF1 molecule;

wherein:

(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 ORF1 molecule are part of a β-strands;

(ii) the secondary structure of the ORF1 molecule comprises at least three (e.g., at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20) β- strands;

(iii) the secondary structure of the ORF1 molecule comprises a ratio of β-strands to α-helices of at least 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1; and

(c) a genetic element comprising a promoter element, a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence;

wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1073. An anellosome comprising:

(a) a proteinaceous exterior;

(b) an ORF1 molecule or a nucleic acid encoding the ORF1 molecule;

(c) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence;

wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1074. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element, a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a region comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

(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);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto; and

wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell; and

optionally, wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1075. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element, a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a sequence comprising at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%; and

wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell; and optionally, wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1076. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector),

wherein the genetic element comprises a region (e.g., a packaging region, e.g., positioned 3' relative to the nucleic acid sequence encoding the effector) having:

at least 95% (e.g., at least 95, 96, 97, 98, 99, or 100%) sequence identity to the nucleic acid sequence: CGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGC (SEQ ID NO: 160);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell.

1076A. An anellosome comprising:

(i) a genetic element comprising a promoter element and a nucleic acid sequence encoding a therapeutic exogenous effector, wherein the genetic element comprises a sequence having at least 95% sequence identity to the 5' UTR nucleotide sequence from an Anellovirus described herein (e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17); and/or

(ii) a proteinaceous exterior comprising a polypeptide having at least 95% sequence identity to a polypeptide encoded by the ORF1 gene of an Anellovirus described herein (e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17);

wherein the genetic element is enclosed within the proteinaceous exterior, and

optionally wherein the anellosome is capable of delivering the genetic element into a mammalian cell.

1076B. An anellosome comprising:

(I) a genetic element comprising: (a) a promoter element, and (b) a nucleic acid sequence encoding an exogenous effector (e.g., an exogenous effector as described herein), wherein the nucleic acid sequence is operably linked to the promoter element; and (c) a 5' UTR domain comprising one of:

(c)(i) a nucleic acid sequence of nucleotides 323 – 393 of SEQ ID NO: 54, or a nucleic acid sequence at least 85% identical thereto;

(c)(ii) a nucleic acid sequence of any of SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115 , SEQ ID NO: 116, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119 or a nucleic acid sequence at least 85% identical thereto; or

(c)(iii) a nucleic acid sequence of nucleotides 117 – 187 of SEQ ID NO: 61, or a nucleic acid sequence at least 85% identical thereto;

(II) a proteinaceous exterior comprising an ORF1 molecule;

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the synthetic anellosome is capable of delivering the genetic element into a mammalian, e.g., a human, cell.

1077. The anellosome of any of the preceding embodiments, wherein the proteinaceous exterior comprises the ORF1 molecule.

1078. The anellosome of any of the preceding embodiments, wherein at least 60% (e.g., at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) of protein in the proteinaceous exterior comprises an ORF1 molecule.

1079. The anellosome of any of the preceding embodiments, wherein no more than 1% (e.g., no more than 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, or 40%) of protein in the proteinaceous exterior comprises an ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 molecule.

1080. The anellosome of any of the preceding embodiments, wherein the ORF1 molecule comprises an amino acid sequence having at least 70% (e.g., at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to an ORF1 protein listed in, or encoded by a sequence listed in any of Tables A1-A12, B1-B5, C1-C5, 1-18, 20-37, or D1-D10.

1081. The anellosome of any of the preceding embodiments, wherein the ORF1 molecule comprises a polypeptide of any of the preceding embodiments.

1082. The anellosome of any of the preceding embodiments, wherein the genetic element further comprises a nucleic acid sequence encoding the ORF1 molecule.

1083. The anellosome of any of the preceding embodiments, wherein the genetic element does not comprise a nucleic acid sequence encoding the ORF1 molecule.

1084. The anellosome of any of the preceding embodiments, wherein the genetic element comprises at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 80%.

1085. The anellosome of any of the preceding embodiments, wherein the genetic element comprises at least 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%.

1086. The anellosome of any of the preceding embodiments, wherein the genetic element comprises at least 36 consecutive nucleotides having a GC content of at least 80%.

1087. An isolated nucleic acid composition (e.g., comprising one, two, or more nucleic acid molecules) comprising a nucleic acid encoding the polypeptide of any of the preceding embodiments;

optionally wherein the isolated nucleic acid composition further comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region); and

optionally wherein the nucleic acid molecule does not comprise:

(i) a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1088. An isolated nucleic acid composition (e.g., comprising one, two, or more nucleic acid molecules), wherein the isolated nucleic acid composition comprises a genetic element encoding an ORF1 molecule;

wherein:

(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 ORF1 molecule are part of a β-sheet;

(ii) the secondary structure of the ORF1 molecule comprises at least three (e.g., at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20) β-sheets;

(iii) the secondary structure of the ORF1 molecule comprises a ratio of β-sheets to α-helices of at least 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1; and

wherein the genetic element comprises a promoter element, a nucleic acid sequence encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence;

wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region); and

optionally wherein the nucleic acid molecule does not comprise:

(i) a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1089. An isolated nucleic acid composition (e.g., comprising one, two, or more nucleic acid molecules) comprising:

(a) a genetic element encoding an ORF1 molecule;

(b) at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

(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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (SEQ ID NO: 168);

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

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

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

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

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto; and

(c) at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

optionally wherein the nucleic acid molecule does not comprise:

(i) a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1090. An isolated nucleic acid composition (e.g., comprising one, two, or more nucleic acid molecules), wherein the isolated nucleic acid composition comprises:

(a) a genetic element encoding an ORF1 molecule;

(b) at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%; and

wherein the isolated nucleic acid composition comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region); and

optionally wherein the nucleic acid molecule does not comprise:

(i) a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1090A. An isolated nucleic acid composition (e.g., comprising one, two, or more nucleic acid molecules), wherein the isolated nucleic acid composition comprises a genetic element comprising a 5' UTR nucleotide sequence from an Anellovirus described herein (e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17).

1091. The isolated nucleic acid composition of any of embodiments 1089-1090, wherein (a) and (b) are part of the same nucleic acid.

1092. The isolated nucleic acid composition of any of embodiments 1089-1091, wherein (a) and (b) are part of different nucleic acids.

1093. The isolated nucleic acid composition of any of the preceding embodiments, wherein the genetic element further comprises one or more of: a TATA box, an initiator element, a cap site, a transcriptional start site, a 5' UTR conserved domain, an ORF1-encoding sequence, an ORF1/1-encoding sequence, an ORF1/2-encoding sequence, an ORF2-encoding sequence, an ORF2/2-encoding sequence, an ORF2/3-encoding sequence, an ORF2/3t-encoding sequence, a three open-reading frame region, a poly(A) signal, and/or a GC-rich region from an Anellovirus described herein (e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

1094. The isolated nucleic acid composition of any of the preceding embodiments, wherein the genetic element further comprises an Anellovirus genome sequence (e.g., as described herein, e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

1095. The isolated nucleic acid composition of embodiment 1094, further comprising at least one additional copy of the Anellovirus genome sequence or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto (e.g., a total of 1, 2, 3, 4, 5, or 6 copies).

1096. The isolated nucleic acid composition of any of the preceding embodiments, further comprising at least one additional copy of the genetic element (e.g., a total of 1, 2, 3, 4, 5, or 6 copies).

1097. An isolated nucleic acid composition (e.g., comprising one, two, or more nucleic acid molecules) comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

(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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (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);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto; and

at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

optionally wherein the nucleic acid molecule does not comprise:

(i) a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1098. An isolated nucleic acid composition (e.g., comprising one, two, or more nucleic acid molecules), wherein the isolated nucleic acid composition comprises at least 20, 25, 30, 31, 32, 33, 34, 35, or 36

consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%; and

wherein the isolated nucleic acid composition comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type Anellovirus genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region); and

optionally wherein the nucleic acid molecule does not comprise:

(i) a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1099. The isolated nucleic acid composition of any of the preceding embodiments, wherein the ORF1 molecule comprises a polypeptide of any of the preceding embodiments.

1100. The isolated nucleic acid composition of any of the preceding embodiments, comprising at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 80%.

1101. The isolated nucleic acid composition of any of the preceding embodiments, comprising at least 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%.

1102. The isolated nucleic acid composition of any of the preceding embodiments, comprising at least 36 consecutive nucleotides having a GC content of at least 80%.

1103. The isolated nucleic acid composition of any of the preceding embodiments, further comprising one or more of a promoter element, a nucleic acid sequence encoding an effector (e.g., an exogenous effector or an endogenous effector), and/or a protein binding sequence (e.g., an exterior protein binding sequence).

1104. The isolated nucleic acid composition of any of the preceding embodiments, comprising at least about 100, 150, 200, 250, 300, 350, 400, 450, or 500 consecutive nucleotides of a wild-type *Anellovirus* genome sequence, or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto.

1105. An isolated nucleic acid molecule (e.g., an expression vector) comprising a nucleic acid sequence having at least 95% (e.g., at least 95, 96, 97, 98, 99, or 100%) sequence identity to the nucleic acid sequence:

(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).

1106. The isolated nucleic acid composition of any of the preceding embodiments, wherein the isolated nucleic acid molecule is circular.

1107. An isolated cell, e.g., a host cell, comprising:

(a) a nucleic acid encoding a polypeptide of any of the preceding embodiments, wherein the nucleic acid is a plasmid, is a viral nucleic acid, or is integrated into a cell chromosome, and

(b) a genetic element, wherein the genetic element comprises a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence, wherein optionally the genetic element does not encode an ORF1 polypeptide (e.g., an ORF1 protein).

1108. An isolated cell, e.g., a host cell, comprising:

(a) a nucleic acid encoding an ORF1 molecule, wherein the nucleic acid is a plasmid, is a viral nucleic acid, or is integrated into a cell chromosome, and

(b) a genetic element, wherein the genetic element comprises a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence.

1109. An isolated cell, e.g., a host cell, comprising:

(a) a nucleic acid encoding an ORF1 molecule (e.g., wherein the nucleic acid is a plasmid, is a viral nucleic acid, or is integrated into a cell chromosome), and

(b) a genetic element that does not encode an ORF1 molecule, wherein the genetic element comprises a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence.

1109A. An isolated cell, e.g., a host cell, comprising:

(i) a nucleic acid molecule (e.g., a first nucleic acid molecule) comprising the nucleic acid sequence of a genetic element of an anellosome as described herein (e.g., a genetic element that does not encode an ORF1 molecule), and

(ii) optionally, a nucleic acid molecule, e.g., 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 listed in any of Table 16, 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.

1110. The isolated cell of any of the preceding embodiments, wherein the genetic element that does not encode an ORF1 molecule encodes a fragment of an ORF1 molecule, e.g., a fragment that does not form a capsid, e.g., a fragment of less than 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100, 50, 20, or 10 nucleotides.

1111. An isolated cell, e.g., a host cell, comprising a nucleic acid encoding an ORF1 molecule (e.g., wherein the nucleic acid is a plasmid, is a viral nucleic acid, or is integrated into a cell chromosome), wherein the isolated cell does not comprise one or more of an ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 molecule.

1112. An isolated cell, e.g., a host cell, comprising the nucleic acid composition of any of the preceding embodiments.

1113. A helper nucleic acid (e.g., a plasmid or viral nucleic acid) encoding an ORF1 molecule, wherein the isolated cell does not comprise one or more of an ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 molecule.

1114. A composition comprising:

(a) an isolated cell described herein, and

(b) an anellosome described herein.

1115. A composition comprising:

(a) a cell comprising a nucleic acid encoding an ORF1 molecule (e.g., wherein the nucleic acid is a plasmid, is a viral nucleic acid, or is integrated into a cell chromosome), and

(b) a genetic element (e.g., inside the cell or outside the cell, e.g., in cell culture medium) that does not encode an ORF1 molecule, wherein the genetic element comprises a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence.

1116. A pharmaceutical composition comprising the polypeptide, complex, anellosome or isolated nucleic acid of any of the preceding embodiments and a pharmaceutically acceptable carrier and/or excipient.

1117. A method of manufacturing an ORF1 molecule, the method comprising:

(a) providing a host cell (e.g., a host cell described herein) comprising a nucleic acid encoding the polypeptide of any of the preceding embodiments, and

(b) maintaining the host cell under conditions that allow the cell to produce the polypeptide; thereby manufacturing the ORF1 molecule.

1118. A method of manufacturing an ORF1 molecule, the method comprising:

(a) providing a host cell (e.g., a host cell described herein) comprising the nucleic acid composition of any of the preceding embodiments, and

(b) maintaining the host cell under conditions that allow the cell to produce the polypeptide; thereby manufacturing the ORF1 molecule.

1119. The method of embodiment 1117 or 1118, wherein the host cell is a helper cell.

1120. The method of embodiment 1119, wherein the helper cell comprises one or more additional nucleic acids encoding one or more additional ORFs (e.g., one or more of ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3) of a wild-type Anellovirus, e.g., as described herein.

1121. The method of any of embodiments 1117-1120, wherein the nucleic acid is integrated into the genome of the host cell.

1122. The method of any of embodiments 1117-1121, wherein the host cell produces at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, 10,000, 50,000, 100,000, 500,000, or 1,000,000 copies (e.g., at least about 60 copies) of the polypeptide per host cell.

1123. The method of any of embodiments 1117-1122, wherein the host cell produces at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, 10,000, or 100,000 copies (e.g., at least about 60 copies) of the polypeptide per anellosome produced by the host cell.

1124. The method of any of embodiments 1117-1123, wherein the method comprises providing a plurality of host cells, and maintaining the host cells under conditions that allow the production of at least 1000 copies of the polypeptide per cell.

1125. The method of embodiment 1124, wherein the plurality of host cells produces at least about $1x10^5$, $1x10^6$, $1x10^7$, $1x10^8$, $9x10^8$, $1x10^9$, $1x10^{10}$, $1x10^{11}$, or $1x10^{12}$ copies of the polypeptide.

1126. A method of manufacturing an anellosome composition, the method comprising:

(a) providing a helper cell, e.g., a helper cell described herein;

(b) introducing a genetic element into the helper cell under conditions that allow the cell to produce anellosomes, and

(c) formulating the anellosomes, e.g., as a pharmaceutical composition suitable for administration to a subject,

thereby making the anellosome composition.

1127. A method of manufacturing an anellosome composition, the method comprising:

(a) providing a host cell;

(b) introducing a helper nucleic acid into the host cell;

(c) introducing a genetic element into the host cell (e.g., before, after, or simultaneously with (b)), under conditions that allow the cell to produce anellosomes; and

(d) formulating the anellosomes, e.g., as a pharmaceutical composition suitable for administration to a subject;

thereby making the anellosome composition.

1128. A method of manufacturing an anellosome composition, the method comprising:

(a) providing a helper cell comprising a nucleic acid encoding an ORF1 molecule (e.g., wherein the nucleic acid is a plasmid, is a viral nucleic acid, or is integrated into a helper cell chromosome);

(b) introducing a genetic element into the helper cell under conditions that allow the cell to produce anellosomes, wherein the genetic element does not encode an ORF1 molecule, wherein the genetic element comprises a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence; and

(c) formulating the anellosomes, e.g., as a pharmaceutical composition suitable for administration to a subject;

thereby making the anellosome composition.

1129. A method of manufacturing an anellosome composition, the method comprising:

(a) providing a host cell;

(b) introducing a helper nucleic acid encoding an ORF1 molecule (e.g., wherein the nucleic acid is a plasmid, or a viral nucleic acid), into the host cell; and

(c) introducing a genetic element into the host cell (e.g., before, after, or simultaneously with (b)), under conditions that allow the cell to produce an anellosome, wherein the genetic element does not encode an ORF1 molecule, wherein the genetic element comprises a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence,

thereby making the anellosome.

1130. The method of any of the preceding embodiments, which further comprises separating the anellosome from the helper cell or host cell.

1131. The method of any of the preceding embodiments, wherein providing a helper cell comprises introducing a helper nucleic acid into the host cell, e.g., wherein the helper nucleic acid encodes an ORF1 molecule (e.g., wherein the nucleic acid is a plasmid, or a viral nucleic acid).

1132. The method of any of the preceding embodiments, wherein the helper cell comprises the ORF1 molecule.

1133. The method of any of the preceding embodiments, wherein the nucleic acid comprises one or more of: a TATA box, an initiator element, a cap site, a transcriptional start site, a 5' UTR conserved domain, an ORF1-encoding sequence, an ORF1/1-encoding sequence, an ORF1/2-encoding sequence, an ORF2-encoding sequence, an ORF2/2-encoding sequence, an ORF2/3-encoding sequence, an ORF2/3t-encoding sequence, a three open-reading frame region, a poly(A) signal, and/or a GC-rich region from an Anellovirus described herein (e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

1134. The method of any of the preceding embodiments, wherein the nucleic acid comprises an Anellovirus genome sequence (e.g., as described herein, e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

1135. The method of any of the preceding embodiments, wherein the nucleic acid comprises at least one additional copy of the Anellovirus genome sequence or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto (e.g., a total of 1, 2, 3, 4, 5, or 6 copies).

1136. The method of any of the preceding embodiments, wherein the host cell or helper cell comprises at least one additional copy of the nucleic acid (e.g., a total of 1, 2, 3, 4, 5, or 6 copies).

1137. The method of any of the preceding embodiments, wherein the nucleic acid is circular.

1137A. A method of making an anellosome, e.g., a synthetic anellosome, comprising:

a) providing a host cell comprising:

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

(ii) a nucleic acid molecule, e.g., 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 listed in any of Table 16, 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) culturing the host cell under conditions suitable to make the anellosome.

1137B. The method of embodiment 1137A, further comprising, prior to step (a), introducing the first nucleic acid molecule and/or the second nucleic acid molecule into the host cell.

1137C. The method of embodiment 1137A or 1137B, wherein the second nucleic acid molecule is introduced into the host cell prior to, concurrently with, or after the first nucleic acid molecule.

1137D. The method of embodiment 1137C, wherein the second nucleic acid molecule is integrated into the genome of the host cell.

1137E. The method of embodiment 1137C, wherein the second nucleic acid molecule is a helper (e.g., a helper plasmid or the genome of a helper virus).

1137F. The method of any of embodiments 1137A-1137E, wherein the first nucleic acid comprises one or more of: a TATA box, an initiator element, a cap site, a transcriptional start site, a 5' UTR conserved domain, and/or a GC-rich region from an Anellovirus described herein (e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

1138. A method of delivering an effector to a subject, comprising administering to the subject an anellosome comprising:

(a) a proteinaceous exterior that comprises an ORF1 molecule;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding the effector (e.g., an exogenous effector or an endogenous effector), and a region comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

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

(ii) GCGCTX$_1$CGCGCGCGCCGGGGGGCTGCGCCCCCCC (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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (SEQ ID NO: 168);

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

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

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

(x) GCGCTACGCGCGCGCGCCGGGGGGCTCTGCCCCCCC (SEQ ID NO: 172);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto; and

wherein the genetic element is enclosed within the proteinaceous exterior; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein,

thereby delivering the effector to a subject.

1139. A method of delivering an effector to a subject, comprising administering to the subject an anellosome comprising:

(a) a proteinaceous exterior that comprises an ORF1 molecule;

(b) a genetic element comprising a promoter element, a nucleic acid sequence (e.g., a DNA sequence) encoding the effector (e.g., an exogenous effector or an endogenous effector), and a sequence comprising at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%;

wherein the genetic element is enclosed within the proteinaceous exterior; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein,

thereby delivering the effector to a subject.

1140. A method of delivering an effector to a subject, comprising administering to the subject an anellosome comprising:

(a) a proteinaceous exterior that comprises an ORF1 molecule;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding the effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence;

wherein the genetic element is enclosed within the proteinaceous exterior; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein,

thereby delivering the effector to a subject.

1141. A method of delivering an effector to a target cell, comprising contacting the target cell with an anellosome comprising:

(a) a proteinaceous exterior that comprises an ORF1 molecule;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding the effector (e.g., an exogenous effector or an endogenous effector), and a region comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

(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);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto; and

wherein the genetic element is enclosed within the proteinaceous exterior; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein,

thereby delivering the effector to the target cell.

1142. A method of delivering an effector to a target cell, comprising contacting the target cell with an anellosome comprising:

(a) a proteinaceous exterior that comprises an ORF1 molecule;

(b) a genetic element comprising a promoter element, a nucleic acid sequence (e.g., a DNA sequence) encoding the effector (e.g., an exogenous effector or an endogenous effector), and a sequence comprising at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%;

wherein the genetic element is enclosed within the proteinaceous exterior; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein,

thereby delivering the effector to the target cell.

1143. A method of delivering an effector to a target cell, comprising contacting the target cell with an anellosome comprising:

(a) a proteinaceous exterior that comprises an ORF1 molecule;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding the effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence;

wherein the genetic element is enclosed within the proteinaceous exterior; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein,

thereby delivering the effector to the target cell.

1143A. A method of delivering an effector to a target cell, comprising contacting the target cell with an anellosome comprising:

(i) a genetic element comprising a promoter element and a nucleic acid sequence encoding a therapeutic exogenous effector, wherein the genetic element comprises a sequence having at least 95% sequence identity to the 5' UTR nucleotide sequence from an Anellovirus described herein (e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17); and/or

(ii) a proteinaceous exterior comprising a polypeptide having at least 95% sequence identity to a polypeptide encoded by the ORF1 gene of an Anellovirus described herein (e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17);

wherein the genetic element is enclosed within the proteinaceous exterior; and

optionally wherein the genetic element:

(i) does not comprise a deletion of nucleotides 3436 to 3607 relative to a wild-type TTV-tth8 genome sequence, e.g., as described herein;

(ii) does not comprise a deletion of nucleotides 1432 to 2210 relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein; and/or

(iii) does not comprise a deletion of at least 101 nucleotides relative to a wild-type TTMV-LY2 genome sequence, e.g., as described herein,

thereby delivering the effector to the target cell.

1144. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element does not encode the amino acid sequence of NCBI Accession No. A7XCE8.1.

1145. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the ORF1 molecule comprises 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 to an ORF1 sequence listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1146. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein 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 ORF1 molecule are part of a β-sheet.

1147. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the secondary structure of the ORF1 molecule comprises at least three (e.g., at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20) β-sheets.

1148. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the secondary structure of the ORF1 molecule comprises a ratio of β-sheets to α-helices of at least 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1.

1149. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the ORF1 molecule comprises an arginine-rich region (e.g., having at least 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to an arginine-rich region sequence listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10).

1150. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of embodiment 1149, wherein the arginine-rich region comprises at least 15, 20, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, or 50 consecutive nucleotides comprising at least 40% (e.g., at least

40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 55%, 60%, 65%, 66%, 67%, 68%, 69%, 70%, 75%, 80%, 85%, 90%, or 95%) arginine residues.

1151. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of embodiment 1149 or 1150, wherein the arginine-rich region is located at the N-terminal or C-terminal end of the ORF1 molecule.

1152. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of embodiments 1149-1151, wherein the arginine-rich region has at least 70% (e.g., at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) sequence identity to the amino acid sequence TVVRRRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQC (SEQ ID NO: 808), RRRYARPYRRRHIRRYRRRRRHFRRRR (SEQ ID NO: 809), MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVR (SEQ ID NO: 216), or MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRWRRGRRKTRTYRRRR RFRRRGRK (SEQ ID NO: 186).

1153. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of embodiments 1149-1152, wherein the arginine-rich region has at least 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to an arginine-rich region sequence listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1154. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the ORF1 molecule comprises a jelly-roll domain, e.g., having at least at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to the amino acid sequence of the jelly-roll domain of an ORF1 molecule described herein, e.g., a jelly-roll domain having the amino acid sequence PTYTTIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLD ALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLM MMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACNLQNPFVKP DKLSNNVTLWSLNT (SEQ ID NO: 217), or a jelly-roll domain sequence listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1155. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the ORF1 molecule comprises an N22 domain, e.g., having at

least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to the amino acid sequence of an N22 domain of an ORF1 molecule described herein, e.g., an N22 domain having the amino acid sequence TMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFADFQ KNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQL EAQNKLLQTGPFTPNIQGQLSDNISMFYKFYFK (SEQ ID NO: 219), or an N22 domain sequence listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1156. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the ORF1 molecule localizes to the nucleus of a cell.

1157. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises no more than 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity relative to about 500, 1000, 1100, 1200, 1210, or 1219 consecutive nucleotides of a wild-type *Anellovirus* genome sequence, e.g., as described herein.

1158. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises no more than 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity relative to about 500, 1000, 1500, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3450, 3460, 3470, 3480, 3490, 3500, 3510, 3520, 3530, 3540, 3550, 3560, 3570, or 3580 consecutive nucleotides of a wild-type *Alphatorquevirus* (e.g., a clade 1, 2, or 3 *Alphatorquevirus*) genome sequence, e.g., as described herein.

1159. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises no more than 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity relative to about 500, 1000, 1100, 1200, 1210, or 1219 consecutive nucleotides of a wild-type *Betatorquevirus* genome sequence, e.g., as described herein.

1160. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises

no more than 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity relative to about 500, 1000, 1500, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3120, 3130, 3140, 3141, or 3142 consecutive nucleotides of a wild-type *Gammatorquevirus* genome sequence, e.g., as described herein.

1161. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity relative to at least about 500, 1000, 1500, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3450, 3460, 3470, 3480, 3490, 3500, 3510, 3520, 3530, 3540, 3550, 3560, 3570, or 3580 consecutive nucleotides (e.g., about 500-3580, 1000-3580, 1500-3580, 2000-3580, or 3000-3580 consecutive nucleotides) of a wild-type *Alphatorquevirus* (e.g., a clade 1, 2, or 3 *Alphatorquevirus*) genome sequence, e.g., as described herein.

1162. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity relative to at least about 500, 1000, 1100, 1200, 1210, or 1219 consecutive nucleotides (e.g., about 500-1000, 500-1100, 500-1200, 500-1219, 1000-1100, 1000-1200, or 1000-1219 consecutive nucleotides) of a wild-type *Betatorquevirus* genome sequence, e.g., as described herein.

1163. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity relative to at least about 500, 1000, 1500, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3120, 3130, 3140, 3141, or 3142 consecutive nucleotides (e.g., about 500-3142, 1000-3142, 1500-3142, 2000-3142, or 2500-3142 consecutive nucleotides) of a wild-type *Gammatorquevirus* genome sequence, e.g., as described herein.

1164. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises no more than 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity relative to about 500, 1000, 1100, 1200, 1210, or 1219 consecutive nucleotides of a wild-type TTMV-LY2 genome sequence, e.g., as described herein.

1165. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises no more than 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity relative to about 500, 1000, 1500, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3550, 3560, 3570, 3580, or 3581 consecutive nucleotides of a wild-type TTV-tth8 genome sequence, e.g., as described herein.

1166. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a deletion of at least 1578, 1579, 1580, 1590, 1600, 1650, 1700, 1750, or 2000 nucleotides relative to a wild-type *Anellovirus* genome sequence, e.g., as described herein.

1167. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a deletion of between 1 and 99, 1 and 90, 1 and 80, 1 and 70, 1 and 60, 1 and 50, 10 and 99, 10 and 90, 10 and 80, 10 and 70, 10 and 60, 10 and 50, 20 and 99, 20 and 90, 20 and 80, 20 and 70, 20 and 60, 20 and 50, 30 and 99, 30 and 90, 30 and 80, 30 and 70, 30 and 60, 30 and 50, 40 and 99, 40 and 90, 40 and 80, 40 and 70, 40 and 60, or 40 and 50 nucleotides relative to a wild-type *Anellovirus* genome sequence, e.g., as described herein.

1168. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule does not have a 100 nucleotide deletion, a 172 nucleotide deletion, or a 1577 nucleotide deletion relative to a wild-type *Anellovirus* genome sequence, e.g., as described herein.

1169. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises three or more deletions relative to a wild-type *Anellovirus* genome sequence, e.g., as described herein.

1170. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a region having at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to the nucleic acid sequence:

(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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (SEQ ID NO: 168);

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

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

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

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

1171. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a region having at least 95% (e.g., at least 95, 96, 97, 98, 99, or 100%) sequence identity to the nucleic acid sequence:

(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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (SEQ ID NO: 168);

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

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

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

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

1172. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a region having at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to the nucleic acid sequence CCGCCATCTTAAGTAGTTGAGGCGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACAC CTACTCAAAATGGTGG (SEQ ID NO: 161).

1173. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a region having at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to the nucleic acid sequence CTTAAGTAGTTGAGGCGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACACCTACTCAAAATGGTGGACAATTTCTTCCGGGTCAAAGGTTACAGCCGCCATGTTAAAACACGTGACGTATGACGTCACGGCCGCCATTTTGTGACACAAGATGGCCGACTTCCTTCC (SEQ ID NO: 162).

1174. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 80%.

1175. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises at least 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%.

1176. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises at least 36 consecutive nucleotides having a GC content of at least 80%.

1177. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, further comprising a nucleic acid sequence encoding an ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 of an *Anellovirus*, e.g., a wild-type *Anellovirus*, e.g., as described herein.

1178. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the promoter element, nucleic acid sequence encoding the effector, or protein binding sequence have at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to a promoter element, nucleic acid sequence encoding an effector, or protein binding sequence, respectively, of an *Anellovirus* of any of Tables A1-A12, B1-B5, C1-C5, or 1-18, e.g., as described herein.

1179. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a packaging region positioned 3' relative to the nucleic acid sequence encoding the effector.

1180. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a packaging region positioned 5' relative to the nucleic acid sequence encoding the effector.

1181. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a nucleic acid sequence encoding an *Anellovirus* protein having at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to the amino acid sequence of an ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 of an *Anellovirus* described herein.

1182. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a single-stranded DNA.

1183. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule 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.

1184. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid has at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 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, e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17), or a portion thereof consisting of about 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, or 3000 consecutive nucleotides therefrom.

1185. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the protein binding sequence has at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to the Consensus 5' UTR sequence shown in Table 20.

1186. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the protein binding sequence has at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to the Consensus GC-rich sequence shown in Table 21.

1187. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the protein binding sequence has at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to a 5' UTR sequence shown in Table 38 and to a GC-rich sequence shown in Table 39.

1188. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a sequence having at least 85% sequence identity to the *Anellovirus* 5' UTR conserved domain of the nucleic acid sequence of any one of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17.

1189. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element or isolated nucleic acid molecule comprises a sequence having at least 85% sequence identity to the *Anellovirus* GC-rich region of the nucleic acid sequence of Table A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17.

1190. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the promoter element comprises 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).

1191. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the effector encodes a therapeutic agent, e.g., a therapeutic peptide or polypeptide or a therapeutic nucleic acid.

1192. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of the any of the preceding embodiments, wherein the effector comprises a regulatory nucleic acid, e.g., an miRNA, siRNA, mRNA, lncRNA, RNA, DNA, an antisense RNA, gRNA; a fluorescent tag or marker, an antigen, a peptide, a synthetic or analog peptide from a naturally-bioactive peptide, an agonist or antagonist peptide, an anti-microbial peptide, a pore-forming peptide, a bicyclic peptide, a targeting or cytotoxic peptide, a degradation or self-destruction peptide, a small molecule, an immune effector (e.g., influences susceptibility to an immune response/signal), a death protein (e.g., an inducer of apoptosis or necrosis), a non-lytic inhibitor of a tumor (e.g., an inhibitor of an oncoprotein), an epigenetic modifying agent, an epigenetic enzyme, a transcription factor, a DNA or protein modification enzyme, a DNA-intercalating agent, an efflux pump inhibitor, a nuclear receptor activator or inhibitor, a proteasome inhibitor, a competitive inhibitor for an enzyme, a protein synthesis effector or inhibitor, a nuclease, a protein fragment or domain, a ligand, an antibody, a receptor, or a CRISPR system or component.

1193. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the anellosome is capable of replicating autonomously.

1194. The isolated nucleic acid molecule of any of the preceding embodiments, wherein the expression vector is selected from the group consisting of a plasmid, a cosmid, an artificial chromosome, a phage and a virus.

1195. An isolated cell comprising the isolated nucleic acid or anellosome of any of the preceding embodiments.

1196. The isolated cell of embodiment 195, further comprising an ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 of an *Anellovirus*, e.g., a wild-type *Anellovirus*, e.g., as described herein.

1197. A method of delivering an effector to a subject, comprising administering the polypeptide, complex, anellosome, isolated nucleic acid, isolated cell, or composition of any of the preceding embodiments to the subject; wherein the genetic element or isolated nucleic acid molecule encodes an effector, and wherein the effector is expressed in the subject.

1198. A method of treating a disease or disorder in a subject in need thereof, comprising administering the polypeptide, complex, anellosome, isolated nucleic acid, isolated cell, or composition of any of the preceding embodiments to the subject; wherein the genetic element or isolated nucleic acid molecule encodes a therapeutic agent, and wherein the therapeutic agent is expressed in the subject.

1199. A method of delivering an effector to a cell or population of cells *ex vivo* (e.g., a cell or population of cells obtained from a subject), comprising introducing the polypeptide, complex, anellosome, isolated nucleic acid, isolated cell, or composition of any of the preceding embodiments to the cell or population of cells; wherein the genetic element or isolated nucleic acid molecule encodes an effector, and wherein the effector is expressed in the cell or population of cells.

1200. The anellosome of any of the preceding embodiments, 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.

1201. The anellosome of any of the preceding embodiments, wherein the genetic element has at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 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, e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17).

1202. The anellosome of any of the preceding embodiments, wherein the protein binding sequence has at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to the Consensus 5' UTR sequence shown in Table 38, or to the Consensus GC-rich sequence shown in Table 39, or both of the Consensus 5' UTR sequence shown in Table 38 and to the Consensus GC-rich sequence shown in Table 39.

1203. The anellosome of any of the preceding embodiments, wherein the promoter element comprises 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).

1204. The anellosome of any of the preceding embodiments, wherein the promoter element comprises a TATA box.

1205. The anellosome of any of the preceding embodiments, wherein the promoter element is endogenous to a wild-type *Anellovirus*, e.g., a wild-type *Anellovirus* sequence as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 6, 9, 11, 13, 15, or 17.

1206. The anellosome of any of the preceding embodiments, wherein the promoter element is exogenous to wild-type *Anellovirus*, e.g., a wild-type *Anellovirus* sequence as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 6, 9, 11, 13, 15, or 17.

1207. The anellosome of any of the preceding embodiments, wherein the effector encodes a therapeutic agent, e.g., a therapeutic peptide or polypeptide or a therapeutic nucleic acid.

1208. The anellosome of any of the preceding embodiments, wherein the effector comprises a regulatory nucleic acid, e.g., an miRNA, siRNA, mRNA, lncRNA, RNA, DNA, an antisense RNA, gRNA; a fluorescent tag or marker, an antigen, a peptide, a synthetic or analog peptide from a naturally-bioactive peptide, an agonist or antagonist peptide, an anti-microbial peptide, a pore-forming peptide, a bicyclic peptide, a targeting or cytotoxic peptide, a degradation or self-destruction peptide, a small molecule, an immune effector (e.g., influences susceptibility to an immune response/signal), a death protein (e.g., an inducer of apoptosis or necrosis), a non-lytic inhibitor of a tumor (e.g., an inhibitor of an oncoprotein), an epigenetic modifying agent, an epigenetic enzyme, a transcription factor, a DNA or protein modification enzyme, a DNA-intercalating agent, an efflux pump inhibitor, a nuclear receptor activator or inhibitor, a proteasome inhibitor, a competitive inhibitor for an enzyme, a protein synthesis effector or inhibitor, a nuclease, a protein fragment or domain, a ligand, an antibody, a receptor, or a CRISPR system or component.

1209. The anellosome of any of the preceding embodiments, wherein the effector comprises a miRNA.

1210. The anellosome 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.

1211. The anellosome of any of the preceding embodiments, wherein the effector comprises an miRNA, and decreases expression of a host gene.

1212. The anellosome 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.

1213. The anellosome 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.

1214. The anellosome of any of the preceding embodiments, wherein the sequence encoding the effector has a size of at least about 100 nucleotides.

1215. The anellosome of any of the preceding embodiments, wherein the sequence encoding the effector has a size of about 100 to about 5000 nucleotides.

1216. The anellosome 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.

1217. The anellosome of any of the preceding embodiments, wherein the sequence encoding the effector is situated at, within, or adjacent to (e.g., 5' or 3' to) one or more of the ORF1 locus (e.g., at the C-terminus of the ORF1 locus), the miRNA locus, the 5' noncoding region upstream of the TATA box, the 5' UTR, the 3' noncoding region downstream of the poly-A region, or a noncoding region upstream of the GC-rich region of the genetic element.

1218. The anellosome of embodiment 1217, wherein the sequence encoding the effector is located between the poly-A region and the GC-rich region of the genetic element.

1219. The anellosome of any of the preceding embodiments, wherein the protein binding sequence comprises a nucleic acid sequence 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 the 5' UTR conserved domain or the GC-rich domain of a wild-type *Anellovirus*, e.g., a wild-type *Anellovirus* sequence as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 6, 9, 11, 13, 15, or 17.

1220. The anellosome of any of the preceding embodiments, wherein the genetic element, e.g., protein binding sequence of the genetic element, comprises least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to:

(i) the Consensus 5' UTR nucleic acid sequence shown in Table 38;

(ii) the exemplary TTV 5' UTR nucleic acid sequence shown in Table 38;

(iii) the TTV-CT30F 5' UTR nucleic acid sequence shown in Table 38;

(iv) the TTV-HD23a 5' UTR nucleic acid sequence shown in Table 38;

(v) the TTV-JA20 5' UTR nucleic acid sequence shown in Table 38.;

(vi) the TTV-TJN02 5' UTR nucleic acid sequence shown in Table 38;

(vii) the TTV-tth8 5' UTR nucleic acid sequence shown in Table 38;

(viii) the Consensus GC-rich region shown in Table 39;

(ix) the exemplary TTV GC-rich region shown in Table 39;

(x) the TTV-CT30F GC-rich region shown in Table 39;

(xi) the TTV-JA20 GC-rich region shown in Table 39;

(xii) the TTV-TJN02 GC-rich region shown in Table 39;

(xiii) the TTV-HD23a GC-rich region shown in Table 39; or

(xiv) the TTV-tth8 GC-rich region shown in Table 39.

1221. The anellosome of any of the preceding embodiments, wherein the proteinaceous exterior comprises an exterior protein capable of specifically binding to the protein binding sequence.

1222. The anellosome of any of the preceding embodiments, wherein the proteinaceous exterior comprises one or more of the following: one or more glycosylated proteins, a hydrophilic DNA-binding 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.

1223. The anellosome of any of the preceding embodiments, wherein 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 substantially non-pathogenic in a host.

1224. The anellosome of any of the preceding embodiments, wherein the proteinaceous exterior comprises at least one functional domain that provides one or more functions, e.g., species and/or tissue

and/or cell selectivity, genetic element binding and/or packaging, immune evasion (substantial non-immunogenicity and/or tolerance), pharmacokinetics, endocytosis and/or cell attachment, nuclear entry, intracellular modulation and localization, exocytosis modulation, propagation, and nucleic acid protection.

1225. The anellosome of any of the preceding embodiments, wherein 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).

1226. The anellosome of any of the preceding embodiments, wherein the genetic element is single-stranded.

1227. The anellosome of any of the preceding embodiments, wherein the genetic element is circular.

1228. The anellosome of any of the preceding embodiments, wherein the genetic element is DNA.

1229. The anellosome of any of the preceding embodiments, wherein the genetic element is a negative strand DNA.

1230. The anellosome of any of the preceding embodiments, wherein the genetic element comprises an episome.

1231. The anellosome of any of the preceding embodiments, wherein the anellosome has a lipid content of less than 10%, 5%, 2%, or 1% by weight, e.g., does not comprise a lipid bilayer.

1232. The anellosome of any of the preceding embodiments, wherein the anellosome is resistant to degradation by a detergent (e.g., a mild detergent, e.g., a biliary salt, e.g., sodium deoxycholate) relative to a viral particle comprising an external lipid bilayer, e.g., a retrovirus.

1233. The anellosome of embodiment 1232, wherein at least about 50% (e.g., at least about 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9%) of the anellosome is not degraded after incubation the detergent (e.g., 0.5% by weight of the detergent) for 30 minutes at 37°C.

1234. The anellosome of any of the preceding embodiments, wherein the genetic element comprises a deletion of at least one element, e.g., an element as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17, relative to a wild-type *Anellovirus* sequence, e.g., a wild-type TTV sequence or a wild-type TTMV sequence.

1235. The anellosome of embodiment 1234, wherein the genetic element comprises a deletion comprising a nucleic acid sequence corresponding to:

(i) nucleotides 3436-3607 of a TTV-tth8 sequence, e.g., the nucleic acid sequence shown in Table 5;

(ii) nucleotides 574-1371 and/or nucleotides 1432-2210 of a TTMV-LY2 sequence, e.g., the nucleic acid sequence shown in Table 15;

(iii) nucleotides 1372-1431 of a TTMV-LY2 sequence, e.g., the nucleic acid sequence shown in Table 15; or

(iv) nucleotides 2610-2809 of a TTMV-LY2 sequence, e.g., the nucleic acid sequence shown in Table 15.

1236. The anellosome of any of the preceding embodiments, wherein the genetic element comprises at least 72 nucleotides (e.g., at least 73, 74, 75, etc. nt, optionally less than the full length of the genome) of 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 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.

1237. The anellosome of any of the preceding embodiments, wherein 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), a sequence that encodes a therapeutic mRNA or protein, and a sequence that encodes a cytolytic/cytotoxic RNA or protein.

1238. The anellosome of any of the preceding embodiments, wherein the anellosome further comprises a second genetic element, e.g., a second genetic element enclosed within the proteinaceous exterior.

1239. The anellosome of embodiment 1238, wherein the second genetic element comprises a protein binding sequence, e.g., an exterior protein binding sequence, e.g., a packaging signal, e.g., a 5' UTR conserved domain or GC-rich region, e.g., as described herein.

1240. The anellosome of any of the preceding embodiments, wherein the anellosome does not detectably infect bacterial cells, e.g., infects less than 1%, 0.5%, 0.1%, or 0.01% of bacterial cells.

1241. The anellosome of any of the preceding embodiments, wherein the anellosome is capable of infecting mammalian cells, e.g., human cells, e.g., immune cells, liver cells, epithelial cells, e.g., in vitro.

1242. The anellosome of any of the preceding embodiments, wherein the genetic element integrates at a frequency of less than 10%, 8%, 6%, 4%, 3%, 2%, 1%, 0.5%, 0.2%, 0.1% of the anellosomes that enters the cell, e.g., wherein the anellosome is non-integrating.

1243. The anellosome of any of the preceding embodiments, wherein the genetic element is capable of replicating (e.g., by rolling circle replication), e.g., capable of generating at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 10^2, 2 x 10^2, 5 x 10^2, 10^3, 2 x 10^3, 5 x 10^3, or 10^4 genomic equivalents of the genetic element per cell, e.g., as measured by a quantitative PCR assay.

1244. The anellosome of any of the preceding embodiments, wherein the genetic element is capable of replicating (e.g., by rolling circle replication), e.g., capable of generating at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 10^2, 2 x 10^2, 5 x 10^2, 10^3, 2 x 10^3, 5 x 10^3, or 10^4 more genomic equivalents of the genetic element in a cell, e.g., as measured by a quantitative PCR assay, than were present in the anellosome prior to delivery of the genetic element into the cell.

1244A. The anellosome of embodiment 1243 or 1244, wherein the proteinaceous exterior is provided in *cis* and/or in *trans* relative to the genetic element.

1244B. The anellosome of any of embodiments 1243-1244A, wherein a helper nucleic acid (e.g., a helper virus) in the cell encodes the proteinaceous exterior or a portion thereof (e.g., an ORF1 molecule).

1244C. The anellosome of any of embodiments 1243-1244B, wherein one or more replication factors (e.g., a replicase) is provided in *cis* and/or in *trans* relative to the genetic element.

1244D. The anellosome of embodiment 1244C, wherein a helper nucleic acid (e.g., a helper virus) in the cell encodes the one or more replication factors.

1245. The anellosome of any of the preceding embodiments, wherein the genetic element is not capable of replicating, e.g., wherein the genetic element is altered at a replication origin or lacks a replication origin.

1246. The anellosome of any of the preceding embodiments, wherein the genetic element is not capable of self-replicating, e.g., capable of being replicated without being integrated into a host cell genome.

1247. The anellosome of any of the preceding embodiments, wherein the anellosome is substantially non-pathogenic, e.g., does not induce a detectable deleterious symptom in a subject (e.g., elevated cell death or toxicity, e.g., relative to a subject not exposed to the anellosome).

1248. The anellosome of any of the preceding embodiments, wherein the anellosome is substantially non-immunogenic, e.g., does not induce a detectable and/or unwanted immune response, e.g., as detected according to the method described in Example 4.

1249. The anellosome of embodiment 1248, wherein the substantially non-immunogenic anellosome has an efficacy in a subject that is a least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of the efficacy in a reference subject lacking an immune response.

1250. The anellosome of embodiment 1248 or 1249, wherein the immune response comprises one or more of an antibody specific to the anellosome or a portion thereof, or a product encoded by a nucleic acid thereof; a cellular response (e.g., an immune effector cell (e.g., T cell- or NK cell) response) against the anellosome or cells comprising the anellosome; or macrophage engulfment of the anellosome or cells comprising the anellosome.

1251. The anellosome of any of the preceding embodiments, wherein the anellosome is less immunogenic than an AAV, elicits an immune response below that detected for a comparable quantity of AAV, e.g., as measured by an assay described herein, induces an antibody prevalence of less than 70% (e.g., less than about 60%, 50%, 40%, 30%, 20%, or 10% antibody prevalence) as measured by an assay described herein, or is substantially non-immunogenic.

1252. The anellosome of any of the preceding embodiments, wherein a population of at least 1000 of the anellosomes is capable of delivering at least about 100 copies (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 copies) of the genetic element into one or more of the eukaryotic cells.

1253. The anellosome of any of the preceding embodiments, wherein a population of the anellosomes (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 genome equivalents of the genetic element per cell) is capable of delivering the genetic element into at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more of a population of the eukaryotic cells, e.g., wherein the eukaryotic cells are HEK293T cells, e.g., as described in Example 22.

1254. The anellosome of any of the preceding embodiments, wherein a population of the anellosomes (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 genome equivalents of the genetic element per cell) is capable of delivering at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 8,000, 1×10^4, 1×10^5, 1×10^6, 1×10^7 or greater copies of the genetic element per cell to a population of the eukaryotic cells, e.g., wherein the eukaryotic cells are HEK293T cells, e.g., as described in Example 22.

1255. The anellosome of any of the preceding embodiments, wherein a population of the anellosomes (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 genome equivalents of the genetic element per cell) is capable of delivering 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 5-10, 10-20, 20-50, 50-100, 100-1000, 1000-10^4, 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 the genetic element per cell to a population of the eukaryotic cells, e.g., wherein the eukaryotic cells are HEK293T cells, e.g., as described in Example 22.

1256. The anellosome of any of the preceding embodiments, wherein the anellosome is present after at least two passages.

1257. The anellosome of any of the preceding embodiments, wherein the anellosome was produced by a process comprising at least two passages.

1258. The anellosome of any of the preceding embodiments, wherein the anellosome selectively delivers the effector to, or is present at higher levels in (e.g., preferentially accumulates in), a desired cell type,

tissue, or organ (e.g., bone marrow, blood, heart, GI, skin, photoreceptors in the retina, epithelial linings, or pancreas).

1259. The anellosome of any of the preceding embodiments, wherein the eukaryotic cell is a mammalian cell, e.g., a human cell.

1260. The anellosome of any of the preceding embodiments, wherein the anellosome, 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 the cell.

1261. The anellosome of any of the preceding embodiments, wherein the anellosome is produced in the cell pellet and the supernatant at at least about 10^8-fold (e.g., about 10^5-fold, 10^6-fold, 10^7-fold, 10^8-fold, 10^9-fold, or 10^{10}-fold) genomic equivalents/mL, e.g., relative to the quantity of the anellosome used to infect the cells, after 3-4 days post infection, e.g., using an infectivity assay, e.g., an assay according to Example 7.

1262. A composition comprising the anellosome of any of the preceding embodiments.

1263. A pharmaceutical composition comprising the anellosome of any of the preceding embodiments, and a pharmaceutically acceptable carrier or excipient.

1264. The composition or pharmaceutical composition of embodiment 1262 or 1263, which comprises at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or more anellosomes, e.g., synthetic anellosomes.

1265. The composition or pharmaceutical composition of any of embodiments 1262-1264, which comprises at least 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, or 10^9 synthetic anellosomes.

1266. The composition or pharmaceutical composition of any of embodiments 1262-1265, having one or more of the following characteristics:

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

b) the pharmaceutical composition was made according to good manufacturing practices (GMP);

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

d) the pharmaceutical composition has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants;

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

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

1267. The composition or pharmaceutical composition of any of embodiments 1262-1266, wherein the pharmaceutical composition has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants.

1268. The composition or pharmaceutical composition of embodiment 1267, wherein the contaminant is selected from the group consisting of: 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 anellosomes (e.g., an anellosome other than the desired anellosome, e.g., a synthetic anellosome as described herein), free viral capsid protein, adventitious agents, and aggregates.

1269. The composition or pharmaceutical composition of embodiment 1268, wherein the contaminant is host cell DNA and the threshold amount is about 10 ng of host cell DNA per dose of the pharmaceutical composition.

1270. The composition or pharmaceutical composition of any of embodiments 1262-1269, wherein the pharmaceutical composition comprises less than 10% (e.g., less than about 10%, 5%, 4%, 3%, 2%, 1%, 0.5%, or 0.1%) contaminant by weight.

1271. Use of the anellosome, composition, or pharmaceutical composition of any of the preceding embodiments for treating a disease or disorder (e.g., as described herein) in a subject.

1272. The anellosome, composition, or pharmaceutical composition of any of the preceding embodiments for use in treating a disease or disorder (e.g., as described herein) in a subject.

1273. A method of treating a disease or disorder (e.g., as described herein) in a subject, the method comprising administering the anellosome (e.g., a synthetic anellosome) or the pharmaceutical composition of any of the preceding embodiments to the subject.

1274. A method of modulating, e.g., enhancing or inhibiting, a biological function (e.g., as described herein) in a subject, the method comprising administering the anellosome (e.g., a synthetic anellosome) or the pharmaceutical composition of any of the preceding embodiments to the subject.

1275. The method of any of embodiments 1273-1274, wherein the anellosome does not comprise an exogenous effector.

1276. The method of any of embodiments 1273-1275, wherein the anellosome comprises a wild-type wild-type Anellovirus, e.g., as described herein.

1277. The method of any of embodiments 1273-1276, wherein the administration of the anellosome, e.g., synthetic anellosome, results in delivery of the genetic element into at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more of a population of target cells in the subject.

1278. The method of any of embodiments 1273-1277, wherein the administration of the anellosome, e.g., synthetic anellosome, results in delivery of the effector into at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more of a population of target cells in the subject.

1279. The method of embodiment 1277 or 1278, wherein the target cells comprise mammalian cells, e.g., human cells, e.g., neurons (e.g., motor neurons), myocytes, retinal cell, liver cells, blood cells, skin cells, nerve cells, adipose cells, or endothelial cells, e.g., in vitro.

1280. The method of any of embodiments 1277-1279, wherein the target cells are present in the spinal cord, muscle, retina, or liver.

1281. The method of any of embodiments 1277-1280, wherein the target cells into which the genetic element is delivered each receive at least 10, 50, 100, 500, 1000, 10,000, 50,000, 100,000, or more copies of the genetic element.

1282. The method of any of embodiments 1273-1281, wherein the effector comprises a miRNA and wherein the miRNA reduces the level of a target protein or RNA in a cell or in a population of cells, e.g., into which the anellosome is delivered, e.g., by at least 10%, 20%, 30%, 40%, or 50%.

1283. A method of delivering an anellosome, e.g., a synthetic anellosome, to a cell, comprising contacting the anellosome of any of the preceding embodiments with a cell, e.g., a eukaryotic cell, e.g., a mammalian cell.

1284. The method of embodiment 1283, further comprising contacting a helper virus with the cell, wherein the helper virus comprises a polynucleotide, e.g., 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.

1285. The method of embodiment 1284, wherein the helper virus is contacted with the cell prior to, concurrently with, or after contacting the anellosome with the cell.

1286. The method of embodiment 1283, further comprising contacting a helper polynucleotide with the cell.

1287. The method of embodiment 1286, wherein the helper polynucleotide comprises a sequence polynucleotide encoding an exterior protein, e.g., an exterior protein capable of binding to the exterior protein binding sequence and a lipid envelope.

1288. The method of embodiment 1286, wherein the helper polynucleotide is an RNA (e.g., mRNA), DNA, plasmid, viral polynucleotide, or any combination thereof.

1289. The method of any of embodiments 1286-1288, wherein the helper polynucleotide is contacted with the cell prior to, concurrently with, or after contacting the anellosome with the cell.

1290. The method of any of embodiments 1283-1289, further comprising contacting a helper protein (e.g., a growth factor) with the cell.

1291. The method of embodiment 1290, wherein the helper protein comprises a viral replication protein or a capsid protein.

1292. A host cell comprising the anellosome of any of the preceding embodiments.

1293. A nucleic acid molecule comprising a promoter element, a sequence encoding an effector (e.g., a payload), and an exterior protein binding sequence,

wherein the nucleic acid molecule is a single-stranded DNA, and wherein the nucleic acid molecule is circular and/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 nucleic acid molecule that enters a cell;

wherein the effector does not originate from TTV and is not an SV40-miR-S1;

wherein the nucleic acid molecule does not comprise the polynucleotide sequence of TTMV-LY;

wherein the promoter element is capable of directing expression of the effector in a eukaryotic cell.

1294. A genetic element comprising:

(i) a promoter element and a sequence encoding an effector, e.g., a payload, optionally wherein the effector is exogenous relative to a wild-type *Anellovirus* sequence;

(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% sequence identity to a wild-type *Anellovirus* sequence; or at least 100 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 and/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 a cell.

1295. A method of manufacturing an anellosome composition, comprising:

a) providing a host cell comprising one or more nucleic acid molecules encoding the components of an anellosome, e.g., a synthetic anellosome described herein, e.g., wherein the anellosome comprises a proteinaceous exterior and a genetic element, e.g., a genetic element comprising a promoter element, 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);

b) producing an anellosome from the host cell, thereby making an anellosome; and

c) formulating the anellosomes, e.g., as a pharmaceutical composition suitable for administration to a subject.

1296. A method of manufacturing a synthetic anellosome composition, comprising:

a) providing a plurality of anellosomes, compositions, or pharmaceutical 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 anellosomes, e.g., as a pharmaceutical composition suitable for administration to a subject, e.g., if one or more of the paramaters of (b) meet a specified threshold.

1297. The method of embodiment 1296, wherein the anellosome 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} anellosomes, or wherein the anellosome 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} anellosome genomes per mL.

1298. The method of embodiment 1296 or 1297, wherein the anellosome composition comprises at least 10 ml, 20 ml, 50 ml, 100 ml, 200 ml, 500 ml, 1 L, 2 L, 5 L, 10 L, 20 L, or 50 L.

1299. A reaction mixture comprising the anellosome of any of the preceding embodiments and a helper virus, wherein the helper virus comprises a polynucleotide, e.g., 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.

1300. A reaction mixture comprising the anellosome of any of the preceding embodiments and a second nucleic acid sequence encoding one or more of an amino acid sequence chosen from ORF2, ORF2/2, ORF2/3, ORF2t/3, ORF1, ORF1/1, or ORF1/2 of any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18, 20-37, or D1-D10, or an amino acid sequence having at least 75% (e.g., 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) sequence identity thereto.

1301. The reaction mixture of embodiment 1300, wherein the second nucleic acid sequence is part of the genetic element.

1302. The reaction mixture of embodiment 1301, wherein the second nucleic acid sequence is not part of the genetic element, e.g., the second nucleic acid sequence is comprised by a helper cell or helper virus.

1303. A synthetic anellosome 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 an effector, e.g., a regulatory nucleic acid; and

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

1304. A pharmaceutical composition comprising

a) an anellosome 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 an effector, e.g., 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.

1305. A pharmaceutical composition comprising

a) at least 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, or 10^9 anellosomes (e.g., synthetic anellosomes described herein) 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 an effector, e.g., a regulatory nucleic acid; and

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

b) a pharmaceutical excipient, and, optionally,

c) less than a pre-determined amount of: 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 anellosomes, free viral capsid protein, adventitious agents, endogenous agents, and/or aggregates.

1306. The anellosome or composition of any one of the previous embodiments, further comprising at least one of the following characteristics: the genetic element is a single-stranded DNA; the genetic

element is circular; the anellosome is non-integrating; the anellosome has a sequence, structure, and/or function based on an anellovirus or other non-pathogenic virus, and the anellosome is non-pathogenic.

1307. The anellosome or composition of any one of the previous embodiments, wherein the proteinaceous exterior comprises the non-pathogenic exterior protein.

1308. The anellosome or composition of any one of the previous embodiments, wherein 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.

1309. The anellosome or composition of any one of the previous embodiments, wherein 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, comprises one or more desired carbohydrates (e.g., glycosylations), is pH and temperature stable, is detergent resistant, and is non-immunogenic or non-pathogenic in a host.

1310. The anellosome or composition of any one of the previous embodiments, wherein the sequence encoding the non-pathogenic exterior protein comprise a sequence at least 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identical to one or more sequences or a fragment thereof listed in Table 19.

1311. The anellosome or composition of any one of the previous embodiments, wherein the non-pathogenic exterior protein comprises at least one functional domain that provides one or more functions, e.g., species and/or tissue and/or cell tropism, 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.

1312. The anellosome or composition of any one of the previous embodiments, wherein the effector comprises a regulatory nucleic acid, e.g., an miRNA, siRNA, mRNA, lncRNA, RNA, DNA, an antisense RNA, gRNA; a therapeutic, e.g., fluorescent tag or marker, antigen, peptide therapeutic, synthetic or analog peptide from naturally-bioactive peptide, agonist or antagonist peptide, anti-microbial peptide,

pore-forming peptide, a bicyclic peptide, a targeting or cytotoxic peptide, a degradation or self-destruction peptide, and degradation or self-destruction peptides, small molecule, immune effector (e.g., influences susceptibility to an immune response/signal), a death protein (e.g., an inducer of apoptosis or necrosis), a non-lytic inhibitor of a tumor (e.g., an inhibitor of an oncoprotein), an epigenetic modifying agent, epigenetic enzyme, a transcription factor, a DNA or protein modification enzyme, a DNA-intercalating agent, an efflux pump inhibitor, a nuclear receptor activator or inhibitor, a proteasome inhibitor, a competitive inhibitor for an enzyme, a protein synthesis effector or inhibitor, a nuclease, a protein fragment or domain, a ligand or a receptor, and a CRISPR system or component.

1312b. The anellosome or composition of any one of the previous embodiments, wherein the effector comprises Factor IX or a functional variant thereof.

1313. The anellosome or composition of any one of the previous embodiments, wherein the effector comprises a sequence having at least 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to one or more of the miRNA sequences described herein.

1314. The anellosome or composition of the previous embodiment, wherein the effector, e.g., miRNA, targets a host gene, e.g., modulates expression of the gene.

1315. The anellosome or composition of the previous embodiment, wherein the miRNA comprises a sequence having at least 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to one or more of the miRNA sequences described herein.

1316. The anellosome or composition of any one of the previous embodiments, wherein 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), a sequence that encodes a therapeutic mRNA or protein, and a sequence that encodes a cytolytic/cytotoxic RNA or protein.

1317. The anellosome or composition of any one of the previous embodiments, wherein the genetic element has one or more of the following characteristics: is non-integrating with a host cell's genome, is an episomal nucleic acid, is a single stranded DNA, is about 1 to 10 kb, exists within the nucleus of the cell, is capable of being bound by endogenous proteins, and produces a microRNA that targets host genes.

1318. The anellosome or composition of any one of the previous embodiments, wherein the genetic element comprises at least one viral sequence or at least 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to one or more sequences listed in Table 23, or a fragment thereof (e.g., a fragment encoding an an ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 molecule, and/or a fragment comprising one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

1319. The anellosome or composition of the previous embodiment, wherein the viral sequence is from at least one of a single stranded DNA virus (e.g., Anellovirus, Bidnavirus, Circovirus, Geminivirus, Genomovirus, Inovirus, Microvirus, Nanovirus, Parvovirus, and Spiravirus), a double stranded DNA virus (e.g., Adenovirus, Ampullavirus, Ascovirus, Asfarvirus, Baculovirus, Fusellovirus, Globulovirus, Guttavirus, Hytrosavirus, Herpesvirus, Iridovirus, Lipothrixvirus, Nimavirus, and Poxvirus), a RNA virus (e.g., Alphavirus, Furovirus, Hepatitis virus, Hordeivirus, Tobamovirus, Tobravirus, Tricornavirus, Rubivirus, Birnavirus, Cystovirus, Partitivirus, and Reovirus).

1320. The anellosome or composition of the previous embodiment, wherein the viral sequence is from one or more non-anelloviruses, e.g., adenovirus, herpes virus, pox virus, vaccinia virus, SV40, papilloma virus, an RNA virus such as a retrovirus, e.g., lenti virus, a single-stranded RNA virus, e.g., hepatitis virus, or a double-stranded RNA virus e.g., rotavirus.

1321. The anellosome or composition of any one of the previous embodiments, wherein the protein binding sequence interacts with the arginine-rich region of the proteinaceous exterior.

1322. The anellosome or composition of any one of the previous embodiments, wherein the anellosome is capable of replicating in a mammalian cell, e.g., human cell.

1323. The anellosome or composition of the previous embodiment, wherein the anellosome is non-pathogenic and/or non-integrating in a host cell.

1324. The anellosome or composition of any one of the previous embodiments, wherein the anellosome is non-immunogenic in a host.

1325. The anellosome or composition of any one of the previous embodiments, wherein the anellosome inhibits/enhances one or more viral properties, e.g., selectivity, e.g., infectivity, e.g., immunosuppression/activation, in a host or host cell.

1326. The anellosome or composition of the previous embodiment, wherein the anellosome 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).

1327. The composition of any one of the previous embodiments further comprising at least one virus or vector comprising a genome of the virus, e.g., a variant of the anellosome, e.g., a commensal/native virus.

1328. The composition of any one of the previous embodiments further comprising a heterologous moiety, at least one small molecule, antibody, polypeptide, nucleic acid, targeting agent, imaging agent, nanoparticle, and a combination thereof.

1329. A vector 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 an effector, e.g., a regulatory nucleic acid.

1330. The vector of the previous embodiment, wherein the genetic element fails to integrate with a host cell's genome.

1331. The vector of any one of the previous embodiments, wherein the genetic element is capable of replicating in a mammalian cell, e.g., human cell.

1332. The vector of any one of the previous embodiments further comprising an exogenous nucleic acid sequence, e.g., selected to modulate expression of a gene, e.g., a human gene.

1333. A pharmaceutical composition comprising the vector of any one of the previous embodiments and a pharmaceutical excipient.

1334. The composition of the previous embodiment, wherein the vector is non-pathogenic and/or non-integrating in a host cell.

1335. The composition of any one of the previous embodiments, wherein the vector is non-immunogenic in a host.

1336. The composition of the previous embodiment, wherein the vector is in an amount sufficient to modulate (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).

1337. The composition of any one of the previous embodiments further comprising at least one virus or vector comprising a genome of the virus, e.g., a variant of the anellosome, a commensal/native virus, a helper virus, a non-anellovirus.

1338. The composition of any one of the previous embodiments further comprising a heterologous moiety, at least one small molecule, antibody, polypeptide, nucleic acid, targeting agent, imaging agent, nanoparticle, and a combination thereof.

1339. A method of producing, propagating, and harvesting the anellosome of any one of the previous embodiments.

1340. A method of designing and making the vector of any one of the previous embodiments.

1341. A method of administering to a subject an effective amount of the composition of any one of the previous embodiments.

1342. A method of delivering a nucleic acid or protein payload to a target cell, tissue or subject, the method comprising contacting the target cell, tissue or subject with a nucleic acid composition that comprises (a) a first DNA sequence derived from a virus wherein the first DNA sequence is suffient to enable the production of a particle capable of infecting the target cell, tissue or subject and (a) a second DNA sequence encoding the nucleic acid or protein payload, the improvement comprising:

the first DNA sequence comprises at least 500 (at least 600, 700, 800, 900, 1000, 1200, 1400, 1500, 1600, 1800, 2000) nucleotides having at least 80% (at least 85%, 90%, 95%, 97%, 99%, 100%) sequence identity to a corresponding sequence listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17, or

the first DNA sequence encodes a sequence having at least 80% (at least 85%, 90%, 95%, 97%, 99%, 100%) sequence identity to an ORF listed in Table A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10, or

the first DNA sequence comprises a sequence having at least 90% (at least 95%, 97%, 99%, 100%) sequence identity to a consensus sequence listed in Table 19.

1343. A method of delivering a nucleic acid or protein effector to a target cell, tissue or subject, the method comprising contacting the target cell, tissue or subject with an anellosome of any of the preceding embodiments or a nucleic acid composition that comprises (a) a first DNA sequence derived from a virus wherein the first DNA sequence is sufficient to enable the production of an anellosome of any of the preceding embodiments that can infect the target cell, tissue or subject and (a) a second DNA sequence encoding the nucleic acid or protein effector.

1344. A codon-optimized nucleic acid molecule encoding an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a wild-type *Anellovirus* ORF1, ORF2, or ORF3 amino acid sequence.

1345. The codon-optimized nucleic acid molecule of embodiment 1344, encoding an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a wild-type *Anellovirus* ORF1 amino acid sequence, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10.

1346. A pharmaceutical composition comprising:

(a) an anellosome, e.g., an anellosome of any of the preceding embodiments, and

(b) a carrier chosen from a vesicle, lipid nanoparticle (LNP), red blood cell, exosome (e.g., a mammalian or plant exosome), or fusosome.

2001. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence),

wherein the genetic element has at least:

(i) 72.2% (e.g., at least 72.2, 72.3, 72.4, 72.5, 73, 74, 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A1;

(ii) 68.4% (e.g., at least 68.4, 68.5, 68.6, 68.7, 68.8, 68.9, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A3;

(iii) 81.7% (e.g., at least 81.7, 81.8, 81.9, 82, 83, 84, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A5;

(iv) 92.6% (e.g., at least 92.6, 92.7, 92.8, 92.9, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A7;

(v) 65% (e.g., at least 65, 66, 67, 68, 69, 70, 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A9; or

(vi) 65% (e.g., at least 65, 66, 67, 68, 69, 70, 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A11;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell.

2002. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence),

wherein the genetic element comprises no more than about:

(i) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1010, 1011, 1012, 1013, 1014, 1015, 1016, or 1017 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A1;

(ii) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1100, 1110, 1120, 1130, 1140, 1150, 11160, 1170, 1171, 1172, 1173, or 1174 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A3;

(iii) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 610, 620, 630, 640, 650, 660, 670, 671, or 672 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A5;

(iv) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 260, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, or 280 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A7;

(v) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, or 1000 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A9; or

(vi) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, or 1000 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A11;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell.

2002. An anellosome comprising:

(a) a proteinaceous exterior;

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence),

wherein the genetic element comprises no more than about:

(i) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1010, 1011, 1012, 1013, 1014, 1015, 1016, or 1017 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B1;

(ii) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1100, 1110, 1120, 1130, 1140, 1150, 11160, 1170, 1171, 1172, 1173, or 1174 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B2;

(iii) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 610, 620, 630, 640, 650, 660, 670, 671, or 672 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B3;

(iv) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 260, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, or 280 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B4; or

(v) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, or 1000 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B5;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region);

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the anellosome is configured to deliver the genetic element into a eukaryotic cell.

2003. The anellosome of any of the preceding embodiments, wherein the genetic element is not a naturally occurring sequence (e.g., comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region)), relative 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, e.g., as listed in any of Tables B1-B5, A1, A3, A5, A7, A9, A11, 1, 3, 5, 7, 9, 11, or 13).

2004. The anellosome of any of the preceding embodiments, comprising a polypeptide comprising an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to the amino acid sequence of an Anellovirus ORF1 molecule (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12).

2005. The anellosome of embodiment 2004, wherein the proteinaceous exterior comprises the polypeptide.

2006. The anellosome of embodiment 2005, wherein at least 60% (e.g., at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) of protein in the proteinaceous exterior comprises the polypeptide.

2007. The anellosome of any of the preceding embodiments, wherein at least 60% (e.g., at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) of protein in the proteinaceous exterior comprises an ORF1 molecule.

2008. The anellosome of any of the preceding embodiments, comprising a nucleic acid molecule (e.g., in the genetic element) encoding an amino acid sequence having at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to the amino acid sequence of an Anellovirus ORF1 molecule (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12).

2009. The anellosome of any of the preceding embodiments, wherein the genetic element comprises a region comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

(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);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto.

2010. The anellosome of any of the preceding embodiments, wherein the genetic element comprises a 5' UTR region and/or a GC-rich region as described herein (e.g., as listed in Table 38 or 39, respectively).

2011. An isolated nucleic acid molecule (e.g., an expression vector) comprising a genetic element comprising at least:

(i) 72.2% (e.g., at least 72.2, 72.3, 72.4, 72.5, 73, 74, 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A1;

(ii) 68.4% (e.g., at least 68.4, 68.5, 68.6, 68.7, 68.8, 68.9, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A3;

(iii) 81.7% (e.g., at least 81.7, 81.8, 81.9, 82, 83, 84, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A5;

(iv) 92.6% (e.g., at least 92.6, 92.7, 92.8, 92.9, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A7;

(v) 65% (e.g., at least 65, 66, 67, 68, 69, 70, 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A9; or

(vi) 65% (e.g., at least 65, 66, 67, 68, 69, 70, 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to an *Anellovirus* sequence as listed in Table A11;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

2012. An isolated nucleic acid molecule (e.g., an expression vector) comprising a genetic element comprising no more than about:

(i) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1010, 1011, 1012, 1013, 1014, 1015, 1016, or 1017 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A1;

(ii) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1100, 1110, 1120, 1130, 1140, 1150, 11160, 1170, 1171, 1172, 1173, or 1174 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A3;

(iii) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 610, 620, 630, 640, 650, 660, 670, 671, or 672 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A5;

(iv) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 260, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, or 280 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A7;

(v) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, or 1000 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A9; or

(vi) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, or 1000 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table A11;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

2012A. An isolated nucleic acid molecule (e.g., an expression vector) comprising a genetic element comprising no more than about:

(i) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1010, 1011, 1012, 1013, 1014, 1015, 1016, or 1017 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B1;

(ii) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1100, 1110, 1120, 1130, 1140, 1150, 11160, 1170, 1171, 1172, 1173, or 1174 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B2;

(iii) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 610, 620, 630, 640, 650, 660, 670, 671, or 672 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B3;

(iv) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 260, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, or 280 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B4; or

(v) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, or 1000 nucleotide differences, e.g., substitutions, insertions or deletions, relative to an *Anellovirus* sequence as listed in Table B5;

optionally, wherein the genetic element comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type *Anellovirus* genome sequence (e.g., as described herein), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region).

2013. The isolated nucleic acid molecule of any of the preceding embodiments, wherein the genetic element is not a naturally occurring sequence (e.g., comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration), e.g., an insertion, substitution, enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of a TATA box, cap site, transcriptional start site, 5' UTR, open reading frame (ORF), poly(A) signal, or GC-rich region)), relative 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, e.g., as listed in any of Tables B1-B5, A1, A3, A5, A7, A9, A11, 1, 3, 5, 7, 9, 11, or 13).

2014. The isolated nucleic acid molecule of any of the preceding embodiments, wherein the isolated nucleic acid molecule comprises a genetic element encoding an ORF1 molecule (e.g., an ORF1 molecule as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12, or a polypeptide having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto);

wherein:

(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 ORF1 molecule are part of a β-sheet;

(ii) the secondary structure of the ORF1 molecule comprises at least three (e.g., at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20) β-sheets;

(iii) the secondary structure of the ORF1 molecule comprises a ratio of β-sheets to α-helices of at least 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1; and

2015. The isolated nucleic acid molecule of any of the preceding embodiments, comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

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

(ii) GCGCTX$_1$CGCGCGCGCGCCGGGGGGGCTGCGCCCCCCC (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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (SEQ ID NO: 168);

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

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

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

(x) GCGCTACGCGCGCGCGCCGGGGGGCTCTGCCCCCCC (SEQ ID NO: 172);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto.

2016. The isolated nucleic acid molecule of any of the preceding embodiments, comprising at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%.

2017. The isolated nucleic acid molecule of any of the preceding embodiments, wherein the genetic element further comprises one or more of: a TATA box, an initiator element, a cap site, a transcriptional start site, a 5' UTR conserved domain, an ORF1-encoding sequence, an ORF1/1-encoding sequence, an ORF1/2-encoding sequence, an ORF2-encoding sequence, an ORF2/2-encoding sequence, an ORF2/3-encoding sequence, an ORF2/3t-encoding sequence, a three open-reading frame region, a poly(A) signal, and/or a GC-rich region from an Anellovirus described herein (e.g., as listed in any of Tables B1-B5, A1, A3, A5, A7, A9, or A11), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

2018. The isolated nucleic acid molecule of any of the preceding embodiments, wherein the genetic element further comprises at least one or two copies (e.g., 1, 2, 3, 4, 5, or 6 copies) of an Anellovirus genome sequence (e.g., as described herein, e.g., as listed in any of Tables B1-B5, A1, A3, A5, A7, A9, A11, 1, 3, 5, 7, 9, 11, 13, 15, or 17), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

2019. The isolated nucleic acid molecule of any of the preceding embodiments, further comprising at least one additional copy of the genetic element (e.g., a total of 1, 2, 3, 4, 5, or 6 copies).

2020. The isolated nucleic acid molecule of any of the preceding embodiments, wherein the isolated nucleic acid molecule is circular.

2021. An isolated nucleic acid composition (e.g., comprising one, two, or more nucleic acid molecules) comprising the isolated nucleic acid of any of the preceding embodiments.

2022. The isolated nucleic acid of any of the preceding embodiments, wherein the genetic element further comprises a promoter element, a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an endogenous effector or an exogenous effector), and/or a protein binding sequence (e.g., an exterior protein binding sequence).

2022A. The isolated nucleic acid molecule of any of the preceding embodiments, wherein the genetic element comprises an insertion or substitution in the hyper-variable domain (HVD) of the ORF1.

2023. The anellosome or isolated nucleic acid molecule of any of the preceding embodiments, wherein the genetic element comprises one or more of a TATA box, initiator site, 5' UTR conserved domain,

ORF1, ORF2, ORF2 downstream sequence, ORF2, ORF3, and/or GC-rich region, or sequences having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity thereto, e.g., as shown in any of Tables B1-B5, A1, A3, A5, A7, A9, or A11.

2024. The anellosome or isolated nucleic acid of any of the preceding embodiments, which comprises (e.g., in the proteinaceous exterior) or encodes one or more polypeptides comprising an amino acid sequence chosen from ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 of any of Tables C1-C5, A2, A4, A6, A8, A10, or A12, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

2025. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element comprises a sequence comprising at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%.

2026. The anellosome or isolated nucleic acid of embodiment 2025, wherein the genetic element comprises at least 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides having a GC content of at least 80%.

2027. The anellosome or isolated nucleic acid of embodiment 2025, wherein the genetic element comprises at least 36 consecutive nucleotides having a GC content of at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, or 80.6%.

2028. The anellosome or isolated nucleic acid of embodiment 2025, wherein the genetic element comprises at least 36 consecutive nucleotides having a GC content of at least 80%.

2029. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element comprises a region (e.g., a packaging region) comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of the nucleic acid sequence:

(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);

or a nucleic acid sequence having at least 75, 76, 77, 78, 79, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity thereto.

2030. The anellosome or isolated nucleic acid of embodiment 2029, wherein the packaging region is positioned 3' relative to the nucleic acid sequence encoding the effector.

2031. A polypeptide comprising one or more of:

(a) a first region comprising an amino acid sequence having 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 of an Anellovirus ORF1 molecule described herein (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12);

(b) a second region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to a jelly-roll region sequence of an Anellovirus ORF1 molecule described herein (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12);

(c) a third region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an N22 domain sequence of an Anellovirus ORF1 molecule described herein (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12); and/or

(d) a fourth region comprising an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an Anellovirus ORF1 C-terminal domain (CTD) sequence of an Anellovirus ORF1 molecule described herein (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12);

wherein the 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), e.g., an insertion, substitution, chemical or enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of an arginine-rich region, jelly-roll domain, HVR, N22, or CTD, e.g., as described herein).

2031A. The polypeptide of embodiment 2031, comprising one or more of:

(a) a first region comprising an amino acid sequence having at least 90% sequence identity to an arginine-rich region sequence of an Anellovirus ORF1 molecule described herein (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12);

(b) a second region comprising an amino acid sequence having at least 90% sequence identity to a jelly-roll region sequence of an Anellovirus ORF1 molecule described herein (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12);

(c) a third region comprising an amino acid sequence having at least 90% sequence identity to an N22 domain sequence of an Anellovirus ORF1 molecule described herein (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12); and/or

(d) a fourth region comprising an amino acid sequence having at least 90% sequence identity to an Anellovirus ORF1 C-terminal domain (CTD) sequence of an Anellovirus ORF1 molecule described herein (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12);

wherein the 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), e.g., an insertion, substitution, chemical or enzymatic modification, and/or deletion, e.g., a deletion of a domain (e.g., one or more of an arginine-rich region, jelly-roll domain, HVR, N22, or CTD, e.g., as described herein).

2032. The polypeptide of embodiment 2031, wherein the polypeptide comprises:

(i) the first region and the second region;

(ii) the first region and the third region;

(iii) the first region and the fourth region;

(iv) the second region and the third region;

(v) the second region and the fourth region;

(vi) the third region and the fourth region;

(vii) the first region, the second region, and the third region;

(viii) the first region, the second region, and the fourth region;

(ix) the first region, the third region, and the fourth region; or

(x) the second region, the third region, and the fourth region.

2033. The polypeptide of embodiment 2031 or 2032, wherein the polypeptide comprises, in N-terminal to C-terminal order, the first region, the second region, the third region, and the fourth region.

2034. The polypeptide of any of the preceding embodiments, further comprising an amino acid sequence, e.g., a hypervariable region (HVR) sequence (e.g., the HVR sequence of an Anellovirus ORF1 molecule, e.g., as described herein), wherein the amino acid sequence 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).

2035. The polypeptide of embodiment 2034, wherein the HVR comprises an amino acid sequence having at least 30% (e.g., at least about 30, 35, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an Anellovirus ORF1 HVR sequence of an Anellovirus ORF1 molecule described herein (e.g., an Anellovirus ORF1 sequence as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12).

2036. The polypeptide of embodiment 2034 or 2035, wherein the HVR sequence is positioned between the second region and the third region.

2037. The polypeptide of any of embodiments 2034-2036, wherein the HVR comprises one or more features of an HVR as described herein.

2038. A polypeptide comprising the amino acid sequence of ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 of any of Tables C1-C5, A2, A4, A6, A8, A10, or A12, or having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto, and wherein the polypeptide further comprises at least one difference (e.g., a mutation or chemical modification) relative to a wild-type *Anellovirus* ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 sequence (e.g., as described herein, e.g., as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12), e.g., a conjugation, addition, insertion, substitution, and/or deletion, e.g., a deletion of a domain.

2039. A polypeptide comprising an amino acid sequence of ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 of any of Tables C1-C5, A2, A4, A6, A8, A10, or A12, or having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

2040. A polypeptide having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, or 98%, but no more than 99%, sequence identity to an amino acid sequence chosen from ORF1, ORF2, ORF2, or ORF3 of any of Tables C1-C5, A2, A4, A6, A8, A10, or A12.

2041. A polypeptide having at least 1, but no more than 2, 5, 10, 20, 50, or 100 amino acid differences, e.g., substitutions, insertions or deletions, relative to an amino acid sequence chosen from ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 of any of Tables C1-C5, A2, A4, A6, A8, A10, or A12.

2042. The polypeptide of any of the preceding embodiments, wherein the polypeptide is an isolated polypeptide.

2043. A complex comprising:

(a) the polypeptide of any of the preceding embodiments, and

(b) a genetic element comprising a promoter element and a nucleic acid sequence (e.g., a DNA sequence) encoding an effector (e.g., an exogenous effector or an endogenous effector), and a protein binding sequence.

2044. The complex of embodiment 2043, wherein the complex comprises one or more features of a complex as described herein.

2045. A fusion protein comprising a first amino acid sequence chosen from the ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 molecule of any of Tables C1-C5, A2, A4, A6, A8, A10, or A12, or having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto, and a heterologous moiety.

2046. A fusion protein comprising a first amino acid sequence chosen from the ORF1 molecule of any of Tables C1-C5, A2, A4, A6, A8, A10, or A12, or having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto, and a heterologous moiety.

2047. The fusion protein of any of the preceding embodiments, wherein the heterologous moiety comprises a targeting moiety.

2048. The fusion protein of any of the preceding embodiments, wherein the first amino acid sequence comprises at least one difference (e.g., a mutation or chemical modification) relative to a wild-type *Anellovirus* ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 sequence (e.g., as described herein, e.g., as listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12), e.g., a conjugation, addition, insertion, substitution, and/or deletion, e.g., a deletion of a domain.

2049. A host cell comprising the anellosome, isolated nucleic acid, fusion protein, or polypeptide of any of the preceding embodiments.

2050. A reaction mixture comprising the anellosome of any of the preceding embodiments and a helper virus, wherein the helper virus comprises a polynucleotide, e.g., a polynucleotide encoding an exterior protein, e.g., an exterior protein that binds to the exterior protein binding sequence and, optionally, a lipid envelope.

2051. A method of treating a disease or disorder in a subject, the method comprising administering an anellosome, isolated nucleic acid molecule, fusion protein, or polypeptide of any of the preceding embodiments or the pharmaceutical composition of any of the preceding embodiments to the subject.

2052. The method of embodiment 2051, wherein the disease or disorder is chosen from an immune disorder, infectious disease, inflammatory disorder, autoimmune condition, cancer (e.g., a solid tumor), and a gastrointestinal disorder.

2053. Use of the anellosome, isolated nucleic acid, fusion protein, or polypeptide of any of the preceding embodiments for treating a disease or disorder in a subject.

2054. The use of embodiment 2053, wherein the disease or disorder is chosen from an immune disorder, infectious disease, inflammatory disorder, autoimmune condition, cancer (e.g., a solid tumor, e.g., lung cancer), and a gastrointestinal disorder.

2055. The anellosome, isolated nucleic acid, composition, or pharmaceutical composition of any of the preceding embodiments for use in treating a disease or disorder in a subject.

2055A. The anellosome, isolated nucleic acid, composition, or pharmaceutical composition of any of the preceding embodiments for use as a medicament.

2056. A method of modulating, e.g., inhibiting or enhancing, a biological function in a subject, the method comprising administering an anellosome, isolated nucleic acid, fusion protein, or polypeptide of any of the preceding embodiments or the pharmaceutical composition of any of the preceding embodiments to the subject.

2057. A method of delivering an anellosome to a cell, comprising contacting the anellosome, isolated nucleic acid, fusion protein, or polypeptide of any of the preceding embodiments with a cell, e.g., a eukaryotic cell, e.g., a mammalian cell.

2058. The method of embodiment 2057, further comprising contacting a helper virus with the cell, wherein the helper virus comprises a polynucleotide, e.g., a polynucleotide encoding an exterior protein, e.g., an exterior protein that binds to the exterior protein binding sequence and, optionally, a lipid envelope.

2059. The method of embodiment 2058, wherein the helper virus is contacted with the cell prior to, concurrently with, or after contacting the anellosome with the cell.

2060. The method of embodiment 2057, further comprising contacting a helper polynucleotide with the cell.

2061. The method of embodiment 2060, wherein the helper polynucleotide comprises a sequence polynucleotide encoding an exterior protein, e.g., an exterior protein that binds to the exterior protein binding sequence and a lipid envelope.

2062. The method of embodiment 2060, wherein the helper polynucleotide is an RNA (e.g., mRNA), DNA, plasmid, viral polynucleotide, or any combination thereof.

2063. The method of any of embodiments 2060-2062, wherein the helper polynucleotide is contacted with the cell prior to, concurrently with, or after contacting the anellosome with the cell.

2064. The method of any of embodiments 2057-2063, further comprising contacting a helper protein with the cell.

2065. The method of embodiment 2064, wherein the helper protein comprises a viral replication protein or a capsid protein.

2066. A method of delivering a nucleic acid or protein effector to a target cell, tissue or subject, the method comprising contacting the target cell, tissue or subject with a nucleic acid composition that comprises (a) a first DNA sequence derived from a virus wherein the first DNA sequence is sufficient to enable the production of a particle that can infect the target cell, tissue or subject and (a) a second DNA sequence encoding the nucleic acid or protein effector, the improvement comprising:

the first DNA sequence comprises at least 500 (at least 600, 700, 800, 900, 1000, 1200, 1400, 1500, 1600, 1800, 2000) nucleotides having at least 80% (at least 85%, 90%, 95%, 97%, 99%, 100%) sequence identity to a corresponding sequence listed in any of Tables B1-B5, A1, A3, A5, A7, A9, or A11, or

the first DNA sequence encodes a sequence having at least 80% (at least 85%, 90%, 95%, 97%, 99%, 100%) sequence identity to an *Anellovirus* ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, and/or ORF3 molecule (e.g., listed in any of Tables C1-C5, A2, A4, A6, A8, A10, or A12).

2067. A method of manufacturing an anellosome composition, comprising:

a) providing a host cell comprising one or more nucleic acid molecules encoding the components of an anellosome of any of the preceding embodiments, wherein the anellosome comprises a proteinaceous exterior and a genetic element, e.g., a genetic element comprising a promoter element, a sequence encoding an effector, (e.g., an endogenous effector or an exogenous effector), and a protein binding sequence (e.g., an exterior protein binding sequence, e.g., a packaging signal);

b) producing an anellosome from the host cell, thereby making an anellosome; and

c) formulating the anellosomes, e.g., as a pharmaceutical composition suitable for administration to a subject;

optionally wherein the one or more nucleic acid molecules encodes a helper protein.

2068. A method of manufacturing an anellosome composition, comprising:

a) providing a plurality of anellosomes 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); and

c) formulating the plurality of anellosomes, 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.

2069. The method of embodiment 2068, wherein the anellosome 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} anellosomes.

2070. The method of embodiment 2068 or 2069, wherein the anellosome composition comprises at least 10 ml, 20 ml, 50 ml, 100 ml, 200 ml, 500 ml, 1 L, 2 L, 5 L, 10 L, 20 L, or 50 L.

2071. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element is configured to replicate in a mammalian cell, e.g., a human cell.

2072. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element further comprises an exogenous nucleic acid sequence, e.g., selected to modulate expression of a gene, e.g., a human gene.

2073. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein at least 60% (e.g., at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the protein binding sequence consists of G or C.

2074. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element comprises a sequence of at least 80, 90, 100, 110, 120, 130, or 140 nucleotides in length, which consists of G or C in at least 70% (e.g., at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) or about 70-100%, 75-95%, 80-95%, 85-95%, or 85-90% of the positions.

2075. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the protein binding sequence binds an arginine-rich region of the proteinaceous exterior.

2076. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the proteinaceous exterior comprises an exterior protein that specifically binds to the protein binding sequence.

2077. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein 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).

2078. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element is single-stranded.

2079. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element is circular.

2080. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element is DNA.

2081. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element is a negative strand DNA.

2082. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the genetic element comprises an episome.

2083. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the anellosome is present at higher levels in (e.g., preferentially accumulates in) a desired organ or tissue relative to other organs or tissues.

2084. The anellosome or isolated nucleic acid of any of the preceding embodiments, wherein the eukaryotic cell is a mammalian cell, e.g., a human cell.

2085. A composition comprising the anellosome or isolated nucleic acid of any of the preceding embodiments.

2086. A pharmaceutical composition comprising the anellosome or isolated nucleic acid of any of the preceding embodiments, and a pharmaceutically acceptable carrier or excipient.

2087. A pharmaceutical composition comprising

a) at least 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, or 10^9 anellosomes of any of the preceding embodiments;

b) a pharmaceutical excipient, and, optionally,

c) less than a pre-determined amount of: 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 anellosomes, free viral capsid protein, adventitious agents, and/or aggregates.

2088. The composition or pharmaceutical composition of embodiment 2085 or 2086, which comprises at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or more anellosomes, e.g., synthetic anellosomes.

2089. The composition or pharmaceutical composition of any of embodiments 2085-2088, which comprises at least 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, or 10^9 anellosomes.

2090. A pharmaceutical composition comprising

a) at least 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, or 10^9 anellosomes of any of the preceding embodiments;

b) a pharmaceutical excipient, and, optionally,

c) less than a pre-determined amount of: 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 anellosomes, free viral capsid protein, adventitious agents, and/or aggregates.

2091. The composition or pharmaceutical composition of any of embodiments 2085-2090, having one or more of the following characteristics:

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

b) the pharmaceutical composition was made according to good manufacturing practices (GMP);

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

d) the pharmaceutical composition has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants;

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

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

2092. The composition or pharmaceutical composition of any of embodiments 2085-2091, wherein the pharmaceutical composition has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants.

2093. The composition or pharmaceutical composition of embodiment 92, wherein the contaminant is selected from the group consisting of: 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 anellosomes (e.g., a anellosome other than the desired anellosome, e.g., a synthetic anellosome as described herein), free viral capsid protein, adventitious agents, and aggregates.

2094. The composition or pharmaceutical composition of embodiment 2093, wherein the contaminant is host cell DNA and the threshold amount is about 500 ng of host cell DNA per dose of the pharmaceutical composition.

2095. The composition or pharmaceutical composition of any of embodiments 2085-2094, wherein the pharmaceutical composition comprises less than 10% (e.g., less than about 10%, 5%, 4%, 3%, 2%, 1%, 0.5%, or 0.1%) contaminant by weight.

2096. The method of any of the preceding embodiments, wherein the anellosome does not comprise an exogenous effector.

2097. The method of any of the preceding embodiments, wherein the administration of the anellosome, e.g., synthetic anellosome, results in delivery of the genetic element into at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more of a population of target cells in the subject.

2098. The method of any of the preceding embodiments, wherein the administration of the anellosome, e.g., synthetic anellosome, results in delivery of the exogenous effector into at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more of a population of target cells in the subject.

2099. The method of embodiment 2097 or 2098, wherein the target cells comprise mammalian cells, e.g., human cells, e.g., immune cells, liver cells, lung epithelial cells, e.g., in vitro.

2100. The method of any of embodiments 2097-2099, wherein the target cells are present in the liver or lung.

2101. The method of any of embodiments 2097-2100, wherein the target cells into which the genetic element is delivered each receive at least 10, 50, 100, 500, 1000, 10,000, 50,000, 100,000, or more copies of the genetic element.

2102. The method of any of the preceding embodiments, wherein the effector comprises a miRNA, and optionally wherein the miRNA reduces the level of a target protein or RNA in a cell or in a population of cells, e.g., into which the anellosome is delivered, e.g., by at least 10%, 20%, 30%, 40%, or 50%.

2103. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element (e.g., the 5' UTR of the genetic element) physically associates with (e.g., binds) to the proteinaceous exterior (e.g., to an ORF1 molecule in a proteinaceous exterior).

2104. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element enclosed within the proteinaceous exterior is resistant to endonuclease digestion, e.g., as determined according to the method described in Martin et al. (2013, *Hum. Gene Ther. Methods* 24(4): 253-269; incorporated herein by reference in its entirety); optionally wherein the amount of DNase used is about 60 U/ml or about 300 U.

2105. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element comprises a sequence of at least 100 nucleotides in length, which consists of G or C at at least 80% of the positions.

2106. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element is circular, single stranded DNA.

2107. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein 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).

2108. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element integrates at a frequency of less than 1% of the anellosomes that enters the mammalian cell.

2109. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the promoter element is exogenous or endogenous to wild-type Anellovirus.

2110. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the exogenous effector is a therapeutic exogenous effector, e.g., a therapeutic peptide, a therapeutic polypeptide, or a therapeutic nucleic acid (e.g., an miRNA).

2111. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein a population of at least 1000 (e.g., at least 1000, 1500, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000, 20,000, 50,000, 75,000, 100,000, 200,000, 500,000, 1,000,000 or more) of the anellosomes delivers at least 100 (e.g., at least 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000, 50,000, 100,000, or more) copies of the genetic element into one or more of the mammalian cells.

2112. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the anellosome comprises one or more polypeptides comprising one or more of an amino acid sequence chosen from an *Anellovirus* ORF2, ORF2/2, ORF2/3, ORF1, ORF1/1, or ORF1/2 (e.g., as described herein) or an amino acid sequence having at least 95% sequence identity thereto.

2113. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element comprises a nucleic acid sequence encoding an amino acid sequence chosen from an *Anellovirus* ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 (e.g., as described herein), or an amino acid sequence having at least 95% sequence identity thereto.

2114. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the anellosome does not comprise a polynucleotide encoding one or both of a replication factor and a capsid protein, or wherein the anellosomes is replication defective.

2115. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the anellosome is contacted to a cell *in vitro* or *in vivo*.

2116. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the anellosome does not comprise a polypeptide having at least 95% sequence identity to an *Anellovirus* ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 (e.g., as described herein).

2117. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element is capable of being amplified by rolling circle replication (e.g., in a cell, e.g., a host cell, e.g., a mammalian cell, e.g., a human cell, e.g., a HEK293T or A549 cell), e.g., to produce at least 2, 4, 8, 16, 32, 64, 128, 256, 518, or 1024 copies.

2118. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element is produced from a double-stranded circular DNA molecule.

2119. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of embodiment 2118, wherein the double-stranded circular DNA molecule is produced by in vitro circularization.

2118. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element is produced from a DNA molecule comprising two copies of the nucleic acid sequence of the genetic element.

2119. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the two copies of the nucleic acid sequence of the genetic element are arranged in tandem in the DNA molecule.

2120. A nucleic acid molecule comprising two copies of a nucleic acid sequence comprising the 5' UTR of an anellosome genetic element (e.g., the genetic element of any of the preceding embodiments).

2121. A nucleic acid molecule comprising a promoter element; a nucleic acid sequence encoding an exogenous effector; a 5' UTR sequence as listed in any of Tables B1-B5, or a nucleic acid sequence

having at least 85% (e.g., at least 85%, 90%, 95% 96%, 97%, 98%, 99%, or 100%) identity thereto; and a GC-rich region as listed in any of Tables B1-B5, or a nucleic acid sequence having at least 85% (e.g., at least 85%, 90%, 95% 96%, 97%, 98%, 99%, or 100%) identity thereto.

2122. The nucleic acid molecule of embodiment 2121, wherein the nucleic acid molecule is single-stranded or double stranded.

2123. The nucleic acid molecule of embodiment 2121, wherein the nucleic acid molecule is circular.

2124. The polypeptide, complex, anellosome, isolated nucleic acid, cell, composition, or method of any of the preceding embodiments, wherein the genetic element comprises a 5' UTR comprising the nucleic acid sequence of:

CGGGAGCCX$_1$CGAGGTGAGTGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGC AGTCTAGCGGAACGGG, wherein X_1 is C or absent,

or a nucleic acid sequence at least 95% identical thereto.

3001. A synthetic anellosome comprising:

(I) a genetic element comprising:

(a) a promoter element,

(b) a nucleic acid sequence encoding an exogenous effector, wherein the nucleic acid sequence is operably linked to the promoter element,

wherein the exogenous effector comprises a polypeptide that, when mutated, causes a human disease, or a functional variant of said polypeptide, chosen from:

(i) a therapeutic enzyme (e.g., Factor IX, UGT1A1, or OCA1),

(ii) a structural protein (e.g., SMN, dystrophin), or

(iii) a complement protein (e.g., C1); and

(c) a 5' UTR domain comprising:

a nucleic acid sequence of nucleotides 323 – 393 of SEQ ID NO: 54, or a nucleic acid sequence at least 85% identical thereto;

a nucleic acid sequence of any of SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115 , SEQ ID NO: 116, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119 or a nucleic acid sequence at least 85% identical thereto; or

a nucleic acid sequence of nucleotides 117 – 187 of SEQ ID NO: 61, or a nucleic acid sequence at least 85% identical thereto;

(II) a proteinaceous exterior comprising an ORF1 molecule;

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the synthetic anellosome is capable of delivering the genetic element into a human cell.

3002. The synthetic anellosome of embodiment 3001, wherein the exogenous effector comprises Factor IX or a functional variant thereof.

3003. The synthetic anellosome of emnodiment 3001, wherein the exogenous effector is an enzyme of Table A or a functional variant thereof.

3004. The synthetic anellosome of embodiment 3001, wherein the effector is a polypeptide of Table B, or a functional variant thereof.

3005. The synthetic anellosome of embodiment 3001, wherein the ORF1 molecule comprises the amino acid sequence of SEQ ID NO: 217, or an amino acid sequence having least 90% identity thereto.

3006. The synthetic anellosome of any of the preceding embodiments, wherein the ORF1 molecule is encoded by nucleotides 612-2612 of SEQ ID NO: 54.

3007. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element comprises the nucleic acid sequence of nucleotides 2868-2929 of SEQ ID NO: 54, or a nucleic acid sequence having at least 85% sequence identity thereto.

3008. The synthetic anellosome of any of the preceding embodiments, wherein the ORF1 molecule comprises an amino acid sequence comprising one or more of the amino acid sequences of an arg-rich region, jelly-roll domain, hypervariable domain, N22 domain, and/or C-terminal domain as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

3009. The synthetic anellosome of any of the preceding embodiments, wherein the ORF1 molecule comprises the amino acid sequence of SEQ ID NO: 58, or a nucleic acid sequence having at least 85% sequence identity thereto.

3010. The synthetic anellosome of any of the preceding embodiments, further comprising a polypeptide comprising the amino acid sequence of an ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

3011. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element encodes the amino acid sequence of an ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

3012. The synthetic anellosome of any of the preceding embodiments, wherein the synthetic anellosome does not comprise a polypeptide comprising the amino acid sequence of an ORF2, ORF2/2, ORF2/3, TAIP, ORF1/1, or ORF1/2 as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

3013. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element does not encode the amino acid sequence of an ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

3014. The synthetic anellosome of any of the preceding embodiments, wherein the ORF1 molecule comprises the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829), wherein X^n is each independently a contiguous sequence of any n amino acids.

3015. The synthetic anellosome of embodiment 3014, wherein the ORF1 molecule further comprises a first beta strand and a second beta strand flanking the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829), e.g., wherein the first beta strand comprises the tyrosine (Y) residue of the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829) and/or wherein the second beta strand comprises the second asparagine (N) residue (from N to C) of the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829).

3016. The synthetic anellosome of any of the preceding embodiments, wherein the ORF1 molecule comprises, 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.

3017. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element is capable of being amplified by rolling circle replication in a host cell, e.g., to produce at least 8 copies.

3018. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element is single-stranded.

3019. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element is circular.

3020. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element is DNA.

3021. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element is a negative strand DNA.

3022. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element integrates at a frequency of less than 10%, 8%, 6%, 4%, 3%, 2%, 1%, 0.5%, 0.2%, 0.1% of the anellosomes that enters the cell, e.g., wherein the synthetic anellosome is non-integrating.

3023. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element comprises a sequence of the Consensus 5' UTR nucleic acid sequence shown in Table 16-1.

3024. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element comprises a sequence of the Consensus GC-rich region shown in Table 16-2.

3025. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element comprises a sequence of at least 100 nucleotides in length, which consists of G or C at at least 70% (e.g., about 70- 100%, 75-95%, 80-95%, 85-95%, or 85-90%) of the positions.

3026. The synthetic anellosome of any of the preceding embodiments, wherein the genetic element comprises the nucleic acid sequence of SEQ ID NO: 120.

3027. The synthetic anellosome of any of the preceding embodiments, wherein the promoter element is exogenous to wild-type Anellovirus.

3028. The synthetic anellosome of any of the preceding embodiments, wherein the promoter element is endogenous to wild-type Anellovirus.

3029. The synthetic anellosome of any of the preceding embodiments, wherein the exogenous effector comprises a peptide, a synthetic or analog peptide from a naturally-bioactive peptide, an agonist or antagonist peptide, a competitive inhibitor for an enzyme, a ligand, an antibody, or a receptor.

3030. The synthetic anellosome of any of the preceding embodiments, wherein the nucleic acid sequence encoding the exogenous effector is about 20-200, 30-180, 40-160, 50-140, 60-120, 200-2000, 200-500, 500-1000, 1000-1500, or 1500-2000 nucleotides in length.

3031. The synthetic anellosome of any of the preceding embodiments, wherein 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.

3032. The synthetic anellosome of any of the preceding embodiments, wherein the synthetic anellosome is capable of infecting human cells, e.g., neurons (e.g., motor neurons), myocytes, retinal cell, liver cells, blood cells, skin cells, nerve cells, adipose cells, or endothelial cells, e.g., in vitro.

3033. The synthetic anellosome of any of the preceding embodiments, which is substantially non-immunogenic, e.g., does not induce a detectable and/or unwanted immune response, e.g., as detected according to the method described in Example 4.

3034. The synthetic anellosome of embodiment 3033, wherein the substantially non-immunogenic anellosome has an efficacy in a subject that is a least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of the efficacy in a reference subject lacking an immune response.

3035. The synthetic anellosome of any of the preceding embodiments, wherein a population of at least 1000 of the anellosomes is capable of delivering at least about 100 copies (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 copies) of the genetic element into one or more human cells.

3036. A pharmaceutical composition comprising the synthetic anellosome of any of the preceding embodiments, and a pharmaceutically acceptable carrier or excipient.

3037. The pharmaceutical composition of embodiment 3036, which comprises at least 10^3, 10^4, 105, 10^6, 10^7, 10^8, or 10^9 synthetic anellosomes.

3038. The pharmaceutical composition of embodiment 3036 or 3037, wherein the pharmaceutical composition has a predetermined ratio of particles:infectious units (e.g., <300:1, < 200:1, <100:1, or <50:1).

3039. A reaction mixture comprising:

(i) a first nucleic acid (e.g., a double-stranded or single-stranded circular DNA) comprising the sequence of the genetic element of the synthetic anellosome of any of the preceding embodiments, and

(ii) a second nucleic acid sequence 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 listed in Table 16, or an amino acid sequence having at least 85% sequence identity thereto.

3040. The reaction mixture of embodiment 3039, wherein the first nucleic acid and second nucleic acid are in the same nucleic acid molecule.

3041. The reaction mixture of embodiment 3039, wherein the first nucleic acid and second nucleic acid are different nucleic acid molecules.

3042. The reaction mixture of embodiment 3039, wherein the first nucleic acid and second nucleic acid are different nucleic acid molecules and wherein the second nucleic acid is provided as double-stranded circular DNA.

3043. The reaction mixture of embodiment 3039, wherein the first nucleic acid and second nucleic acid are different nucleic acid molecules and wherein the first and the second nucleic acid are provided as double-stranded circular DNA.

3044. The reaction mixture of embodiment 3041, wherein the second nucleic acid sequence is comprised by a helper cell or helper virus.

3045. A method of making a synthetic anellosome, the method comprising:

a) providing a host cell comprising:

(i) a first nucleic acid molecule comprising the nucleic acid sequence of a genetic element of a synthetic anellosome of any of the preceding embodiments, and

(ii) 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 listed in any of Table 16, or an amino acid sequence having at least 85% sequence identity thereto; and

b) incubating the host cell under conditions suitable to make a synthetic anellosome;

thereby making the synthetic anellosome.

3046. The method of embodiment 3045, further comprising, prior to step (a), introducing the first nucleic acid molecule and/or the second nucleic acid molecule into the cell.

3047. The method of embodiment 3046, wherein the second nucleic acid molecule is introduced into the host cell prior to, concurrently with, or after the first nucleic acid molecule.

3048. The method of any of embodiments 3045 or 3046, wherein the second nucleic acid molecule is integrated into the genome of the host cell.

3049. The method of any of embodiments 3045-3048, wherein the second nucleic acid molecule is a helper (e.g., a helper plasmid or the genome of a helper virus).

3050. The method of any of embodiments 3045-3048, wherein second nucleic acid molecule encodes an ORF2 molecule comprising 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.

3051. A method of manufacturing a synthetic anellosome preparation, the method comprising:

a) providing a plurality of synthetic anellosomes according to embodiments 3001-3035, a pharmaceutical composition of any of embodiments 3036-3038, or a reaction mixture of any of embodiments 3039-3044;

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); and

c) formulating the plurality of synthetic anellosomes, 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.

3052. A host cell comprising:

(i) a first nucleic acid molecule comprising the nucleic acid sequence of a genetic element of a synthetic anellosome of any of the preceding embodiments, 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 as listed in any of Table 16, or an amino acid sequence having at least 85% sequence identity thereto.

3053. A method of delivering an exogenous effector (e.g., a therapeutic exogenous effector) to a mammalian cell, comprising:

(a) providing a synthetic anellosome of any of the preceding embodiments; and

(b) contacting a mammalian cell with the synthetic anellosome;

wherein the synthetic anellosome is capable of delivering the genetic element into the mammalian cell; and

optionally wherein the synthetic anellosome is produced by introducing the genetic element into a host cell, under conditions suitable for enclosing the genetic element within the proteinaceous exterior in the host cell;

thereby delivering the therapeutic exogenous effector to the mammalian cell.

3054. Use of a synthetic anellosome of any of the embodiments 3001-3035 or the pharmaceutical composition of any of embodiments 3036-3038 for delivering the genetic element to a host cell.

3055. Use of a synthetic anellosome of any of the embodiments 3001-3035 or the pharmaceutical composition of any of embodiments 3036-3038 for treating a disease or disorder in a subject.

3056. The use of embodiment 3055, wherein the disease or disorder is a neuromuscular disease (e.g., spinal muscular atrophy), a muscular dystrophy (e.g., Duchenne muscular dystrophy or Becker muscular dystrophy), an enzymatic insufficiency (e.g., Crigler–Najjar syndrome), or albinism (e.g., ocular albinism), or C1 deficiency.

3057. A synthetic anellosome of any of embodiments 3001-3035 or the pharmaceutical composition of any of embodiments 3036-3038, for use in treating a disease or disorder in a subject.

3058. A method of treating a disease or disorder in a subject, the method comprising administering a synthetic anellosome of any of embodiments 3001-3035 or the pharmaceutical composition of any of embodiments 3036-3038 to the subject, wherein the disease or disorder is a neuromuscular disease (e.g., spinal muscular atrophy), a muscular dystrophy (e.g., Duchenne muscular dystrophy or Becker muscular dystrophy), an enzymatic insufficiency (e.g., Crigler–Najjar syndrome), or albinism (e.g., ocular albinism), or C1 deficiency.

3059. Use of the synthetic anellosome of any of embodiments 3001-3035 or the pharmaceutical composition of any of embodiments 3036-3038, in the manufacture of a medicament for treating a disease or disorder in a subject, optionally wherein the disease or disorder is a neuromuscular disease (e.g., spinal muscular atrophy), a muscular dystrophy (e.g., Duchenne muscular dystrophy or Becker muscular dystrophy), an enzymatic insufficiency (e.g., Crigler–Najjar syndrome), or albinism (e.g., ocular albinism), or C1 deficiency.

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

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.

Figure 1A is an illustration showing percent sequence similarity of amino acid regions of capsid protein sequences.

Figure 1B is an illustration showing percent sequence similarity of capsid protein sequences.

Figure 2 is an illustration showing one embodiment of an anellosome.

Figure 3 depicts a schematic of a kanamycin vector encoding the LY1 strain of TTMiniV ("Anellosome 1").

Figure 4 depicts a schematic of a kanamycin vector encoding the LY2 strain of TTMiniV ("Anellosome 2").

Figure 5 depicts transfection efficiency of synthetic anellosomes in 293T and A549 cells.

Figures 6A and 6B depict quantitative PCR results that illustrate successful infection of 293T cells by synthetic anellosomes.

Figures 7A and 7B depict quantitative PCR results that illustrate successful infection of A549 cells by synthetic anellosomes.

Figures 8A and 8B depict quantitative PCR results that illustrate successful infection of Raji cells by synthetic anellosomes.

Figures 9A and 9B depict quantitative PCR results that illustrate successful infection of Jurkat cells by synthetic anellosomes.

Figures 10A and 10B depict quantitative PCR results that illustrate successful infection of Chang cells by synthetic anellosomes.

Figures 11A-11B are a series of graphs showing luciferase expression from cells transfected or infected with TTMV-LY2Δ574-1371,Δ1432-2210,2610::nLuc. Luminescence was observed in infected cells, indicating successful replication and packaging.

Figure 11C is a diagram depicting the phylogenetic tree of *Alphatorquevirus* (Torque Teno Virus; TTV), with clades highlighted. At least 100 *Anellovirus* strains are represented. Exemplary sequences from several clades is provided herein, e.g., in Tables A1-A12, B1-B5, C1-C5, and 1-18.

Figure 12 is a schematic showing an exemplary workflow for production of anellosomes (e.g., replication-competent or replication-deficient anellosomes as described herein).

Figure 13 is a graph showing primer specificity for primer sets designed for quantification of TTV and TTMV genomic equivalents. Quantitative PCR based on SYBR green chemistry shows one

distinct peak for each of the amplification products using TTMV or TTV specific primer sets, as indicated, on plasmids encoding the respective genomes.

Figure 14 is a series of graphs showing PCR efficiencies in the quantification of TTV genome equivalents by qPCR. Increasing concentrations of primers and a fixed concentration of hydrolysis probe (250nM) were used with two different commercial qPCR master mixes. Efficiencies of 90-110% resulted in minimal error propagation during quantification.

Figure 15 is a graph showing an exemplary amplification plot for linear amplification of TTMV (Target 1) or TTV (Target 2) over a 7 log10 of genome equivalent concentrations. Genome equivalents were quantified over 7 10-fold dilutions with high PCR efficiencies and linearity (R^2 TTMV: 0.996; R^2 TTV: 0.997).

Figures 16A-16B are a series of graphs showing quantification of TTMV genome equivalents in an anellosome stock. (A) Amplification plot of two stocks, each diluted 1:10 and run in duplicate. (B) The same two samples as shown in panel A, here shown in the context of the linear range. Shown are the upper and lower limits in the two representative samples. PCR Efficiency: 99.58%, R^2: 0988.

Figure 17 is a graph showing fold change in miR-625 expression in HEK293T cells transfected with the indicated plasmid.

Figure 18 is a diagram showing pairwise identity for alignments of representative sequences from each *Alphatorquevirus* clade. DNA sequences for TTV-CT30F, TTV-P13-1, TTV-tth8, TTV-HD20a, TTV-16, TTV-TJN02, and TTV-HD16d were aligned. Pairwise percent identity across a 50-bp sliding window is shown along the length of the alignment. Brackets above indicate non-coding and coding regions with pairwise identities are indicated. Brackets below indicate regions of high or low sequence conservation.

Figure 19 is a diagram showing pairwise identity for amino acid alignments for putative proteins across the seven *Alphatorquevirus* clades. Amino acid sequences for putative proteins from TTV-CT30F, TTV-P13-1, TTV-tth8, TTV-HD20a, TTV-16, TTV-TJN02, and TTV-HD16d were aligned. Pairwise percent identity across a 15-aa sliding window is shown along the length of each alignment. Pairwise identity for both open reading frame DNA sequence and protein amino acid sequence is indicated. (*) Putative ORF2t/3 amino acid sequences were aligned for TTV-CT30F, TTV-tth8, TTV-16, and TTV-TJN02.

Figure 20 is a diagram showing that a domain within the 5' UTR is highly conserved across the seven *Alphatorquevirus* clades (SEQ ID NOS 810-817, respectively, in order of appearance). The 71-bp 5'UTR conserved domain sequences for each representative *Alphatorquevirus* were aligned. The sequence has 95.2% pairwise identity between the seven clades.

Figure 21 is a diagram showing an alignment of the GC-rich domains from the seven *Alphatorquevirus* clades. Each Anellovirus has a region downstream of the ORFs with greater than 70% GC content. Shown is an alignment of the GC-rich regions from TTV-CT30F, TTV-P13-1, TTV-tth8, TTV-HD20a, TTV-16, TTV-TJN02, and TTV-HD16d. The regions vary in length, but where they do align they have 75.4% pairwise identity.

Figure 22 is a diagram showing infection of Raji B cells with anellosomes encoding a miRNA targeting n-myc interacting protein (NMI). Shown is quantification of genome equivalents of anellosomes detected after infection of Raji B cells (arrow) or control cells with NMI miRNA-encoding anellosomes.

Figure 23 is a diagram showing infection of Raji B cells with anellosomes encoding a miRNA targeting n-myc interacting protein (NMI). The Western blot shows that anellosomes encoding the miRNA against NMI reduced NMI protein expression in Raji B cells, whereas Raji B cells infected with anellosomes lacking the miRNA showed comparable NMI protein expression to controls.

Figure 24 is a series of graphs showing quantification of anellosome particles generated in host cells after infection with an anellosome comprising an endogenous miRNA-encoding sequence and a corresponding anellosome in which the endogenous miRNA-encoding sequence was deleted.

Figures 25A-25C are a series of diagrams showing intracellular localization of ORFs from TTMV-LY2 fused to nano-luciferase. (A) In Vero cells, ORF2 (top row) appeared to localize to the cytoplasm while ORF1/1 (bottom row) appeared to localize to the nucleus. (B) In HEK293 cells, ORF2 (top row) appeared to localize to the cytoplasm while ORF1/1 (bottom row) appeared to localize to the nucleus. (C) Localization patterns for ORF1/2 and ORF2/2 in cells.

Figure 26 is a series of diagrams showing sequential deletion controls in the 3' non-coding region (NCR) of TTV-tth8. The top row shows the structure of the wild-type TTV-tth8 *Anellovirus*. The second row shows TTV-tth8 with a deletion of 36 nucleotides in the GC-rich region of the 3' NCR (Δ36nt (GC)). The third row shows TTV-tth8 with the 36 nucleotide deletion and an additional deletion of the miRNA sequence, resulting in a total deletion of 78 nucleotides (Δ36nt (GC) ΔmiR). The fourth row shows TTV-tth8 with a deletion of 171 nucleotides from the 3' NCR, which includes both the 36 nucleotide deletion region and the miRNA sequence (Δ3' NCR).

Figures 27A-27D are a series of diagrams showing that sequential deletions in the 3' NCR of TTV-tth8 have significant effects on *Anellovirus* ORF transcript levels. Shown are expression of ORF1 and ORF2 at day 2 (A), ORF1/1 and ORF2/2 at day 2 (B), ORF1/2 and ORF2/3 at day 2 (C), and ORF2t3 at day 2 (D).

Figures 28A-28B are a series of diagrams showing constructs used to produce anellosomes expressing nano-luciferase (A) and a series of anellosome/plasmid combinations used to transfect cells (B)

Figures 29A-29C are a series of diagrams showing nano-luciferase expression in mice injected with anellosomes. (A) Nano-luciferase expression in mice at days 0-9 after injection. (B) Nano-luciferase expression in mice injected with various anellosome/plasmid construct combinations, as indicated. (C) Quantification of nano-luciferase luminescence detected in mice after injection. Group A received a TTMV-LY2 vector ± nano-luciferase. Group B received a nano-luciferase protein and TTMV-LY2 ORFs.

Figure 29D is a schematic of the genomic organization of representative anellos from seven different *Alphatorquevirus* clades. Sequences for TTV-CT30F, TTV-P13-1, TTV-tth8, TTV-HD20a, TTV-16, TTV-TJN02, and TTV-HD16d were aligned, with key regions annotated. Putative open reading frames (ORFs) are represented in light gray, TATA boxes are represented in dark gray, and key putative regulatory regions are represented in medium gray, including the initiator element, the 5'UTR conserved domain, and the GC-rich region (e.g., as indicated).

Figure 30 is a schematic showing an exemplary workflow for determining the endogenous target of Anellovirus pre-miRNAs.

Figures 31A-31B are a series of diagrams showing that a tandem Anellovirus plasmid can increase anellovirus or anellosome production. (A) Plasmid map for an exemplary tandem Anellovirus plasmid. (B) Transfection of HEK293T cells with a tandem Anellovirus plasmid resulted in production of four times the number of viral genomes compared to single-copy harboring plasmids.

Figure 31C is a gel electrophoresis image showing circularization of TTMV-LY2 plasmids pVL46-063 and pVL46-240.

Figure 31D is a chromatogram showing copy numbers for linear and circular TTMV-LY2 constructs, as determined by size exclusion chromatography (SEC).

Figure 32 is a diagram showing an alignment of 36-nucleotide GC-rich regions from nine Anellovirus genome sequences, and a consensus sequence based thereon (SEQ ID NOS 818-827, respectively, in order of appearance).

Figure 33 is a series of diagrams showing ORF1 structures from Anellovirus strains LY2 and CBD203. Putative domains are labeled: arginine-rich region (arg-rich), core region comprising a jelly-roll domain, hypervariable region (HVR), N22 region, and C-terminal domain (CTD), as indicated.

Figure 34 is a diagram showing an ORF1 structure from Betatorquevirus strain CBS203. Residues showing high similarity among a set of 110 betatorqueviruses are indicated. Indicated are

residues of 60-79.9% similarity, residues of 80-99.9% similarity, and residues of 100% similarity among all strains evaluated.

Figure 35 is a diagram showing the consensus sequence (SEQ ID NO: 828) from alignment of 258 sequences of Alphatorqueviruses with residues with high similarity scores highlighted dark gray (100%), medium gray (80-99.9%), light gray (60-80%). Putative domains are indicated in boxes. Percent identity is also indicated by the box graph below the consensus sequence, with medium-gray boxes indicating 100% identity, light gray boxes indicating 30-99% identity, and dark gray boxes indicating below 30% identity.

Figure 36 is a schematic showing the domains of an Anellovirus ORF1 molecule and the hypervariable region to be replaced with a hypervariable domain from a different Anellovirus.

Figure 37 is a schematic showing the domains of ORF1 and the hypervariable region that will be replaced with a protein or peptide of interest (POI) from a non-anellovirus source.

Figure 38 is a series of diagrams showing the design of an exemplary anellosome genetic element based on an Anellovirus genome. The protein-coding region was deleted from the anellovirus genome (left), leaving the anelloviral non-coding region (NCR), including the viral promoter, 5'UTR conserved domain (5CD), and GC-rich region. Payload DNA was inserted into the non-coding region at the protein-coding locus (right). The resulting anellosome harbored the payload DNA (including open reading frames, genes, non-coding RNAs, etc.) and the essential anellovirus cis replication and packaging elements, but lacked the essential protein elements for replication and packaging.

Figure 39 is a bar graph showing that anellosomes comprising a genetic element encoding an exogenous human immunoadhesin successfully transduced the human lung-derived cell line EKVX.

Figure 40 is a graph showing that anellosomes based on tth8 or LY2, engineered to contain a sequence encoding human erythropoietin (hEpo), could deliver a functional transgene to mammalian cells.

Figures 41A and 41B are a series of graphs showing that engineered anellosomes administered to mice were detectable seven days after intravenous injection.

Figure 42 is a graph showing that hGH mRNA was detected in the cellular fraction of whole blood seven days after intravenous administration of an engineered anellosome encoding hGH.

Figures 43A-43D are a series of diagrams illustrating a highly conserved motif in Anellovirus ORF2. Figure 43 discloses SEQ ID NO: 949.

Figures 44A and 44B are a series of diagrams showing evidence of full-length ORF1 mRNA expression in human tissues.

Figure 45 is a graph showing the ability of an in vitro circularized (IVC) TTV-tth8 genome (IVC TTV-tth8) compared to a TTV-tth8 genome in a plasmid to yield TTV-tth8 genome copies at the expected density in HEK293T cells.

Figure 46 is a series of graphs showing the ability of an in vitro circularized (IVC) LY2 genome (WT LY2 IVC) and a wild-type LY2 genome in plasmid (WT LY2 Plasmid) to yield LY2 genome copies at the expected density in Jurkat cells.

Figure 47 is a diagram showing an alignment of secondary structure of the jelly roll domain of Anellovirus ORF1 proteins from Alphatorquevirus, Betatorquevirus, and Gammatorquevirus (SEQ ID NOs: 950-975). These secondary structural elements are highly conserved.

Figure 48 is a diagram showing the conserved sequence and secondary structure of the ORF1 motif located in the N22 domain (SEQ ID NOS 976-1000 and 851, respectively, in order of appearance). The conserved YNPXXDXGXXN (SEQ ID NO: 829) Motif of human TTV ORF1 has a conserved secondary structure. In particular, the tyrosine in the motif breaks a beta strand, and a second beta strand starts on the terminal asparagine of the motif.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS

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.

As used herein, the term "anellosome" refers to a vehicle comprising a genetic element, e.g., an episome, e.g., circular DNA, enclosed in a proteinaceous exterior. A "synthetic anellosome," as used herein, generally refers to an anellosome 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 anellosome 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, e.g., prior

to entry into a host cell. In some embodiments, the anellosome is purified, e.g., it is separated from its original source and/or substantially free (>50%, >60%, >70%, >80%, >90%) of other components.

As used herein, the term "anellovector" refers to a vector that comprises sufficient nucleic acid sequence derived from or highly similar to (e.g., at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to) an *Anellovirus* genome sequence or a contiguous portion thereof to allow packaging into a proteinaceous exterior (e.g., a capsid), and further comprises a heterologous sequence. In some embodiments, the anellovector is a viral vector or a naked nucleic acid. In some embodiments, the anellovector comprises at least about 50, 60, 70, 71, 72, 73, 74, 75, 80, 90, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2500, 3000, or 3500 consecutive nucleotides of a native *Anellovirus* sequence or a sequence highly similar (e.g., at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical) thereto. In some embodiments, the anellovector further comprises one or more of an *Anellovirus* ORF1, ORF2, or ORF3. In some embodiments, the heterologous sequence comprises a multiple cloning site, comprises a heterologous promoter, comprises a coding region for a therapeutic protein, or encodes a therapeutic nucleic acid. In some embodiments, the capsid is a wild-type *Anellovirus* capsid. In embodiments, an anellovector comprises a genetic element described herein, e.g., comprises a genetic element comprising a promoter, a sequence encoding a therapeutic effector, and a capsid binding sequence.

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 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 anellosome are based.

As used herein, the term "genetic element" refers to a nucleic acid sequence, generally in an anellosome. 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 anellosome 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, 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, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10), 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 anellosome 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 (e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17). 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), e.g., an ORF1 protein having the 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, 18, 20-37, or D1-D10, or as encoded by the ORF1 gene as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17.

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, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10), 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), e.g., an ORF2 protein having the 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, 18, 20-37, or D1-D10, or as encoded by the ORF2 gene as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17.

As used herein, the term "proteinaceous exterior" refers to an exterior component that is predominantly (e.g., >50%, >60%, > 70%, >80%, > 90%) 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 "replication protein" refers to a protein, e.g., a viral protein, that is utilized during infection, viral genome replication/expression, viral protein synthesis, and/or assembly of the viral components.

As used herein, a "substantially non-pathogenic" organism, particle, or component, refers to an organism, particle (e.g., a virus or an anellosome, e.g., as described herein), or component thereof that does not cause or induce a detectable disease or pathogenic condition, e.g., in a host organism, e.g., a mammal, e.g., a human. In some embodiments, administration of an anellosome 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 a detectable 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 anellosome, 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.

As used herein, a "substantially non-immunogenic" organism, particle, or component, refers to an organism, particle (e.g., a virus or anellosome, 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 detectable immune response. In embodiments, the substantially non-immunogenic anellosome does not produce a detectable immune response against a protein comprising an amino acid sequence or encoded by a nucleic acid sequence shown in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17. In embodiments, an immune response (e.g., an undesired or untargeted immune response) is detected by assaying antibody presence or level (e.g., presence or level of an anti-anellosome antibody, e.g., presence or level of an antibody against an anellosome 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 against an *Anellovirus* or an anellosome 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, "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).

As used herein, the term "virome" refers to viruses in a particular environment, e.g., a part of a body, e.g., in an organism, e.g. in a cell, e.g. in a tissue.

This invention relates generally to anellosomes, e.g., synthetic anellosomes, and uses thereof. The present disclosure provides anellosomes, compositions comprising anellosomes, and methods of making or using anellosomes. Anellosomes are generally useful as delivery vehicles, e.g., for delivering a therapeutic agent to a eukaryotic cell. Generally, an anellosome 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 anellosome 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). Anellosomes 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. Anellosomes
- A. Anelloviruses
- B. ORF1 molecules
- C. ORF2 molecules
- D. Genetic elements
- E. Protein binding sequences
- F. 5' UTR Regions
- G. GC-rich regions
- H. Effectors
- I. Proteinaceous exterior

II. Vectors

III. Compositions

IV. Host cells

V. Methods of use

VI. Methods of production

VII. Administration/ Delivery

I. Anellosomes

In some aspects, the invention described herein comprises compositions and methods of using and making an anellosome, anellosome preparations, and therapeutic compositions. In some embodiments, the anellosome has a sequence, structure, and/or function that is based on an *Anellovirus* (e.g., an *Anellovirus* as described herein, e.g., an *Anellovirus* comprising a nucleic acid or polypeptide comprising a sequence as shown in any of Tables A1-A12, B1-B5, C1-C5, 1-18, 20-37, or D1-D10), 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 anellosome 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 anellosome. In some embodiments, an *Anellovirus*-based anellosome 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 anellosome 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 anellosome 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 anellosome is capable of replicating in a eukaryotic cell, e.g., a mammalian cell, e.g., a human cell. In some embodiments, the anellosome is substantially non-pathogenic and/or substantially non-integrating in the mammalian (e.g., human) cell. In some embodiments, the anellosome is substantially non-immunogenic in a mammal, e.g., a human. In some embodiments, the anellosome is replication-deficient. In some embodiments, the anellosome is replication-competent.

In some embodiments the anellosome 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 an aspect, the invention includes an anellosome 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 anellosome is capable of delivering the genetic element into a eukaryotic cell.

In some embodiments of the anellosome 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 anellosomes 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 anellosomes, 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 anellosome 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 listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17); and (ii) a proteinaceous exterior; wherein the genetic element is enclosed within the proteinaceous exterior; and wherein the anellosome is capable of delivering the genetic element into a eukaryotic cell.

In one aspect, the invention includes an anellosome 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 anellosome 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 anellosome 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 anellosomes, compositions comprising anellosomes, methods using such anellosomes, 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 anellosomes 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 anellosomes 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 anellosome, or the genetic element comprised in the anellosome, 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 anellosome, is expressed in a cell (e.g., a human cell), e.g., once the anellosome or the genetic element has been introduced into the cell. In embodiments, introduction of the anellosome, 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 anellosome, or genetic element comprised therein, decreases level of interferon produced by the cell. In embodiments, introduction of the anellosome, or genetic element comprised therein, into a cell modulates (e.g., increases or decreases) a function of the cell. In embodiments, introduction of the anellosome, or genetic element comprised therein, into a cell modulates (e.g., increases or decreases) the viability of the cell. In embodiments, introduction of the anellosome, or genetic element comprised therein, into a cell decreases viability of a cell (e.g., a cancer cell).

In some embodiments, an anellosome (e.g., a synthetic anellosome) 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 anellosome 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 anellosome 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 anellosome, e.g., as described herein, comprises sequences or expression products derived from an *Anellovirus*. In some embodiments, an anellosome includes one or more sequences or expression products that are exogenous relative to the *Anellovirus*. In some embodiments, an anellosome includes one or more sequences or expression products that are endogenous relative to the *Anellovirus*. In some embodiments, an anellosome includes one or more sequences or expression products that are heterologous relative to one or more other sequences or expression products in the anellosome. *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-Inigo 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 anellosome 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 anellosome comprises a nucleic acid sequence selected from a sequence as shown in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the anellosome comprises a polypeptide comprising a sequence as shown in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10, 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 anellosome 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 A1-A12, B1-B5, C1-C5, or 1-18). 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 A1-A12 or 1-18). 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, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10, or an ORF1 or ORF2 amino acid sequence encoded by a nucleic acid sequence as shown in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17). 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, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10, or an ORF1 amino acid sequence encoded by a nucleic acid sequence as shown in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17).

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 nucleotide sequence of Table A1 (e.g., nucleotides 574 – 2775 of the 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 (e.g., nucleotides 574 – 699 and/or 2326 – 2775 of the 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* ORF1/2 nucleotide sequence of Table A1 (e.g., nucleotides 574 – 699 and/or 2552 – 2759 of the 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* ORF2 nucleotide sequence of Table A1 (e.g., nucleotides 335 – 703 of the 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* ORF2/2 nucleotide sequence of Table A1 (e.g., nucleotides 335 – 699 and/or 2326 – 2759 of the 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* ORF2/3 nucleotide sequence of Table A1 (e.g., nucleotides 335 – 699 and/or 2552 – 2957 of the 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* ORF2t/3 nucleotide sequence of Table A1 (e.g., nucleotides 335 – 465 and/or 2552 – 2957 of the 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* TATA box nucleotide sequence of Table A1 (e.g., nucleotides 77 – 81 of the 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* initiator element nucleotide sequence of Table A1 (e.g., nucleotides 95 – 110 of the 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* transcriptional start site nucleotide sequence of Table A1 (e.g., nucleotide 105 of the 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* 5' UTR conserved domain nucleotide sequence of Table A1 (e.g., nucleotides 165 – 235 of the 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* three open-reading frame region nucleotide sequence of Table A1 (e.g., nucleotides 2535 – 2746 of the 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* poly(A) signal nucleotide sequence of Table A1 (e.g., nucleotides 2953 – 2958 of the 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* GC-rich nucleotide sequence of Table A1 (e.g., nucleotides 3620 – 3648 of the 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 nucleotide sequence of Table A3 (e.g., nucleotides 599 – 2887 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 599 – 724 and/or 2414 – 2887 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 599 – 724 and/or 2643 – 2849 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 342 – 728 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 342 – 724 and/or 2414 – 2849 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 342 – 724 and/or 2643 – 3057 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 87 – 91 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 105 – 120 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotide 115 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 175 – 245 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 2626 – 2846 of the nucleic acid sequence of Table A3). 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 A3 (e.g., nucleotides 3052 – 3058 of the nucleic acid sequence of Table A3).

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 nucleotide sequence of Table A5 (e.g., nucleotides 556 – 2904 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 556 – 687 and/or 2422 – 2904 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 556 – 687 and/or 2564 – 2878 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 305 – 691 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 305 – 687 and/or 2422 – 2878 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 305 – 687 and/or 2564 – 3317 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 305 – 360 and/or 2564 – 3317 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 50 – 55 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 68 – 83 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotide 78 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 138 – 208 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 2626 – 2846 of the nucleic acid sequence of Table A5). 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 A5 (e.g., nucleotides 3316 – 3319 of the nucleic acid sequence of Table A5).

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 nucleotide sequence of Table A7 (e.g., nucleotides 589 – 2889 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 589 – 711 and/or 2362 – 2889 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 589 – 711 and/or 2555 – 2863 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 353 – 715 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 353 – 711 and/or 2362 – 2863 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 353 – 711 and/or 2555 – 3065 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 353 – 432 and/or 2555 – 3065 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 86 – 90 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 104 – 119 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotide 114 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 174 – 244 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 2555 – 2863 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 3062 – 3066 of the nucleic acid sequence of Table A7). 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 A7 (e.g., nucleotides 3720 – 3742 of the nucleic acid sequence of Table A7).

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 nucleotide sequence of Table A9 (e.g., nucleotides 511 – 2793 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 511 – 711 and/or 2326 – 2793 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 511 – 711 and/or 2525 – 2767 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 272 – 637 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 272 – 633 and/or 2326 – 2767 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 272 – 633 and/or 2525 – 2984 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 272 – 633 and/or 2525 – 2984 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 12 – 17 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 30 – 45 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotide 40 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 100 – 171 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 2525 –

2767 of the nucleic acid sequence of Table A9). 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 A9 (e.g., nucleotides 2981 – 2985 of the nucleic acid sequence of Table A9).

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 nucleotide sequence of Table A11 (e.g., nucleotides 704 – 3001 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 704 – 826 and/or 2534 – 3001 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 704 – 826 and/or 2721 – 2975 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 465 – 830 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 465 – 826 and/or 2534 – 2975 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 465 – 826 and/or 2721 – 3192 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 465 – 595 and/or 2721 – 3192 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 206 – 210 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 224 – 239 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotide 234 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 294 – 364 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 2721 – 2975 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 3189 – 3193 of the nucleic acid sequence of Table A11). 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 A11 (e.g., nucleotides 3844 – 3895 of the nucleic acid sequence of Table A11).

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 nucleotide sequence of Table B1 (e.g., nucleotides 574 – 2775 of the 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 (e.g., nucleotides 574 – 699 and/or 2326 – 2775 of the 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* ORF1/2 nucleotide sequence of Table B1 (e.g., nucleotides 574 – 699 and/or 2552 – 2759 of the 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* ORF2 nucleotide sequence of Table B1 (e.g., nucleotides 335 – 703 of the 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* ORF2/2 nucleotide sequence of Table B1 (e.g., nucleotides 335 – 699 and/or 2326 – 2759 of the 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* ORF2/3 nucleotide sequence of Table B1 (e.g., nucleotides 335 – 699 and/or 2552 – 2957 of the 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* ORF2t/3 nucleotide sequence of Table B1 (e.g., nucleotides 335 – 465 and/or 2552 – 2957 of the 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* TATA box nucleotide sequence of Table B1 (e.g., nucleotides 77 – 81 of the 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* initiator element nucleotide sequence of Table B1 (e.g., nucleotides 95 – 110 of the 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* transcriptional start site nucleotide sequence of Table B1 (e.g., nucleotide 105 of the 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* 5' UTR conserved domain nucleotide sequence of Table B1 (e.g., nucleotides 165 – 235 of the 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* three open-reading frame region nucleotide sequence of Table B1 (e.g., nucleotides 2535 – 2746 of the 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* poly(A) signal nucleotide sequence of Table B1 (e.g., nucleotides 2953 – 2958 of the 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* GC-rich nucleotide sequence of Table B1 (e.g., nucleotides 3620 – 3648 of the 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 nucleotide sequence of Table B2 (e.g., nucleotides 574 – 2775 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 574 – 699 and/or 2326 – 2775 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 574 – 699 and/or 2552 – 2759 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 335 – 703 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 335 – 699 and/or 2326 – 2759 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 335 – 699 and/or 2552 – 2957 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 335 – 465 and/or 2552 – 2957 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 77 – 81 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 95 – 110 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotide 105 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 165 – 235 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 2535 – 2746 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 2953 – 2958 of the nucleic acid sequence of Table B2). 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 B2 (e.g., nucleotides 3620 – 3648 of the nucleic acid sequence of Table B2).

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 nucleotide sequence of Table B3 (e.g., nucleotides 574 – 2775 of the 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 (e.g., nucleotides 574 – 699 and/or 2326 – 2775 of the 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* ORF1/2 nucleotide sequence of Table B3 (e.g., nucleotides 574 – 699 and/or 2552 – 2759 of the 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* ORF2 nucleotide sequence of Table B3 (e.g., nucleotides 335 – 703 of the 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* ORF2/2 nucleotide sequence of Table B3 (e.g., nucleotides 335 – 699 and/or 2326 – 2759 of the 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* ORF2/3 nucleotide sequence of Table B3 (e.g., nucleotides 335 – 699 and/or 2552 – 2957 of the 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* ORF2t/3 nucleotide sequence of Table B3 (e.g., nucleotides 335 – 465 and/or 2552 – 2957 of the 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* TATA box nucleotide sequence of Table B3 (e.g., nucleotides 77 – 81 of the 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* initiator element nucleotide sequence of Table B3 (e.g., nucleotides 95 – 110 of the 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* transcriptional start site nucleotide

sequence of Table B3 (e.g., nucleotide 105 of the 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* 5' UTR conserved domain nucleotide sequence of Table B3 (e.g., nucleotides 165 – 235 of the 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* three open-reading frame region nucleotide sequence of Table B3 (e.g., nucleotides 2535 – 2746 of the 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* poly(A) signal nucleotide sequence of Table B3 (e.g., nucleotides 2953 – 2958 of the 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* GC-rich nucleotide sequence of Table B3 (e.g., nucleotides 3620 – 3648 of the 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 nucleotide sequence of Table B4 (e.g., nucleotides 574 – 2775 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 574 – 699 and/or 2326 – 2775 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 574 – 699 and/or 2552 – 2759 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 335 – 703 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 335 – 699 and/or 2326 – 2759 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 335 – 699 and/or 2552 – 2957 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 335 – 465 and/or 2552 – 2957 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 77 – 81 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 95 – 110 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotide 105 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 165 – 235 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 2535 – 2746 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 2953 – 2958 of the nucleic acid sequence of Table B4). 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 B4 (e.g., nucleotides 3620 – 3648 of the nucleic acid sequence of Table B4).

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 nucleotide sequence of Table B5 (e.g., nucleotides 574 – 2775 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 574 – 699 and/or 2326 – 2775 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides

574 – 699 and/or 2552 – 2759 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 335 – 703 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 335 – 699 and/or 2326 – 2759 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 335 – 699 and/or 2552 – 2957 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 335 – 465 and/or 2552 – 2957 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 77 – 81 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 95 – 110 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotide 105 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 165 – 235 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 2535 – 2746 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 2953 – 2958 of the nucleic acid sequence of Table B5). 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 B5 (e.g., nucleotides 3620 – 3648 of the nucleic acid sequence of Table B5).

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 nucleotide sequence of Table 1 (e.g., nucleotides 571 – 2613 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 571 – 587 and/or 2137 – 2613 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 571 – 687 and/or 2339 – 2659 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 299 – 691 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 299 – 687 and/or 2137 – 2659 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 299 – 687 and/or 2339 – 2831 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 299 – 348 and/or 2339 – 2831 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 84 – 90 of the nucleic acid sequence of Table 1). 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* Cap site nucleotide sequence of Table 1 (e.g., nucleotides 107 – 114 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotide 114 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 177 – 247 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 2325 – 2610 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 2813 – 2818 of the nucleic acid sequence of Table 1). 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 1 (e.g., nucleotides 3415 – 3570 of the nucleic acid sequence of Table 1).

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 nucleotide sequence of Table 3 (e.g., nucleotides 729 – 2972 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 729 – 908 and/or 2490 – 2972 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 729 – 908 and/or 2725 – 3039 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 412 – 912 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 412 – 908 and/or 2490 – 3039 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 412 – 908 and/or 2725 – 3208 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 112 – 119 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 128 – 148 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotide 148 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 204 – 273 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 2699 – 2969 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 3220 – 3225 of the nucleic acid sequence of Table 3). 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 3 (e.g., nucleotides 3302 – 3541 of the nucleic acid sequence of Table 3).

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 nucleotide sequence of Table 5 (e.g., nucleotides 599 – 2830 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 599 – 715 and/or 2363 – 2830 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 599 – 715 and/or 2565 – 2789 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 336 – 719 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 336 – 715 and/or 2363 – 2789 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 336 – 715 and/or 2565 – 3015 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 336 – 388 and/or 2565 – 3015 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 83 – 88 of the nucleic acid sequence of Table 5). 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* Cap site nucleotide sequence of Table 5 (e.g., nucleotides 104 – 111 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotide 111 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 170 – 240 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 2551 – 2786 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 3011 – 3016 of the nucleic acid sequence of Table 5). 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 5 (e.g., nucleotides 3632 – 3753 of the nucleic acid sequence of Table 5).

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 nucleotide sequence of Table 7 (e.g., nucleotides 586 – 2928 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 586 – 717 and/or 2446 – 2928 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 586 – 717 and/or 2675 – 2902 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 335 – 721 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 335 – 717 and/or 2446 – 2902 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 335 – 717 and/or 2675 – 3109 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 82 – 87 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 95 – 115 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotide 115 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 170 – 238 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 2640 – 2899 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 3106 – 3114 of the nucleic acid sequence of Table 7). 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 7 (e.g., nucleotides 3768 – 3878 of the nucleic acid sequence of Table 7).

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 nucleotide sequence of Table 9 (e.g., nucleotides 588 – 2873 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 588 – 722 and/or 2412 – 2873 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 588 – 722 and/or 2638 – 2847 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 331 – 726 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 331 – 722 and/or 2412 – 2847 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 331 – 722 and/or 2638 – 3058 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 331 – 380 and/or 2638 – 3058 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 82 – 86 of the nucleic acid sequence of Table 9). 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 9 (e.g.,

nucleotides 100 – 115 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotide 115 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 170 – 240 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 2699 – 2969 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 3220 – 3225 of the nucleic acid sequence of Table 9). 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 9 (e.g., nucleotides 3302 – 3541 of the nucleic acid sequence of Table 9).

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 nucleotide sequence of Table 11 (e.g., nucleotides 599 – 2839 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 599 – 727 and/or 2381 – 2839 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 599 – 727 and/or 2619 – 2813 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 357 – 731 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 357 – 727 and/or 2381 – 2813 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 357 – 727 and/or 2619 – 3021 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 357 – 406 and/or 2619 – 3021 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 89 – 90 of the nucleic acid sequence of Table 11). 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* Cap site nucleotide sequence of Table 11 (e.g., nucleotides 107 – 114 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotide 114 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 174 – 244 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 2596 – 2810 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 3017 – 3022 of the nucleic acid sequence of Table 11). 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 11 (e.g., nucleotides 3691 – 3794 of the nucleic acid sequence of Table 11).

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 nucleotide sequence of Table 13 (e.g., nucleotides 599 – 2896 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 599 – 724 and/or 2411 –

2896 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 599 – 724 and/or 2646 – 2870 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 357 – 728 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 357 – 724 and/or 2411 – 2870 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 357 – 724 and/or 2646 – 3081 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 82 – 86 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 94 – 115 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotide 115 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 170 – 240 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 2629 – 2867 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 3076 – 3086 of the nucleic acid sequence of Table 13). 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 13 (e.g., nucleotides 3759 – 3866 of the nucleic acid sequence of Table 13).

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 nucleotide sequence of Table 15 (e.g., nucleotides 612 – 2612 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 612 – 719 and/or 2274 – 2612 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 612 – 719 and/or 2449 – 2589 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 424 – 723 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 424 – 719 and/or 2274 – 2589 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 424 – 719 and/or 2449 – 2812 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 237– 243 of the nucleic acid sequence of Table 15). 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* Cap site nucleotide sequence of Table 15 (e.g., nucleotides 260 – 267 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotide 267 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 323 – 393 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 2441 – 2586 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 2808 – 2813 of the nucleic acid sequence of Table 15). 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 15 (e.g., nucleotides 2868 – 2929 of the nucleic acid sequence of Table 15).

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 nucleotide sequence of Table 17 (e.g., nucleotides 432 – 2453 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 432 – 584 and/or 1977 – 2453 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 432 – 584 and/or 2197 – 2388 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 283 – 588 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 283 – 584 and/or 1977 – 2388 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 283 – 584 and/or 2197 – 2614 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 21– 25 of the nucleic acid sequence of Table 17). 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* Cap site nucleotide sequence of Table 17 (e.g., nucleotides 42 – 49 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotide 49 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 117 – 187 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 2186 – 2385 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 2676 – 2681 of the nucleic acid sequence of Table 17). 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 17 (e.g., nucleotides 3054 – 3172 of the nucleic acid sequence of Table 17).

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A2.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A4.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A6.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A8.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A10.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table A12.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C1.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C2. In embodiments, the

nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C2.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C3. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C3. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C3. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C3. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C3. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C3. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C3.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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

amino acid sequence of Table C4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C4.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C5. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C5. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C5. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C5. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C5. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table C5. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C5.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%

sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 2. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 2.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 4. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 4.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 6. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 6.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 8. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 8.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 10. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 10.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 12. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 12.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 14. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 14. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 14. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 14. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 14. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 14.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 16. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 16. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 16. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence

encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 16. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 16. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 16.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 18. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 18. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 18. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 18. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 18. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding an amino 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 amino acid sequence of Table 18.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A2. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence

having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A2. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 574-2775 of the nucleic acid sequence of Table A1. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table A2 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A4. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A4. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 599-2887 of the nucleic acid sequence of Table A3. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an

Anellovirus ORF1 protein of Table A4 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A6. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A6. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 556-2904 of the nucleic acid sequence of Table A5. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table A6 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A8. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A8. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A8. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence

having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A8. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A8. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A8. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A8. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 589-2889 of the nucleic acid sequence of Table A7. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table A8 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A10. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A10. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the

Anellovirus ORF1 nucleic acid sequence of nucleotides 511-2793 of the nucleic acid sequence of Table A9. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table A10 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table A12. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A12. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A12. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table A12. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A12. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A12. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table A12. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 704-3001 of the nucleic acid sequence of Table A11. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table A12 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C1. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C1. In embodiments, the anellosome described herein comprises a

protein having an amino 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 amino acid sequence of Table C1. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C1. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C1. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C1. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C1. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C1. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 512-2545 of the nucleic acid sequence of Table B1. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table C1 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C2. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C2. In embodiments, the anellosome described herein comprises a protein having an

amino 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 amino acid sequence of Table C2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C2. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C2. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 501-2489 of the nucleic acid sequence of Table B2. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table C2 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C3. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C3. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C3. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C3. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C3. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C3. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C3. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C3. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide

encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 572-2758 of the nucleic acid sequence of Table B3. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table C3 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C4. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C4. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C4. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 581-2884 of the nucleic acid sequence of Table B4. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table C4 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table C5. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C5. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C5. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table C5. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C5. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C5. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table C5. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP amino acid sequence of Table C5. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 614-2911 of the nucleic acid sequence of Table B5. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table C5 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 2. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 2. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 2. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 571-2613 of the nucleic acid sequence of Table 1. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table 2 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 4. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 4. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 4. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 729-2972 of the nucleic acid sequence of Table 3. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table 4 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%

sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 6. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 6. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 6. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 599-2830 of the nucleic acid sequence of Table 5. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table 6 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 8. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 8. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 8. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 8. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid

sequence of Table 8. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 8. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 586-2928 of the nucleic acid sequence of Table 7. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table 8 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 10. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 10. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 10. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 588-2873 of the nucleic acid sequence of Table 9. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table 10 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 12. In embodiments, the

anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 12. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 12. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 12. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 12. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 12. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 12. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 599-2839 of the nucleic acid sequence of Table 11. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table 12 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 14. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 14. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 14. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 14. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 14. In embodiments, the anellosome described herein comprises a protein having an

amino 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 amino acid sequence of Table 14. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 599-2896 of the nucleic acid sequence of Table 13. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table 14 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 16. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 16. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 16. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 16. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 16. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 16. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 612-2612 of the nucleic acid sequence of Table 15. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table 16 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 amino acid sequence of Table 18. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 18. In embodiments, the anellosome described herein comprises a

protein having an amino 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 amino acid sequence of Table 18. In embodiments, the anellosome described herein comprises a protein having an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 amino acid sequence of Table 18. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 18. In embodiments, the anellosome described herein comprises a protein having an amino 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 amino acid sequence of Table 18. In some embodiments, an ORF1 molecule (e.g., comprised in the anellosome) comprises a polypeptide encoded by the Anellovirus ORF1 nucleic acid sequence of nucleotides 432-2453 of the nucleic acid sequence of Table 17. In some embodiments, the ORF1 molecule (e.g., comprised in the anellosome) comprises an Anellovirus ORF1 protein of Table 18 or a splice variant or post-translationally processed (e.g., proteolytically processed) variant thereof.

In some embodiments, the polypeptide 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 an *Anellovirus* ORF1 amino acid sequence described herein. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table 2. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table 4. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table 6. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table 8. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table 10. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table 12. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table 14. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table 16. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table 18. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table A2. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table A4. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table A6. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table A8. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table A10. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table A12. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table C1. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table C2. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table C3. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table C4. In embodiments, the polypeptide 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 the *Anellovirus* ORF1 amino acid sequence of Table C5.

In some embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid described herein. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table 1. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table 3. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table 5. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table 7. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table 9. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table 11. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table 13. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table 15. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table 17. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table A1. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table A3. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table A5. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table A7. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table A9. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table A11. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table B1. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table B2. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table B3. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table B4. In embodiments, the polypeptide 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 an ORF1 molecule encoded by an *Anellovirus* ORF1 nucleic acid as listed in Table B5.

In some embodiments, the polypeptide comprises an amino acid sequence (e.g., an ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3 sequence) as shown in any of Tables 2, 4, 6, 8, 10, 12, 14, 16, or 18, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

Table A1. Novel *Anellovirus* nucleic acid sequence (*Alphatorquevirus*)

Name	TTV-RTx1
Genus/Clade	*Alphatorquevirus*, Clade 6
Accession Number	SRR2167793

Full Sequence: 3648 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
CGTCACTAACCACGTGACTCCCACAGGCCAACCACAGTGTACGTGATTCA
CTTCCTGGGAGTGGTTTACATTATAATATAAGCAACTGCACTTCCGAATG
GCTGAGTTTTCCACGCCCGTCCGCAGCGAGAACACCACGGAGGGGAGTCC
GCGCGTCCCGTGGGCGGGTGCCGAAGGTGAGTTTACACACCGCAGTCAAG
GGGCAATTCGGGCACGGGACTGGCCGGGCTATGGGCAAGGCTCTTAAAAA
GCTATGTTTCTTGGTAGGCCGTACCGAAAGAAAAGGAAACTGCTACTGCT
ACCACTGCATTCTACACCGAAAACTAGCCGGGTTATGAGCTGGTCTAGGC
CTGTACATAATGCCACAGGCATTGAAAGAAACTGGTGGGAGTCCTGTCTT
AGATCCCACGCAAGTTCTTGTGGCTGCGGTAATTTTGTTAATCATATTAA
TGTACTGGCTAATCGGTATGGCTTTGCTGGTTCCACGGAGACGCCGGGTA
ATCCTCGGCCGAGGCCCCCGGTACTGAGCTCCACCACCAGCACTCCTACC
GATCAATCCAGACCAGCTCTACCATGGCATGGGGATACTGGTGGAGAAGG
CGCTTCTGGAGACCCCGCAGGAGATGGAGAACGTGGCGCCGCAGAAGGAG
ACTACGGCCCAGAAGATCTAGACGCACTTTTCGACGCACTCGACGAAGAG
TAAGGAGGCGACGGTGGGGGAGGCGTGCACGCAGGCGGGGATGGCGACGC
AGGACTTATATTAGAGCCAGGCGACGCAGGAGACGAAAAAGACTTGTACT
GACTCAGTGGCATCCCGCAGTTAGAAGAAAATGTAAAATTACAGGCTACA
TGCCTATAGTATACTGTGGACATGGCAGAGCTAGTTTTAACTATGCCTGG
CACTCTGATGACTGTATAAAACAACCACTACCCTTTGGAGGCTCACTATC
TACAGTGTCCTTCAACCTAAAAGTACTATTTGACGAAAACCAAAGAGGAC
TAAACAAATGGAGCTACCCAAATGACCAACTAGACCTCGCCAGATACAAA
GGCTGTAGACTAACATTTTACAGAAAAAAAAACACAGACTACATAGCTCA
ATATGACATATCAGAACCTTATCAACTAGACAAATATAGCTGTGCAAACT
ATCACCCCTCAAAAATGATGTTTGCAAAAAACAAAATTTTAATTCCTAGC
TATGATACAAAACCTAGAGGCAGACAAAGAGTTAGAGTTAGAATAGGGCC
CCCTAAACTATTTACAGACAAGTGGTACAGTCAATCAGACTTATGCAAGG
TAAACCTTGTGTCACTTGCGGTTTCTGCGGCTTCCTTTCTCCACCCATTC
GGCTCACCACAAACTGCCAACTTTTGTGCAACCTTCCAGGTGCTGCAACC
GTTCTACTACCAGGCTATAGGCATTAGTTCTACAAAACACTCAGAAGTTA
TAGACATTTTATATAAGAAAAATACATACTGGCAAAGCAACATTACCTCT
TGGTTTTTAACTAATGTTAAAAACCCAAAAAATATGTCCACAAAAATGTT
TGAGGACATTAATGTTAAATCAAACAAAGACAGTAATTATGACTGGTTTC
CATTTACCCCATACACTACAGAAAACTATTCAAAAATTCAAAATGCAGCT
CAAGAATACTGGAAATATTTAACTAGTGACCACCCACAAGCTACTAATAG
CAATGAAGGCCTAGTACAACCATGGACTAATGCCACTATAAAACAATATG
AATACCACCTCGGTATGTTTAGTCCTATATTTATAGGACCTACCAGAGCT
AAAACTAAATTTAAAACAGCATACTTTGACTGCACTTATAACCCACTACT
AGACAAAGGAATGGGAAACAGAATATGGTATCAATACGCAACCAAAGCTG
ACACACAAATATCAAAAACAGGGTGCTACTGCATGTTAGAAGACATTCCA
ATATATGCAGCATTTTATGGATACGTAGACTTTATAGAAATGGAAATAGG
TAAAGGACAAGACATTAAAGAGAACGGACTTATTTGCTGCATATGTAGAT
ACACAGACCCCCCAATGTACAATGAACAACATCCAGACATGGGATTTGTA
TTTTATAACACTAACTTTGGAAATGGAAAATGGATAGATGGACGGGGCGA
CATACCTACTTACTGGATGCAAAGATGGAGACCTGTTGTATTATTTCAAA
CTGATGTTATTAGAGACTTAGTAGAAACTGGACCTTTTAGTTACAAAGAT
GACCTAGCAAATACCTCACTGACTATGAAATATGAATTCTATTTTACCTG
GGGCGGAAACCAGGCGTACCACCAGACAATCAAAAACCCTTGTAAAGACG
AAGGTACCGGACCCCATAGACAGCCTAGAGACGTACAAGTTACGGACCCG
ACAACCGTGGGACCTGAATATGTGTTCCACGCGTGGGACTGGAGACGGGG
CTTCCTTAGCGAGCGAGCTCTCAGACGCATGTTCGAAAAACCTCTCAACT
ATGATGAGTATTCTAAAAAACCAAAAAGACCTAGAATATTTCCTCCAACA
```

GAAACAGAGTCCCGAAACCAAGAGCTCGAAGAAAGCTCGCTTTCAGAGGA
AGAAAAGTCGCTACTCTCCACAGAAGAGATCCAGAAAGAGGAGATACAGC
GACAGTTCAAGCGACAGCTCAAGCGACAGCTGCGCCTCGGGCAGCAGCTC
AAACTCCTCCAACAACAACTCCTCAAGACGCAAGCGGGCCTGCACCTAAA
CCCCCTTTCATATTTCCCGCAATAAATAAAGTGTACCTGTTCCCAGACAG
AGCTCCAAAACCTAAACCCACCTCTGGAGACTGGGAAACAGAGTATGCAG
CTTGCAGTGCCTTTGACAGACCCGCTAGAACCAACCTTAGCTCACCCCCT
TACTACCCAGGAGTACCTACTCCCTGGCAAGTAAAATTCAGCCTTAAATT
TCAATAAAGTGCATTTTTACTACAGCTGGGCCGTGGGAGTTTCACTTGTC
GGTGTCTACCTCTTAAGGTCACTAAGCACTCCGAGCGCAGCGAGGAGTGC
GACCCTTAACCCTGGGTCAACGCCTTCGGAGCCGCGCGCTACGCCTTCGG
CTGCGCGCGGCACCTCAGACCCCCGCTCGTGCTGACGCGCTTGCGCGCGT
CAGACCACTTCGGGCTCGCGGGGGTCGGGAACTTTGCTAACAGACTCCGA
GGTGCCATTGGACACAGAGTGGGCGTTCAGCAACGAAAGTGAGTGGGGCC
AGACTTCGCCATAAGGCCTTTATCTTCTTGCCATTTGTCAGTATAAGGGG
TTGCCATAGGCTTCGGCCTCAATTTTAGGCCTTCCGGACTACCAAAATGG
CCGATTTAGTGACGTCACGGCGGCCATTTTAAGTAAGGCGGAAGTAACTC
CACTATTTACAAAATGGCGGCGGAGCACTTCCGGCTTGCCCAAAATGGCG
GCAAAAAACATCCGGGTCAAAGGTCGTTACCACGTCACAAGTCACGTGGG
AGGGTGGTGCTGTAAACCCGGAAGCAATCCTCTCACGTGGCTAGTCACGT
GACTAACACGTCACACCCGCCATTTTGTTTTACAAAATGGCCGACTTCCT
TCCGCTTTTTTAAAAATAACGGCTCAGCGGCGGCGCGCGCTACGCG
(SEQ ID NO: 830)

Annotations:

Putative Domain	*Base range*
TATA Box	77 – 81
Initiator Element	95-110
Transcriptional Start Site	105
5' UTR Conserved Domain	165 - 235
ORF2	335 - 703
ORF2/2	335 – 699 ; 2326 – 2759
ORF2/3	335 – 699 ; 2552 – 2957
ORF2t/3	335-465 ; 2552 - 2957
ORF1	574 – 2775
ORF1/1	574 – 699 ; 2326 – 2775
ORF1/2	574 – 699 ; 2552 – 2759
Three open-reading frame region	2535 – 2746
Poly(A) Signal	2953 - 2958
GC-rich region**	3620 – 3648

Table A2. Novel *Anellovirus* amino acid sequences (*Alphatorquevirus,* Clade 6)

TTV-RTx1 (*Alphatorquevirus* Clade 6)	
ORF2	MSWSRPVHNATGIERNWWESCLRSHASSCGCGNFVNHINVLANRYGFAGS TETPGNPRPRPPVLSSTTSTPTDQSRPALPWHGDTGGEGASGDPAGDGERGA AEGDYGPEDLDALFDALDEE (SEQ ID NO: 831)
ORF2/2	MSWSRPVHNATGIERNWWESCLRSHASSCGCGNFVNHINVLANRYGFAGS TETPGNPRPRPPVLSSTTSTPTDQSRPALPWHGDTGGEGASGDPAGDGERGA AEGDYGPEDLDALFDALDEEQSKTLVKTKVPDPIDSLETYKLRTRQPWDLN MCSTRGTGDGASLASELSDACSKNLSTMMSILKNQKDLEYFLQQKQSPETK SSKKARFQRKKSRYSPQKRSRKRRYSDSSSDSSSDSCASGSSSNSSNNNSSRR KRACT (SEQ ID NO: 832)
ORF2/3	MSWSRPVHNATGIERNWWESCLRSHASSCGCGNFVNHINVLANRYGFAGS TETPGNPRPRPPVLSSTTSTPTDQSRPALPWHGDTGGEGASGDPAGDGERGA AEGDYGPEDLDALFDALDEENRVPKPRARRKLAFRGRKVATLHRRDPERG DTATVQATAQATAAPRAAAQTPPTTTPQDASGPAPKPPFIFPAINKVYLFPD RAPKPKPTSGDWETEYAACSAFDRPARTNLSSPPYYPGVPTPWQVKFSLKF Q (SEQ ID NO: 833)
ORF2t/3	MSWSRPVHNATGIERNWWESCLRSHASSCGCGNFVNHINVLANRNRVPKP RARRKLAFRGRKVATLHRRDPERGDTATVQATAQATAAPRAAAQTPPTTTP QDASGPAPKPPFIFPAINKVYLFPDRAPKPKPTSGDWETEYAACSAFDRPART NLSSPPYYPGVPTPWQVKFSLKFQ (SEQ ID NO: 834)
ORF1	MAWGYWWRRRFWRPRRRWRTWRRRRRLRPRRSRRTFRRTRRRVRRRRW GRRARRRGWRRRTYIRARRRRRRKRLVLTQWHPAVRRKCKITGYMPIVYC GHGRASFNYAWHSDDCIKQPLPFGGSLSTVSFNLKVLFDENQRGLNKWSYP NDQLDLARYKGCRLTFYRKKNTDYIAQYDISEPYQLDKYSCANYHPSKMM FAKNKILIPSYDTKPRGRQRVRVRIGPPKLFTDKWYSQSDLCKVNLVSLAVS AASFLHPFGSPQTANFCATFQVLQPFYYQAIGISSTKHSEVIDILYKKNTYWQ SNITSWFLTNVKNPKNMSTKMFEDINVKSNKDSNYDWFPFTPYTTENYSKIQ NAAQEYWKYLTSDHPQATNSNEGLVQPWTNATIKQYEYHLGMFSPIFIGPT RAKTKFKTAYFDCTYNPLLDKGMGNRIWYQYATKADTQISKTGCYCMLED IPIYAAFYGGYVDFIEMEIGKGQDIKENGLICCICRYTDPPMYNEQHPDMGFVF YNTNFGNGKWIDGRGDIPTYWMQRWRPVVLFQTDVIRDLVETGPFSYKDD

	LANTSLTMKYEFYFTWGGNQAYHQTIKNPCKDEGTGPHRQPRDVQVTDPT TVGPEYVFHAWDWRRGFLSERALRRMFEKPLNYDEYSKKPKRPRIFPPTET ESRNQELEESSLSEEEKSLLSTEEIQKEEIQRQFKRQLKRQLRLGQQLKLLQQ QLLKTQAGLHLNPLSYFPQ (SEQ ID NO:835)
ORF1/1	MAWGYWWRRRFWRPRRRWRTWRRRRRLRPRRSRRTFRRTRRRTIKNPCK DEGTGPHRQPRDVQVTDPTTVGPEYVFHAWDWRRGFLSERALRRMFEKPL NYDEYSKKPKRPRIFPPTETESRNQELEESSLSEEEKSLLSTEEIQKEEIQRQFK RQLKRQLRLGQQLKLLQQQLLKTQAGLHLNPLSYFPQ (SEQ ID NO: 836)
ORF1/2	MAWGYWWRRRFWRPRRRWRTWRRRRRLRPRRSRRTFRRTRRRKQSPETK SSKKARFQRKKSRYSPQKRSRKRRYSDSSSDSSSDSCASGSSSNSSNNNSSRR KRACT (SEQ ID NO: 837)

Table A3. Novel *Anellovirus* nucleic acid sequence (*Alphatorquevirus*)

Name	TTV-RTx2
Genus/Clade	*Alphatorquevirus*, Clade 6
Accession Number	SRR3479021

Full Sequence: 3704 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
CCCCGAAGTCCGTCACTAACCACGTGACTCCCACAGGCCAATCAGATGCT
ATGTCGTGCACTTCCTGGGCTGTGTCTACGTCCTCATATAAGTAACTGCA
CTTCCGAATGGCTGAGTTTTCCACGCCCGTCCGCAGCGGCAGCACCACGG
AGGGTGATCCCCGCGTCCCGTGGGCGGGTGCCGGAGGTGAGTTTACACAC
CGCAGTCAAGGGGCAATTCGGGCACGGGACTGGCCGGGCTATGGGCAAGG
CTCTTAAAAAGCTATGTTCTTCGGTAGGTGCTGGAGAAAGAAAAGGAAAG
TGCTTCTGCAAGATCTGTCAACTCCACCGAAAAAACCTGCTATGAGTGTG
TGGCTTCCTCCCATAGACAATGTTACCGAGCGTGAGAGGAGCTGGCTCTC
TAGCATTCTTCAGTCTCACAGAGCTTTTTGTGGGTGCCATGATGCTATCT
ATCATCTTAGCAGTCTGGCTGCTCGCTTTAATATGCAACCAGGGCCGTCG
CCGGGTGGTGATTCTAGGCCGCCGCGACCGCCACTAAGACGCCTGCCCGC
GCTCCCGGGTCCCAGAGACCCCCCTAGCGACACCAACAACCGCAGGTCAT
GGCCTACTGGGGATGGTGGAGACGGAGGCGCTGGCCAAGGCGCAGGTGGA
GGCGCTACCGCTACCGAAGAAGACTACCGCGCCGAAGACCTAGACGAGCT
GTACGCCGCCCTCGAAGGAGACGAGTAAGGAGGCGCCGCGGTAGGGGGTG
GTACAGAGGGCGACGCTACTCCCGCAGACGGTACAGACGTAGATATGTGA
GGCGAAAGAGAAAGACTCTAGTTTGGAGACAGTGGCAGCCTCAAAATATC
AGAAAATGCAGGATCAGGGGCATAATTCCCATCCTGATATGCGGACACGG
GAGGGGGGCCAGAAACTATGCGCTCCACAGCGACGACATAACCCCCCAGA
```

ACACCCCCTTCGGGGGAGGACTGAGCACCACCTCCTGGAGCCTAAAAGTG
CTATATGACCAGCACACCAGGGGACTCAACAGGTGGTCTGCCAGTAACGA
GAGCCTAGACCTTGCCAGATACAATGGCTGTAGTTTCACTTTCTACAGAG
ACAAAAAGACTGACTTTATAGTGACCTATGACACCTCTGCTCCCTACAAA
CTAGACAAATACAGCTCCCCCAGCTACCACCCAGGGTCCATGATGCTCAT
GACAAAACACAAAATCCTGATCCCCAGTTTTGACACAAAACCCAAAGGTC
CTGCCAAAATTAGAGTCAGAATCAAGCCCCCCAAAATGTTCTTAGATAAA
TGGTACACTCAAGACGACCTCTGTTCCGTTAATCTTGTGTCACTTGCGGT
TAGCGCAGCTTCCTTTACACATCCGTTCTGCCCACCACTAACTGACACTC
CTTGTGTAACGCTGCAGGTGTTGAAAGACTTCTACTACACAACCATAGGC
TACTCCTCTAATGCAGACAAAGTAGAGTCTGTATTCACTAACACTCTCTA
CAAACACTGCTGCTACTATCAGTCCTTTCTCACCACTCAATTTATAGCCA
AAATCACTCGCACACCAGATGGACAACCAGTAGCCACATTCTCTCCTCCT
ACCTCTTTCCCTGGCACAACTGTAACAAAAAGTTCCATAGAATCATTTAA
CCAATGGGTAACTTCCACAGGTACAAGTGGCTGGCTAACAAATGCAAACC
AACACTTTCATTTCTGTAACTATAAACCAGATGCCACAAAGCTAAAATGG
CTCAGACAGTACTACTTTGACTGGGAAACATACAAATTAGCAGATGTAAA
GCCAGACGGCCTTACACCCTCAGTAAACTGGTATGAGTACAGAATAGGCC
TCTTTAGTCCTATTTTCCTGAGCCCCTTCAGATCTAGCAGTCTAGACTTT
CCCAGAGCCTACCAGGATGTGAACTACAACCCCCTGGTAGACAAAGGAGT
GGGCAACATCATATGGTTCCAATACAACACAAAACCAGACACACAGCTGT
CAGTACCCAGCTGCAAGTGTGTCATAGAAGACAAACCCCTATGGGCAGCC
TTCTATGGCTACAGTGACTTTGTACAACAAGAGATAGGAGACTACACAGA
CGCAGAGGCCGTGGGCTTCGTCTGTGTCATCTGTCCATACACCAAACCCC
CTCTAAAAAACCCAGACAACCCCATGCAAGGGTTCATATTCTATGACAGC
CTTTTTGGCAATGGCAAGTGGATAGATGGCACGGGGCACGTCCCCCTTTA
CTGGCAGAGCAGGTGGAGGCCAGAGATGCTCTTCCAAGAAAACACCATGA
GAGACATCACACTATCTGGGCCCTTCAGCTACAAGGACGACTATAAGAAC
TGTGTACTGACTTGCAAATACAAATTTAACTTTCGATTCGGGGGCAATCT
TCTCCACGAACAGACGATCAGAAACCCATGCCCCACGGACGGACATCCCA
GTACCGGTAGACAGCCTAGAGACGTACAAGTGGTTGACCCGATCAAAGTG
GGCCCCCGGTTCGTGTTCCACTCCTGGGACTGGCGCAGAGGCTACCTTAG
CCCAGCAGCTCTCAAAAGAATTGGAGAGCAACCGCTCGATTATGAAGCTT
ATTCGTACCGCCCAAAGAGACCTAGAATCTTTCCTCCCACAGAAGGAGAC
CAGCTCGCCCGAAGTCGAGAAGAAGACTCATTTTCAGAGGAAGAAAGTCC
CCATATCTCGTTCGAAGAGGGGCAGGAACCGAAAGCCCAGGCGGTACAGC
AGCACCTCCTCCGACACCTCAGAAAGCAGCGAGAACTCCGAAAGCGACTC
CGAGCCCTGTTCCAAAGCCTCCAAAAGACGCAGGCGGGTCTCCACGTAAA
TCCATTATTATTCAACCAGCCTGCAATCAGGTTCTGATGTTCCCAGAGAT
GGGGCCTAAGCCAGCTCCCACTGCCCAAGACTGGCAGTGCGAATACGAGA
CATGTAAGCACTGGGATAGACCCCCCAGAAAGTTTCTCACAGACCCCCCT
TTCTATCCCTGGGCCCCTACTACTTACAATGTATCTTTCAAGCTAAACTT
CAAATAAACTAGGCCGTGGGAGTCTCACTTGTCGGTGTCTACCTCTTAAG
GTCACTAAGCACTCCGAGCGTCAGCGAGGAGTGCGACCCTTCCCCCTGGT
GCAACGCCCTCGGCGGCCGCGCGCTACGCCTTCGGCTGCGCGCGGCACCT
CGGACCCCCGCTCGTGCTGACGCGCTCGCGCGCGTCAGACCACTTCGGGC
TCGCGGGGGTCGGGAAATTTGCTAAACAGACTCCGAGTTGCCATTGGACA
CAGGAGCTGTGAATCAGTAACGAAAGTGAGTGGGGCCAGACTTCGCCATA
AGGCCTTTATCTTCTTGCCATTTGTCCGTGAGGAGGGGTCGCCAAGACGC
GGACCCCGTTTTCGGACCTTCCGAACTACCAAAATGGCCGATTCAGTGAC
GTCACGGCAGCCATTTTGTGTAAGCACCGCCCAGGACAGACGTCACAGTT
CAAAGGTCATCCTCGAGCGGAACTTACAGAAAATGGCGGTCAATTGCTTC
CGGGTCAAAGGTCACGCCTACGTCATAAGTCACGTGGTGGAGGCTACTGC
GCATACACGGAAGTAGGCCCCGCCACGTGACCGACCACGTGGGTGCTGCG
TCACGGCCGCCATTTTGTATCACAAAATGGCCGACTTCCTTCCTCTTTTT
CAAA (SEQ ID NO: 838)

Annotations:

Putative Domain	*Base range*
TATA Box	87 – 91
Initiator Element	105-120
Transcriptional Start Site	115
5' UTR Conserved Domain	175 - 245
ORF2	342 - 728
ORF2/2	342 – 724 ; 2414 – 2849
ORF2/3	342 – 724 ; 2643 – 3057
ORF1	599 – 2887
ORF1/1	599 – 724 ; 2414 – 2887
ORF1/2	599 – 724 ; 2643 – 2849
Three open-reading frame region	2626 – 2846
Poly(A) Signal	3052 - 3058

Table A4. Novel *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 6)

TTV-RTx2 (*Alphatorquevirus* Clade 6)	
ORF2	MSVWLPPIDNVTERERSWLSSILQSHRAFCGCHDAIYHLSSLAARFNMQPGP SPGGDSRPPRPPLRRLPALPGPRDPPSDTNNRRSWPTGDGGDGGAGQGAGG GATATEEDYRAEDLDELYAALEGDE (SEQ ID NO: 839)
ORF2/2	MSVWLPPIDNVTERERSWLSSILQSHRAFCGCHDAIYHLSSLAARFNMQPGP SPGGDSRPPRPPLRRLPALPGPRDPPSDTNNRRSWPTGDGGDGGAGQGAGG GATATEEDYRAEDLDELYAALEGDERSETHAPRTDIPVPVDSLETYKWLTR SKWAPGSCSTPGTGAEATLAQQLSKELESNRSIMKLIRTAQRDLESFLPQKE TSSPEVEKKTHFQRKKVPISRSKRGRNRKPRRYSSTSSDTSESSENSESDSEPC SKASKRRRRVST (SEQ ID NO: 840)
ORF2/3	MSVWLPPIDNVTERERSWLSSILQSHRAFCGCHDAIYHLSSLAARFNMQPGP SPGGDSRPPRPPLRRLPALPGPRDPPSDTNNRRSWPTGDGGDGGAGQGAGG GATATEEDYRAEDLDELYAALEGDERRPARPKSRRRLIFRGRKSPYLVRRG AGTESPGGTAAPPPTPQKAARTPKATPSPVPKPPKDAGGSPRKSIIIQPACNQ

	VLMFPEMGPKPAPTAQDWQCEYETCKHWDRPPRKFLTDPPFYPWAPTTYN VSFKLNFK (SEQ ID NO: 841)
ORF1	MAYWGWWRRRRWPRRRWRRYRYRRRLPRRRPRRAVRRPRRRRVRRRRG RGWYRGRRYSRRRYRRRYVRRKRKTLVWRQWQPQNIRKCRIRGIIPILICGH GRGARNYALHSDDITPQNTPFGGGLSTTSWSLKVLYDQHTRGLNRWSASNE SLDLARYNGCSFTFYRDKKTDFIVTYDTSAPYKLDKYSSPSYHPGSMMLMT KHKILIPSFDTKPKGPAKIRVRIKPPKMFLDKWYTQDDLCSVNLVSLAVSAA SFTHPFCPPLTDTPCVTLQVLKDFYYTTIGYSSNADKVESVFTNTLYKHCCY YQSFLTTQFIAKITRTPDGQPVATFSPPTSFPGTTVTKSSIESFNQWVTSTGTS GWLTNANQHFHFCNYKPDATKLKWLRQYYFDWETYKLADVKPDGLTPSV NWYEYRIGLFSPIFLSPFRSSSLDFPRAYQDVNYNPLVDKGVGNIIWFQYNTK PDTQLSVPSCKCVIEDKPLWAAFYGYSDFVQQEIGDYTDAEAVGFVCVICPY TKPPLKNPDNPMQGFIFYDSLFGNGKWIDGTGHVPLYWQSRWRPEMLFQE NTMRDITLSGPFSYKDDYKNCVLTCKYKFNFRFGGNLLHEQTIRNPCPTDGH PSTGRQPRDVQVVDPIKVGPRFVFHSWDWRRGYLSPAALKRIGEQPLDYEA YSYRPKRPRIFPPTEGDQLARSREEDSFSEEESPHISFEEGQEPKAQAVQQHLL RHLRKQRELRKRLRALFQSLQKTQAGLHVNPLLFNQPAIRF (SEQ ID NO: 842)
ORF1/1	MAYWGWWRRRRWPRRRWRRYRYRRRLPRRRPRRAVRRPRRRRTIRNPCP TDGHPSTGRQPRDVQVVDPIKVGPRFVFHSWDWRRGYLSPAALKRIGEQPL DYEAYSYRPKRPRIFPPTEGDQLARSREEDSFSEEESPHISFEEGQEPKAQAV QQHLLRHLRKQRELRKRLRALFQSLQKTQAGLHVNPLLFNQPAIRF (SEQ ID NO: 843)
ORF1/2	MAYWGWWRRRRWPRRRWRRYRYRRRLPRRRPRRAVRRPRRRRKETSSPE VEKKTHFQRKKVPISRSKRGRNRKPRRYSSTSSDTSESSENSESDSEPCSKAS KRRRRVST (SEQ ID NO: 844)

Table A5. Novel *Anellovirus* nucleic acid sequence (*Alphatorquevirus*)

Name	TTV-RTx3
Genus/Clade	*Alphatorquevirus*, Clade 4

Accession Number SRR3479781

Full Sequence: 3653 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
CCAACCAGAGTCTATGTCGTGCACTTCCTGGGCATGGTCTACGTAATAAT
ATAAAGCGGTGCACTTCCGAATGGCTGAGTTTTCCACGCCCGTCCGCAGC
GAGATCGCGACGGAGGAGCGATCGAGCGTCCCGAGGGCGGGTGCCGGAGG
TGAGTTTACACACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGCCGG
GCTATGGGCAAGGCTCTTAAAAAGCCATGTTTCTCGGTAAACTTTACAGG
CAGAAAAGGAAACTGCTACTGCAGCCTGTGCGTGCTCCACAGACGCCATC
TTCCATGAGCTCTACCTGGCGAGTGCCCCGCGGCGATGTCTCCGCCCGCG
AGCTATGTTGGTACCGCTCAGTTCGAGAGAGCCACGATGCTTTTTGTGGC
TGTCGTGATCCTGTTTTTCATCTTTCTCGTCTGGCTGCACGTTCTAACCA
TCAGGGACCTCCGACGCCCCCCACGGACGAGCGCCCGTCGGCGTCTACCC
CAGTGAGGCGCCTGCTGCCGCTGCCCTCCTACCCCGGCGAGGGTCCCCAG
GCTAGATGGCCTGGTGGGGATGGAGAAGGCGCTGGTGGCGCCCGCGGAGG
CGCTGGAGATGGCGGCGCCCGCGCAGGCGAAGAAGAGTACCGGCCCGAAG
ACCTCGACGAGCTGTTCGACGCTATCGAACAAGAACAGTAAGGAGACGGA
GGCGAGGGTGGCGGAGGGGCTACAGGCGCCGTTACAGACTGAGACGCTAC
CGTAGAAGGGGCAGGCGACGCAAAAAAATAGTACTGACTCAGTGGAACCC
CCAGACTGTCAGAAAGTGCTTTATCAGAGGACTGATGCCAGTACTATGGG
CGGGCATGGGCACGGGGGGCCACAACTACGCCGTCCGCTCAGATGACTTT
GTGGTAGACAGAGGCTTCGGGGGCTCCTTCGCCACAGAAACTTTCTCCCT
GAGGGTCCTCTTTGACCAGTACCAGAGAGGATTTAATAGGTGGTCTCACA
CCAACGAAGACCTAGACCTGGCCCGCTACACGGGCTGCAAATGGACATTT
TACAGACACCAAGACACAGACTTTATAGTGTACTTTACAAACAATCCCCC
CATGAAAACCAACCAGCACACAGCCCCTCTCACAACTCCAGGCATGCTCA
TGAGGAGCAAGTATAAAATACTAGTGCCCAGTTTTAAAACAAGACCAAAG
GGCAGAAAAACAGTGTCAGTGAGAGTTAGACCCCCCAAACTGTTTCAGGA
CAAATGGTATACTCAACAGGACCTCTGTCCAGTACCCCTCGTCCAACTGA
ACGTGACCGCAGCGGATTTCACACATCCGTTCGGCTCACCACTAACTGAC
ACGCCTTGCATAAGATTCCAAGTTTTAGGGAACTTATACAACAAGTGCCT
AAATATAGATCTTCCGCAATTTGATGAGGACGGTGAGATACTCACTTCAA
CACCTTATAACAGAGAAAACAAAGAAGATCTTAAAAAGCTTTATAAAACT
CTATTTGTAGATGAACACGCAGGCAATTATTGGCAGACATTCTTAACCAA
CACAATGGTAAAGTCACACATAGATGCAAACCAAGCAAAGACATACGATC
AAGAAAAAACTGCTGCAGAACAAGGTAAAGACCCCTTCCCAACAAACCCA
CCAAAAGACCAATTCACTACCTGGAACAAGAAACTAGTAGACCCTAGAGA
CAGCAACTTTCTCTTTGCCACATATCACCCAAAAAACATTAAAAAAGCTA
TAAAAACCATGAGAGACAACAACTTTGCTCTCACCACAGGCAAAAATGAC
ATATATGGAGACTACACCGCGGCCTACACCAGAAACACCCACATGCTAGA
CTACTACCTAGGCTTTTATAGCCCCATATTTCTTTCCAGCGGTAGGTCCA
ACACAGAGTTCTGGACCGCCTACAGAGACATAGTATATAATCCCCTCTTA
GACAAAGGCACAGGCAACATGATCTGGTTCCAATATCACACAAAAACAGA
CAATATATACAAAAAACCAGAGTGCCACTGGGAGATACTAGACATGCCCC
TGTGGGCCCTCTGCAACGGGTATGTAGAGTACCTAGAGAGCCAAATAAAG
TACGGGGACATCCTAGTAGAGGGCAAAGTCCTCATCAGATGCCCCTACAC
CAAACCCGCACTGGTAGACCCCAATAACAGCCTAGCTGGTTACGTGGTAT
TCAACACCACCTTCGGCCAGGGAAAATGGATAGATGGCAAAGGCTACATC
CCCCTACACGAGAGGAGCAAGTGGTACGTCATGCTCAGATACCAGACCGA
CGTACTCCATGACATAGTGACTTGTGGACCCTGGCAGTACAGAGACGATA
```

ACAAAAACTCTCAGCTAATAGCCAAGTACAGATTCAAGTTCTACTGGGGA
GGTAACATGGTACATTCTCAGGTCATCAGAAACCCGTGCAAAGACACCCA
AGTATCCGGACCCCGTCGACAGCCTCGCGAAGTACAAGTCGTTGACCCGC
AACTCATTACGCCGCCGTGGGTCCTCCACTCGTTCGACCAGAGACGAGGA
ATGTTTACTGCAGGAGCTATCAAACGTCTGCTCAAGCAACCAATACCTGG
CGAGTATGCTCCTACACCACTCAGGGTCCCGCTCCTCTTTCCCTCCTCAG
AGTTCCAGCGAGAGGGAGAAGATGCAGAAAGCGGCTCAGGTTCACCACCC
AAGAGACCGCGACTCTGGCAGGAAGAGGCCAACCAGACGCAAACGGAGTC
CTCGGAGGGGCCGGCGGAGACGACGAGGGAGCTCCTCGAGCGAAAGCTCA
GAGAGCAGCGAGTCCTCAACCTCCAACTCCAGCATGTCGCAGTACAACTC
GCCAAAACCCAAGCGAACCTCCACATAAACCCCCTATTATACTCCCAGCC
TTAAACAAAGTGTATCTATTCCCCCCTGACAAGCCCACTCCCATACAGNN
NNNNNNNNNNNNNNNNAACACAGAGTTCGAAGCCTGCCAGGCCTTCGACA
GACCACCTAGAAAATACCTCTCAGACACACCTACCTACCCTTGGCTCCCC
GTCCCCAATCCTGAAATAAAGGTCAGCTTTAAGCTCGGTTTCAAATCTTA
CAAGGCCGTGGGAGTTTCACTGGTCGGTGTCTACCTCTTAAGGTCACTAA
GCACTCCGAGCGTCAGCGAGGAGTGCGACCCTTCCCCCTGGTGCAACGCC
CTCGGCGGCCGCGCGCTACGCCTTCGGCTGCGCGCGGCACCTCGGACCCC
CGCTCGTGCTGACGCGCTCGCGCGCGTCAGACCACTTCGGGCTCGCGGGG
GTCGGGAATTTTGCTAAACAGACTCCGAGTTGCCATTGGACACTGTAGCT
GTGAATCAGTAACGAAAGTGAGTGGGGCCAGACTTCGCCATAAGGCCTTT
ATCTTCTTGCCATTGGTCCGTGTAGGGGGTCGCCATAGGCTTCGGGTTCG
GTTTTAGGCCTTCCGGACTACAAAAATGGCGGATTTAGTGACGTCACGGC
CGCCATTTTAAGTAGGTGCCGTCCAGGACTGCTGTTCCGGGTCACAGGGC
ATCCTCGGCGGAACTTACACAAAATGGCGGTCAAAAACATCCGGGTCAAA
GGTCGCAGCTACGTCATAAGTCACGTGCAGGGGTCCTGCTGCGTCATATG
CGG (SEQ ID NO: 845)

Annotations:

Putative Domain	*Base range*
TATA Box	50 – 55
Initiator Element	68-83
Transcriptional Start Site	78
5’ UTR Conserved Domain	138 - 208
ORF2	305 - 691
ORF2/2	305 – 687 ; 2422 – 2878
ORF2/3	305 – 687 ; 2564 – 3317
ORF2t/3	305 – 360 ; 2564 – 3317
ORF1	556 – 2904
ORF1/1	556 – 687 ; 2422 – 2904
ORF1/2	556 – 687 ; 2564 – 2878
Three open-reading frame region	2626 – 2846
Poly(A) Signal	3316 - 3319

Table A6. Novel *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 4)

TTV-RTx3 (*Alphatorquevirus* Clade 4)	
ORF2	MSSTWRVPRGDVSARELCWYRSVRESHDAFCGCRDPVFHLSRLAARSNHQ GPPTPPTDERPSASTPVRRLLPLPSYPGEGPQARWPGGDGEGAGGARGGAG DGGARAGEEEYRPEDLDELFDAIEQEQ (SEQ ID NO: 846)
ORF2/2	MSSTWRVPRGDVSARELCWYRSVRESHDAFCGCRDPVFHLSRLAARSNHQ GPPTPPTDERPSASTPVRRLLPLPSYPGEGPQARWPGGDGEGAGGARGGAG DGGARAGEEEYRPEDLDELFDAIEQEQSSETRAKTPKYPDPVDSLAKYKSLT RNSLRRRGSSTRSTRDEECLLQELSNVCSSNQYLASMLLHHSGSRSSFPPQSS SEREKMQKAAQVHHPRDRDSGRKRPTRRKRSPRRGRRRRRGSSSSESSESSE SSTSNSSMSQYNSPKPKRTST (SEQ ID NO: 847)
ORF2/3	MSSTWRVPRGDVSARELCWYRSVRESHDAFCGCRDPVFHLSRLAARSNHQ GPPTPPTDERPSASTPVRRLLPLPSYPGEGPQARWPGGDGEGAGGARGGAG DGGARAGEEEYRPEDLDELFDAIEQEQSYQTSAQATNTWRVCSYTTQGPAP LSLLRVPARGRRCRKRLRFTTQETATLAGRGQPDANGVLGGAGGDDEGAP RAKAQRAASPQPPTPACRSTTRQNPSEPPHKPPIILPALNKVYLFPPDKPTPIQ XXXXXXNTEFEACQAFDRPPRKYLSDTPTYPWLPVPNPEIKVSFKLGFKSYK AVGVSLVGVYLLRSLSTPSVSEECDPSPWCNALGGRALRLRLRAAPRTPAR ADALARVRPLRARGGREFC (SEQ ID NO: 848)
ORF2t/3	MSSTWRVPRGDVSARELCWSYQTSAQATNTWRVCSYTTQGPAPLSLLRVP ARGRRCRKRLRFTTQETATLAGRGQPDANGVLGGAGGDDEGAPRAKAQR AASPQPPTPACRSTTRQNPSEPPHKPPIILPALNKVYLFPPDKPTPIQXXXXXX NTEFEACQAFDRPPRKYLSDTPTYPWLPVPNPEIKVSFKLGFKSYKAVGVSL VGVYLLRSLSTPSVSEECDPSPWCNALGGRALRLRLRAAPRTPARADALAR VRPLRARGGREFC (SEQ ID NO: 849)
ORF1	MAWWGWRRRWWRPRRRWRWRRPRRRRRVPARRPRRAVRRYRTRTVRR RRRGWRRGYRRRYRLRRYRRRGRRRKKIVLTQWNPQTVRKCFIRGLMPVL WAGMGTGGHNYAVRSDDFVVDRGFGGSFATETFSLRVLFDQYQRGFNRW SHTNEDLDLARYTGCKWTFYRHQDTDFIVYFTNNPPMKTNQHTAPLTTPG MLMRSKYKILVPSFKTRPKGRKTVSVRVRPPKLFQDKWYTQQDLCPVPLVQ LNVTAADFTHPFGSPLTDTPCIRFQVLGNLYNKCLNIDLPQFDEDGEILTSTP

	YNRENKEDLKKLYKTLFVDEHAGNYWQTFLTNTMVKSHIDANQAKTYDQ EKTAAEQGKDPFPTNPPKDQFTTWNKKLVDPRDSNFLFATYHPKNIKKAIKT MRDNNFALTTGKNDIYGDYTAAYTRNTHMLDYYLGFYSPIFLSSGRSNTEF WTAYRDIVYNPLLDKGTGNMIWFQYHTKTDNIYKKPECHWEILDMPLWAL CNGYVEYLESQIKYGDILVEGKVLIRCPYTKPALVDPNNSLAGYVVFNTTFG QGKWIDGKGYIPLHERSKWYVMLRYQTDVLHDIVTCGPWQYRDDNKNSQ LIAKYRFKFYWGGNMVHSQVIRNPCKDTQVSGPRRQPREVQVVDPQLITPP WVLHSFDQRRGMFTAGAIKRLLKQPIPGEYAPTPLRVPLLFPSSEFQREGED AESGSGSPPKRPRLWQEEANQTQTESSEGPAETTRELLERKLREQRVLNLQL QHVAVQLAKTQANLHINPLLYSQP (SEQ ID NO: 850)
ORF1/1	MAWWGWRRRWWRPRRRWRWRRPRRRRRVPARRPRRAVRRYRTRTVIRN PCKDTQVSGPRRQPREVQVVDPQLITPPWVLHSFDQRRGMFTAGAIKRLLK QPIPGEYAPTPLRVPLLFPSSEFQREGEDAESGSGSPPKRPRLWQEEANQTQT ESSEGPAETTRELLERKLREQRVLNLQLQHVAVQLAKTQANLHINPLLYSQP (SEQ ID NO: 852)
ORF1/2	MAWWGWRRRWWRPRRRWRWRRPRRRRRVPARRPRRAVRRYRTRTELSN VCSSNQYLASMLLHHSGSRSSFPPQSSSEREKMQKAAQVHHPRDRDSGRKR PTRRKRSPRRGRRRRRGSSSSESSESSESSTSNSSMSQYNSPKPKRTST (SEQ ID NO: 853)

Table A7. Novel *Anellovirus* nucleic acid sequence (*Alphatorquevirus*)

Name	TTV-RTx4
Genus/Clade	*Alphatorquevirus*, Clade 4
Accession Number	SRR3481579

Full Sequence: 3742 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
AAAGTGCTACGTCACTAACCACGTGACACCCACAGGCCAACCGAATGCTA
TGTCGTGCACTTCCTGGGCCGGGTCTACGTCCTCATATAACTACCTGCAC
TTCCGAATGGCTGAGTTTTCCACGCCCGTCCGCAGCGGTGAAGCCACGGA
GGGAGATCAGCGCGTCCCGAGGGCGGGTGCCGAAGGTGAGTTTACACACC
GAAGTCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTATGGGCAAGGC
```

TCTGAAAAAAGCATGTTTATTGGCAGGCATTACAGAAAGAAAAGGGCGCT
GCCACTGTGTGCTGTGCGATCAACAAAGAAGGCTTGCAAACTACTAATAG
TAATGTGGACCCCACCTCGCAATGACCAACAGTACCTTAACTGGCAATGG
TACTCAAGTATACTTAGCTCCCACGCTGCTATGTGCGGGTGTCCCGACGT
TGTTGCTCATTTTAATCATCTTGCTTCTGTGCTTCGCGCCCCGCAAAATC
CACCCCCACCCGGTCCCCAGCGAAACCTGCCCCTCCGACGGCTGCCGGCT
CTCCCGGCTGCGCCAGAGGCGCCCGGAGATAGAGCACCATGGCCTATGGC
TGGTGGCGCCGGAGGAGAAGACGGTGGCGCAGGTGGAGACGCAGACCATG
GAGGCGCCGCTGGAGGACCAGAAGACGCAGACCTGTTAGACGCCGTGGCC
GCCGCAGAAACGTAAGGAGACGCCGCAGAGGAGGGAGGTGGAGGAGGAGG
TACAGGAGATGGAAAAGAAAGGGCAGACGCAGAAAAAAAGCTAAAATAAT
AATAAGACAATGGCAACCTAACTACAGAAGGAGATGTAACATAGTAGGCT
ATATTCCTGTACTGATATGTGGCGAAAATACTGTCAGCAGAAACTATGCC
ACACACTCAGACGATACTAACTACCCAGGACCCTTTGGGGGGGGTATGAC
TACAGACAAATTTACCTTAAGAATTCTGTATGACGAGTACAAAAGGTTTA
TGAACTATTGGACAGCATCTAATGAAGACCTAGACCTCTGTAGATATCTA
GGAGTAAACCTGTACTTTTTTAGACACCCAGAAGTAGACTTTATTATAAA
AATAAATACCATGCCCCCTTTTCTAGACACAGAACTAACAGCTCCTAGCA
TACACCCAGGAATGCTAGCCTTAGACAAAAGAGCAAGATGGATACCTAGC
TTAAAATCTAGACCAGGAAAAAAACACTATATTAAAATAAGAGTAGGGGC
GCCTAAAATGTTCACAGATAAATGGTACCCCCAAACAGATCTTTGTGACA
TGGTGCTGCTAACTGTCTATGCAACCGCAGCGGATATGCAATATCCGTTC
GGCTCACCACTAACTGACTCTGTGGTTGTGAACTTCCAGGTTCTGCAATC
CATGTATGATGAAACCATTAGCATATTACCAGATCAAAAGGAGAAAAGAA
TAACGCTGCTCACTAGTATAGCCTTTTATAACACCACACAAACTATAGCC
CAATTAAAGCCATTTATAGATGCAGGCAATATGACTTCAACTACAACAGC
AACAACATGGGGATCATACATAAACACAACCAAATTTAATACAGCAGCCA
CTACAACATACACATACCCAGGCAGTACTACAACTACAGTAACTATGTTA
ACTTGTAATGACTCCTGGTACAGAGGAACAGTATATAACGACCAAATTAA
AAATTTACCAAAGGAAGCAGCTCAATTATACTTAAAAGCAACAAAAACCT
TACTAGGAAACACCTTCACAAATGACGACCACACACTAGAATACCATGGA
GGACTGTACAGCTCAATTTGGCTGTCCCCCGGCAGATCTTACTTTGAAAC
ACCAGGAGCATACACAGACATAAAATACAACCCATTTACAGACAGAGGAG
AAGGAAACATGCTATGGATAGACTGGCTAAGCAAAAAAAATATGAACTAT
GACAAACTACAAAGTAAATGTTTAATATCAGACCTACCTTTATGGGCAGC
AGCATATGGATATTTAGAATTTTGTGCAAAAAGTACAGGAGACCAAAATA
TACACATGAATGCCAGACTACTAATAAGAAGTCCCTTTACAGACCCCCAA
CTACTAGTACACACAAACCCCACAAAAGGCTTTGTTCCCTACTCTTTAAA
CTTTGGAAATGGTAAAATGCCAGGAGGTAGTAGTAATGTTCCTATTAGAA
TGAGAGCTAAATGGTATCCAACATTGTTTCACCAGCAAGAAGTACTAGAG
GCCTTAGCACAGTCAGGCCCCTTTGCATACCACTCAGACATTAAAAAAGT
ATCTCTGGGTATGAAATACCGTTTTAAGTGGATCTGGGGTGGAAACCCCG
TTCGCCAACAGGTTGTTAGAAATCCCTGCAAAGACTCCCACTCCTCGGTC
AATAGAGTCCCTAGAAGCTTACAAATCGTTGACCCGAAATACAACTCACC
GGAACTCACATTCCATACGTGGGACTTCAGACGTGGCCTCTTTGGCCAGA
AAGCTATTGAGAGAATGCAACAACAACCAACAACTACTGACATTTTTTCA
GCAGGCCGCAAGAGACCCAGGAGGGACACCGAGGTGTACCACTCCAGCCA
AGAAGGGGAGCAAAAAGAAAGCTTACTTTTCCCCCCAGTCAAGCTCCTCA
GACGAGTCCCCCCGTGGGAAGACTCGCAGCAGGAGGAAAGCGGGTCGCAA
AGCTCAGAGGAAGAGACGCAGACCGTCTCCCAGCAGCTCAAGCAGCAGCT
GCAGCAACAGCGAATCCTGGGAGTCAAACTCATACTCCTGTTCAACCAAG
TCCAAAAAATCCAACAAAATCAAGATATCAACCCTACCTTGTTACCAAGG
GGGGGGGATCTAGCATCCTTATTTCAAATAGCACCATAAACATGTTTGGA
GACCCCAAACCTTACAACCCTTCCAGTAATGACTGGAAAGAGGAGTATGA
GGCCTGTAGAATATGGGACAGACCCCCAAGAGGCAATCTAAGAGACACCC
CCTTTTACCCCTGGGCCCCCAAAGAAAACCAGTACCGTGTAAACTTTAAA
CTTGGATTTCAATAAAGCTAGGCCGTGGGACTTTCACTTGTCGGTGTCTG

CTTATAAAAGTAACCAAGCACTCCGAGCGAAGCGAGGAGTGCGACCCTTG
GGGGCTCAACGACTTCGGAGCCGCGCGTTAAGCCTTCGGCTGCGCGCGGC
ACCTCAGACCCCCGCTCGTGCTGACACGCTTGCGCGTGTCAGACCACTTC
GGGCTCGCGGGGGTCGGGAAATTTATTAAACAGACTCCGAGTTGCCATTG
GACACAGTAGTCTATGAACAGCAACGAAAGTGAGTGGGGCCAGACTTCGC
CATAAGGCCTTTATCTTCTTGCCATTTGTCAGTATAGAGGGTCGCCATAG
GCTTCGGTCTCCATTTTAACCTGTAAAAACTACCAAAATGGCCGTTCCAG
TGACGTGACAGCCGCCATTTTAAGTAGCTGACGTCAAGGATTGACGTAAA
GGTTAAAGGTCATCCTCGGCGGAAGCTACACAAAATGGTGGACAACATCT
TCCGGGTCAAAGGTCGTGCACACGTCAAAAGTCACGTGGTGGGGACCCGC
TGTAACCCGGAAGTAGGCCCCGTCACGTGATTTGTCACGTGTGTACACGT
CACAGCCGCCATTTTGTTTTACAAAATGGCTGACTTCCTTCCTCTTTTTT
GAAAAAAGGCGCCAAAAAAGGCTCCGCCCCCCGGCCCCCCCC (SEQ ID NO: 854)

Annotations:

Putative Domain	*Base range*
TATA Box	86 – 90
Initiator Element	104 – 119
Transcriptional Start Site	114
5' UTR Conserved Domain	174 – 244
ORF2	353 – 715
ORF2/2	353 – 711 ; 2362 – 2863
ORF2/3	353 – 711 ; 2555 – 3065
ORF2t/3	353 – 432; 2555 – 3065
ORF1	589 – 2889
ORF1/1	589 – 711 ; 2362 – 2889
ORF1/2	589 – 711 ; 2555 – 2863
Three open-reading frame region	2555 – 2863
Poly(A) Signal	3062 – 3066
GC-rich region, or a portion thereof**	3720 – 3742

Table A8. Novel *Anellovirus* amino acid sequences (*Alphatorquevirus,* Clade 4)

TTV-RTx4 (*Alphatorquevirus* Clade 4)	
ORF2	MWTPPRNDQQYLNWQWYSSILSSHAAMCGCPDVVAHFNHLASVLRAPQN PPPPGPQRNLPLRRLPALPAAPEAPGDRAPWPMAGGAGGEDGGAGGDADH GGAAGGPEDADLLDAVAAAET (SEQ ID NO: 855)
ORF2/2	MWTPPRNDQQYLNWQWYSSILSSHAAMCGCPDVVAHFNHLASVLRAPQN PPPPGPQRNLPLRRLPALPAAPEAPGDRAPWPMAGGAGGEDGGAGGDADH

	GGAAGGPEDADLLDAVAAAETLLEIPAKTPTPRSIESLEAYKSLTRNTTHRN SHSIRGTSDVASLARKLLRECNNNQQLLTFFQQAARDPGGTPRCTTPAKKGS KKKAYFSPQSSSSDESPRGKTRSRRKAGRKAQRKRRRPSPSSSSSSCSNSESW ESNSYSCSTKSKKSNKIKISTLPCYQGGGI (SEQ ID NO: 856)
ORF2/3	MWTPPRNDQQYLNWQWYSSILSSHAAMCGCPDVVAHFNHLASVLRAPQN PPPPGPQRNLPLRRLPALPAAPEAPGDRAPWPMAGGAGGEDGGAGGDADH GGAAGGPEDADLLDAVAAAETPQETQEGHRGVPLQPRRGAKRKLTFPPSQ APQTSPPVGRLAAGGKRVAKLRGRDADRLPAAQAAAAATANPGSQTHTPV QPSPKNPTKSRYQPYLVTKGGGSSILISNSTINMFGDPKPYNPSSNDWKEEYE ACRIWDRPPRGNLRDTPFYPWAPKENQYRVNFKLGFQ (SEQ ID NO: 857)
ORF2t/3	MWTPPRNDQQYLNWQWYSSILSSHAAMPQETQEGHRGVPLQPRRGAKRK LTFPPSQAPQTSPPVGRLAAGGKRVAKLRGRDADRLPAAQAAAAATANPGS QTHTPVQPSPKNPTKSRYQPYLVTKGGGSSILISNSTINMFGDPKPYNPSSND WKEEYEACRIWDRPPRGNLRDTPFYPWAPKENQYRVNFKLGFQ (SEQ ID NO: 858)
ORF1	MAYGWWRRRRRRWRRWRRRPWRRRWRTRRRRPVRRRGRRRNVRRRRR GGRWRRRYRRWKRKGRRRKKAKIIIRQWQPNYRRRCNIVGYIPVLICGENT VSRNYATHSDDTNYPGPFGGGMTTDKFTLRILYDEYKRFMNYWTASNEDL DLCRYLGVNLYFFRHPEVDFIIKINTMPPFLDTELTAPSIHPGMLALDKRARW IPSLKSRPGKKHYIKIRVGAPKMFTDKWYPQTDLCDMVLLTVYATAADMQ YPFGSPLTDSVVVNFQVLQSMYDETISILPDQKEKRITLLTSIAFYNTTQTIAQ LKPFIDAGNMTSTTTATTWGSYINTTKFNTAATTTYTYPGSTTTTVTMLTCN DSWYRGTVYNDQIKNLPKEAAQLYLKATKTLLGNTFTNDDHTLEYHGGLY SSIWLSPGRSYFETPGAYTDIKYNPFTDRGEGNMLWIDWLSKKNMNYDKLQ SKCLISDLPLWAAAYGYLEFCAKSTGDQNIHMNARLLIRSPFTDPQLLVHTN PTKGFVPYSLNFGNGKMPGGSSNVPIRMRAKWYPTLFHQQEVLEALAQSGP FAYHSDIKKVSLGMKYRFKWIWGGNPVRQQVVRNPCKDSHSSVNRVPRSL QIVDPKYNSPELTFHTWDFRRGLFGQKAIERMQQQPTTTDIFSAGRKRPRRD TEVYHSSQEGEQKESLLFPPVKLLRRVPPWEDSQQEESGSQSSEEETQTVSQ QLKQQLQQQRILGVKLILLFNQVQKIQQNQDINPTLLPRGGDLASLFQIAP (SEQ ID NO: 859)

ORF1/1	MAYGWWRRRRRRWRRWRRRPWRRRWRTRRRRPVRRRGRRRNVVRNPC KDSHSSVNRVPRSLQIVDPKYNSPELTFHTWDFRRGLFGQKAIERMQQQPTT TDIFSAGRKRPRRDTEVYHSSQEGEQKESLLFPPVKLLRRVPPWEDSQQEES GSQSSEEETQTVSQQLKQQLQQQRILGVKLILLFNQVQKIQQNQDINPTLLPR GGDLASLFQIAP (SEQ ID NO: 860)
ORF1/2	MAYGWWRRRRRRWRRWRRRPWRRRWRTRRRRPVRRRGRRRNAARDPG GTPRCTTPAKKGSKKKAYFSPQSSSSDESPRGKTRSRRKAGRKAQRKRRRPS PSSSSSSCSNSESWESNSYSCSTKSKKSNKIKISTLPCYQGGGI (SEQ ID NO: 861)

Table A9. Novel *Anellovirus* nucleic acid sequence (*Alphatorquevirus*)

Name	TTV-RTx5b
Genus/Clade	*Alphatorquevirus*, Clade 5
Accession Number	SRR3481639

Full Sequence: 3553 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
ATACCTCATCATATAAAGCGGCGCACTTCCGAATGGCTGAGTTTTCCACG
CCCGTCCGCAGCGAGATCGCGACGGAGGAGCGATCGAGCGTCCCGAGGGC
GGGTGCCGGAGGTGAGTTTACACACCGCAGTCAAGGGGCAATTCGGGCTC
GGGACTGGCCGGGCTATGGGGCAAGACTCTTAAAAAAGCCATGTTTCTCG
GTAAACTTTACAGAAAGAAAAGGGCACTGTCACTGCTACGCGTGCGAGCT
CCAGAGGCGAAACCACCTGCTATGAGTTGGAGACCCCCGGTGCACAACCC
CAATGGGATCGAGAGAAACCTGTGGGAGGCATTCTTTCGCATGCATGCTT
CAGCTTGTGGTTGTGGCGATCTTGTTGGCCATCTTACTGTACTGGCTGGT
CGGTATGGTGCTCCTCCTCGTCCCCCGGCCCCCGGCGCTCCCAGACCACC
GCTGATACGCCAGCTGGCCCTTCCGGCGCCCCCCGCCGATCCTCAACAGG
CTAACCCACAATGGCCTGGTGGGGACGGTGGAGAAGATGGCGCTGGAGGC
CCCGCCGCTGGCGGCGCCGTCGCAGACGCCGAGTACCAAGAAGACGAGCT
CAACGCCCTGTTCGACGCCGTCGAGCAAGAAGAGTAAGGAGGAGGCGATG
GGGGAGGCGGAGGTGGAGACGGGGGTACAGACGCAGACTGAGACTAAGAC
GCAGACGCAGACGAAAGCGAAAGATAGTACTAACTCAGTGGAATCCCGCC
AAAGTGCGGAGGTGTACTATTAAGGGAGTTCTGCCCATGATCCTGTGCGG
GGCCGGGCGCTCGGGGTTTAACTACGGACTGCACAGCGACGACTACACTG
TACAGAAGCCCCTTGGCCAGAACCCCCACGGGGGCGGCATGAGTACAGTG
ACTTTTAGCCTACAGGTGCTCTATGACCAGTACCAGAGGTTTATGAACAA
GTGGTCGTACTCCAACGACCAGCTAGACCTCGCCAGGTACTTTGGCTGCA
CCTTCTGGTTCTACAGACACCCAGAGGTGGACTTTGTAGCTCAGTTTGAC
AACGTTCCCCCCATGAAAATGGACGAGAACACAGCCCCAAACACTCATCC
CTCTTTCTTACTACAGAACAAACACAAGGTTAAAATTCCCAGCTTTAAAA
CAAAGCCTTTTGGTAAAAAAAGAGTTAGAGTTACAGTAGGGCCCCCCAAA
CTGTTTGAAGATAAGTGGTACAGCCAACATGACTTGTGTAAGGTGCCCCT
```

AGTCAGTTGGCGGTTAACCGCAGCTGACTTCAGGTTTCCGTTCTGCTCAC
CACAAACTGACAACCCTTGCTACACCTTCCAGGTATTGCATGAAGAGTAT
TACCCAGTAATAGGCACTTCTGCTTTAGAAAACGGCAGTAACTACAATAG
CTCAGCTATAACAGCCTTAGAAAAATTCTTATATGAAAAATGCACACACT
ATCAAACATTTGCCACAGACACCAGACTTAATCCTCAGCGACCAGTGTCA
TCTACAAATGCAAACAAAACATACACCCCCTCAGGCTCCCAAGAAACAAT
AGTGTGGGGGCAGTCAGATTTTAATTTATTTAAAAAGCACACAGACAGCA
ACTATGGCTACTGCACCTACTGTCCTACCAATGACTTAGCTACAAAAATT
AAAAAGTACAGAGACAAAAGATTCGACTGGCTAACAAACATGCCAGTAAC
AAACACCTGCCACATAAATGCCACCTTCGCCCGAGGCAAAATTAAAGAAT
GGGAGTACCACCTAGGGTGGTTCTCAAACATCTTTATAGGCAACCTGAGA
CACAACCTAGCATTCCGGGCCGCATACATAGACATCACCTANACAGACAA
GGGAGAAGGCAACATTATCTGGTTCCAGTACCTCACTAAACCCACCACAG
AGTACATAGAAGCCCAAGCAAAGTGCTCCATCACAAACATACCCCTGTAT
GCTGCTTTTTATGGCTACGAAGACTACCTCCAGAGAACACTAGGCCCCTA
CCAAGATGTAGAAACCCTAGGTATAATCTGTGTTAAATGTCCCTACACAG
ATCCCCCTCTAGTTCACAAGTCTACAGATAAAAAGAACTGGGGCTACGTG
TTCTACGACGTGCACTTTGGCAACGGAAAGACCCCAGAGGGACTGGGCCA
GGTGCACCCTTACTGGATGCAGAGGTGGAGACCCTACGTACAGTTTCAGA
AAGACACTATGAACAAAATAGCCAGGACGGGACCGTTCAGCTACAGAGAC
GAGACGCCTTCCATCACCCTGACCGCCGGGTACAAGTTTCATTTTAACTG
GGGGGGCGACTCTATATTTCCACAGATTATTAAAAACCCCTGCCCAGACA
GCGGGGTACGACCTTCATCCAGTAGAGAGCGTCGCTCAGTACAAGTCGTT
AGCCCGCTCACAATGGGGCCAGAGTACATATTCCACCGGTGGGACTGGCG
ACGGGGGTTCTTTAATCAAAAAGCTCTCAAAAGAATGCTTGAAAAATCAA
TTAATGATGGAGAGTATCCAACAGGCCCAAAGGTCCCTCGATGGTTTCCC
CCACTCGACAACCAAGAGCAAGAAGGCGCCTCAGGTTCAGAGGAGACAAG
GTCGCAGTCCTCGCAAGAAGAAGCCGCTCAAGAAGCCCTCCAAGAAGTCC
AAGAGGCGTCGCTACAGCAGCACCTCCTCCAGCAGTACCGAGAGCAGCGA
CGGATCGGAAAGCAACTCCAACTCGTCATGCTGCAGCTCACCAAGACGCA
GAGCAACCTGCACATAAACCCCCGTGTTCTTGGCCATGCATAAATAAAGT
CTACATGTTTCCCCCCGACAAGCCCATGCCCATACACGGGTACCACGGGT
GGGAGACGGAGTACCAGGCCTGCAAGGCCTTCAACAGGCCCCCCAGAAAC
TACCTTTCAGACAAACCCATCTACCCTTGGCTCCCTCGCCCCGAACCCGA
AATAATAGTGAGCTTTAGGTTCGGTTTCAAATAAACAAGGCCGCAAATAA
ACAAGGCCGTGGGAGTTTCACTGGTCGGTGTCTACCTCTTAAGGTCACTA
AGCACTCCGAGCGTTAGCGAGGAGTGCGACCCTTCCCCCTGGTGCCACGC
CCTCGGCGGCCGCGCGCTACGCCTNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNTGAATCAGTAACGAAAGTGAGTGG
GGCCAGACTTCGCCATAAGGCCTTTATCTTCTTGCCATTGGTCCGTGTGG
GGAGTCGCCATAGGCTTCGGGCTCGGTTTTAGGCCTTCCGGACTACAAAA
ACCGCCATTTTAGTGACGTCACGGCGGCCATTTTAAGTAAGCATGGCGGG
CGGTGACGTACAAGTTGAAAGGTCACCGCGCTTCCGTGTTTACTCAAAAT
GGTGGCCAACTGCTTCCGGGTCAAAGGTCGGCGGCCACGTCATAAGTCAC
GTGGGAGGGCTGCGTCACAAACACGGAAGTGGCTGTCCCACGTGACTTGT
CACGTGATTGCTACGTCACGGCCGCCATTTTAGTTCACAAAATGGCGGAC
TTC (SEQ ID NO: 862)

Annotations:

Putative Domain	*Base range*
TATA Box	12 – 17

Initiator Element	30 – 45
Transcriptional Start Site	40
5' UTR Conserved Domain	100 – 171
ORF2	272 – 637
ORF2/2	272 – 633 ; 2326 – 2767
ORF2/3	272 – 633 ; 2525 – 2984
ORF2t/3	272 – 633 ; 2525 – 2984
ORF1	511 – 2793
ORF1/1	511 – 711 ; 2326 – 2793
ORF1/2	511 – 711 ; 2525 – 2767
Three open-reading frame region	2525 – 2767
Poly(A) Signal	2981 – 2985
Unknown sequence	3125-3176
*Note: Modifications made to maintain reading frames:	
-"C" inserted into ORF2	430
-"N" inserted into ORF1	1842

Table A10. Novel *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 5)

TTV-RTx5b (*Alphatorquevirus* Clade 5)	
ORF2	MSWRPPVHNPNGIERNLWEAFFRMHASACGCGDLVGHLTVLAGRYGAPPR PPAPGAPRPPLIRQLALPAPPADPQQANPQWPGGDGGEDGAGGPAAGGAVA DAEYQEDELNALFDAVEQEE (SEQ ID NO: 863)
ORF2/2	MSWRPPVHNPNGIERNLWEAFFRMHASACGCGDLVGHLTVLAGRYGAPPR PPAPGAPRPPLIRQLALPAPPADPQQANPQWPGGDGGEDGAGGPAAGGAVA DAEYQEDELNALFDAVEQEELLKTPAQTAGYDLHPVESVAQYKSLARSQW GQSTYSTGGTGDGGSLIKKLSKECLKNQLMMESIQQAQRSLDGFPHSTTKSK KAPQVQRRQGRSPRKKKPLKKPSKKSKRRRYSSTSSSSTESSDGSESNSNSSC CSSPRRRATCT (SEQ ID NO: 864)
ORF2/3	MSWRPPVHNPNGIERNLWEAFFRMHASACGCGDLVGHLTVLAGRYGAPPR PPAPGAPRPPLIRQLALPAPPADPQQANPQWPGGDGGEDGAGGPAAGGAVA

	DAEYQEDELNALFDAVEQEEPKGPSMVSPTRQPRARRRLRFRGDKVAVLAR RSRSRSPPRSPRGVATAAPPPAVPRAATDRKATPTRHAAAHQDAEQPAHKP PCSWPCINKVYMFPPDKPMPIHGYHGWETEYQACKAFNRPPRNYLSDKPIY PWLPRPEPEIIVSFRFGFK (SEQ ID NO:865)
ORF2t/3	MSWRPPVHNPNGIERNLWEAFFRMHASACGCGDLVGHLTVLAGRPKGPSM VSPTRQPRARRRLRFRGDKVAVLARRSRSRSPPRSPRGVATAAPPPAVPRAA TDRKATPTRHAAAHQDAEQPAHKPPCSWPCINKVYMFPPDKPMPIHGYHG WETEYQACKAFNRPPRNYLSDKPIYPWLPRPEPEIIVSFRFGFK (SEQ ID NO: 866)
ORF1	MAWWGRWRRWRWRPRRWRRRRRRRVPRRRAQRPVRRRRARRVRRRRW GRRRWRRGYRRRLRLRRRRRRKRKIVLTQWNPAKVRRCTIKGVLPMILCG AGRSGFNYGLHSDDYTVQKPLGQNPHGGGMSTVTFSLQVLYDQYQRFMN KWSYSNDQLDLARYFGCTFWFYRHPEVDFVAQFDNVPPMKMDENTAPNT HPSFLLQNKHKVKIPSFKTKPFGKKRVRVTVGPPKLFEDKWYSQHDLCKVP LVSWRLTAADFRFPFCSPQTDNPCYTFQVLHEEYYPVIGTSALENGSNYNSS AITALEKFLYEKCTHYQTFATDTRLNPQRPVSSTNANKTYTPSGSQETIVWG QSDFNLFKKHTDSNYGYCTYCPTNDLATKIKKYRDKRFDWLTNMPVTNTC HINATFARGKIKEWEYHLGWFSNIFIGNLRHNLAFRAAYIDITXTDKGEGNII WFQYLTKPTTEYIEAQAKCSITNIPLYAAFYGYEDYLQRTLGPYQDVETLGII CVKCPYTDPPLVHKSTDKKNWGYVFYDVHFGNGKTPEGLGQVHPYWMQR WRPYVQFQKDTMNKIARTGPFSYRDETPSITLTAGYKFHFNWGGDSIFPQIIK NPCPDSGVRPSSSRERRSVQVVSPLTMGPEYIFHRWDWRRGFFNQKALKRM LEKSINDGEYPTGPKVPRWFPPLDNQEQEGASGSEETRSQSSQEEAAQEALQ EVQEASLQQHLLQQYREQRRIGKQLQLVMLQLTKTQSNLHINPRVLGHA (SEQ ID NO: 867)
ORF1/1	MAWWGRWRRWRWRPRRWRRRRRRRVPRRRAQRPVRRRRARRIIKNPCPD SGVRPSSSRERRSVQVVSPLTMGPEYIFHRWDWRRGFFNQKALKRMLEKSI NDGEYPTGPKVPRWFPPLDNQEQEGASGSEETRSQSSQEEAAQEALQEVQE ASLQQHLLQQYREQRRIGKQLQLVMLQLTKTQSNLHINPRVLGHA (SEQ ID NO: 868)
ORF1/2	MAWWGRWRRWRWRPRRWRRRRRRRVPRRRAQRPVRRRRARRAQRSLDG

	FPHSTTKSKKAPQVQRRQGRSPRKKKPLKKPSKKSKRRRYSSTSSSSTESSD GSESNSNSSCCSSPRRRATCT (SEQ ID NO: 869)

Table A11. Novel *Anellovirus* nucleic acid sequence (*Alphatorquevirus*)

Name	TTV-RTx6
Genus/Clade	*Alphatorquevirus*, Clade 5
Accession Number	SRR3438066

Full Sequence: 3896 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAAACTTCCTCTTTTAATAGGAAACCACAAAATTTGCATTGCCGACCACA
AACGCATATGCAAATTTACTTCCCCAAAAACTCAACCACAAAATTTGCAT
TGCCGCCCACAAACGTCTACTTTAACCACATCCTCTAACATGTTAGAAAC
TCCACCCAACTACTTCATTAGTATACAGCATCACAAGGGAGGAGCCAAAC
AACTATATAACCAAGTGTACTTCCGAATGGCTGAGTTTATGCCGCCAGAC
GGAGACGGGATCGCGACGGAGGAGCGATCGAGCGTCCCGAGGGCGGGTGC
CGGAGGTGAGTTTACACACCGCAGTCAAGGGGCAATTCGGGCTCGGGACT
GGCCGGGCTATGGGCAAGGCTCTTAAAAAAGCCATGTTTCTCGGTCGACC
TTACAGAAAGAAAAGGGCACTGTCACTGCTACGCGTGCGAGCTCCAGAGG
CGAAACCACCTGCTATGAGCTGGAGGCCCCCGGTGCACAACCCTAATGGG
ATCCAGAGAAACCTGTGGGAGGCATTCTTTCGCATGCATGCTGCAGCTTG
TGGTTGTGGCGATCTTGTTGGCCATATTACTGTACTGGCTGGTCGGTATG
GTGCTCCTCCTCGTCCCCCGGCCCCCGGGGCTCCCAGACCACCGCTGATA
CGCCAGCTGGCCCTTCCGGCGCCCCCCGCCGATCCTCAACAGGCTAACCC
ACAATGGCCTGGTGGGGACGGTGGAGAAGATGGCGCTGGAGGCCCCGCCG
CTGGCGGCGCCGTCGCAGACGCCGAGTACCAAGAAGACGAGCTCAACGCC
CTGTTCGACGCCGTCGAGCAAGAAGAGTAAGGAGGAGGCGATGGGGGAGG
CGGAGGTGGAGACGGGGGTACAGACGCAGACTAAGACTGAGACGCAGACG
CAGACGAAAGAAAATAAGACTGACTCAGTGGAACCCAGCCAAAGTCAGGA
GATGTACTATTAAGGGGGTGCTACCCATGATCTTATGCGGCGCCGGCCGC
TCGGGGTTTAACTATGGACTGCACAGCGACGACTACACGGTGCAGAAACC
CCTGGGGCAGAACCCCCACGGGGGCGGCATGAGCACAGTAACTTTTAGCC
TACAAGTACTATTTGACCAGTACCAGAGGTTTATGAACCGGTGGTCGTAC
TCCAACGACCAGCTAGACCTCGCCAGGTACTTTGGCTGCACCTTCTACTT
TTACAGACACCCTGAAATTGACTTTGTAGCTCAGTATGACAATGTACCCC
CAATGAAAATGGACGAGAACACGGCNCCTAACACTCACCCCTCTTTTCTA
CTACAAAACAAACGCAAAATTAAAATCCCCAGCTTTAAAACCAAGCCATT
TGGCAGAAAAAGAGTAAAAGTAACAGTGGGGCCCCCCAAACTGTTTGAAG
ATAAATGGTACAGCCAGCATGACTTGTGTAAGGTGCCCCTAGTCAGTTGG
CGGTTAACCGCATGTGACTTCAGGTTTCCGTTCTGCTCACCACTAACTGA
CAACCCTTGCTACACCTTCCAGGTATTGCATGAAAACTATTACCCAGTCA
TAGGCACTTCCTCTTTAGAAAACGGTACAAACTACAATAACACTGCTATA
ACTACCCTTGAGACATGGCTATATGGAAAATGCACACACTATCAAACATT
TGCCACAGACACCAGACTTAATCCACAGAGACCTGTATCTTCAAGTAATG
CAAATGAAACTTATACTCCTAGTGGTTCTAAAGAATCAATAATATGGGGA
CAGTCTGACTGGGCAAACTTTAAAAAGAACACAGACAGCAACTATGGCTA
CTGTTCCTACTGCCCCTCAAATGGCACTAACGGAACAGTAGATAAAATTA
AAAAATACAGAGACCAAAGATTTAGATGGCTTACAGAAATGCCAGTACCT
```

AACACCTGTCACATACATGCCACCTTCGCCCGAGGCACTATTAAATACTG
GGAGTACCACCTAGGCTGGTACTCAAACATATTTATTGGCAACCTCAGAC
ACAACTTAGCCTTCAGACCAGCCTACATAGACATTACCTACAATCCCATC
ACTGACAAAGGAGAGGGCAACATTATCTGGTTCCAGTACCTCACTAAGCC
CACCACAGAATACATAGAAACCCAGGCAAAATGCACCATTACTAACATTC
CCCTTTATGCTGCTTTCTATGGCTACGAAGACTACCTCCAGAGAACACTA
GGCCCCTACCAAGATGTAGAAACCCTAGGCATAATCTGTGTTAAATGTCC
CTACACAGATCCCCCTCTAGTTCACAAAGACAAAAGTAAAACCAACTGGG
GCTACGTATTCTACGACGCCCACTTTGGCAACGGAAAGACCCCAGAGGGA
CTAGGCCAAGTACACCCTTACTGGATGCAGAGATGGAGACCCTATGTACA
GTTTCAAAAAGACACCATGCACAAAATATCCAGAACGGGACCCTTCAGCT
ACAGAGACGACACGCCTTCCATCACCCTCACTGCCGAATACAAGTTTCGT
TTTAACTGGGGGGGCGACTCTATATTTCCACAGATTATTAAAAACCCCTG
CCCAGACACCGGGGTTCGACCTTCAACCGGTAGAGACCGTCGCTCAGTAC
AAGTCGTTAGCCCGCTCACAATGGGACCCCAGTTTATATTCCACTCATGG
GACTGGAGACGGGGGTTCTTTAATCAAAAAACTCTCAAAAGAATGCTTGA
AAAACCAGTTAATGATGGAGAATATCCAACAGGCCCAAAGGTGCCTCGAT
GGTTTCCCCCACTCGACAACCAAGAGCAAGAAGGCGTCTCAGATACAGAG
ACGACAACCTCGCAGTCCTCGCAAGAAGAAGCCGCTCAAGAAGCCCTCCA
AGAAGTCCAAGAGGCGTCGCTACAGCAGCACCTCCTCCAGCAGTACCGAG
AGCAGCGAAGAATCGGAAAGCAACTCCAACTCGTCATGCTCCAACTCACC
AAGACGCAGAGCAACCTGCACATAAATCCCCGTGTCCTTGGCCATGCATA
AATAAAGTGTACATGTTTCCCCCCGAAAAGCCAATGCCCATACACGGCTA
CCACGGGTGGGAGACAGAGTATCAGGCCTGCAAGGCCTTTGACAGGCCCC
CTAGAAACTACCTATCAGACAAACCCATCTACCCCTGGCTTCCCCGCTCC
CAACCAGAATTTAAAGTGAGTTTTAAGCTTGGCTGTCAATAAACAAGNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNGTTTACACAAAATGGTGGCCAAGTCCTTCCGGGTGAAAGGTCGGC
GCCTACGTCATAAGTCACGTGGGGAGGGCTGCGTCACAACCAGGAAGCAA
TCCTCACCACGTGATTTGTCACGTGATCGCTACGTCACGGCCGCCATTTT
AGTTTACAAAATGGCGGACTTCCTTCCTCTTTTTCAAAAATAACGGCCCT
GCGGCGGCGCGCGCGCTGCGCGCGCGCGCCGGGGGCTGCCGCCCCA (SEQ ID NO: 870)

Annotations:

Putative Domain	*Base range*
TATA Box	206 – 210
Initiator Element	224 – 239
Transcriptional Start Site	234
5’ UTR Conserved Domain	294 – 364
ORF2	465 – 830
ORF2/2	465 – 826 ; 2534 – 2975

ORF2/3	465 – 826; 2721 – 3192
ORF2t/3	465 – 595 ; 2721 – 3192
ORF1	704 – 3001
ORF1/1	704 – 826; 2534 – 3001
ORF1/2	704 – 826; 2721 – 2975
Three open-reading frame region	2721 – 2975
Poly(A) Signal	3189 – 3193
Unknown sequence	3198 – 3655
GC-rich region, or a portion thereof**	3844 – 3895

Table A12. Novel *Anellovirus* amino acid sequences (*Alphatorquevirus,* Clade 5)

TTV-RTx6 (*Alphatorquevirus* Clade 5)	
ORF2	MSWRPPVHNPNGIQRNLWEAFFRMHAAACGCGDLVGHITVLAGRYGAPPR PPAPGAPRPPLIRQLALPAPPADPQQANPQWPGGDGGEDGAGGPAAGGAVA DAEYQEDELNALFDAVEQEE (SEQ ID NO: 871)
ORF2/2	MSWRPPVHNPNGIQRNLWEAFFRMHAAACGCGDLVGHITVLAGRYGAPPR PPAPGAPRPPLIRQLALPAPPADPQQANPQWPGGDGGEDGAGGPAAGGAVA DAEYQEDELNALFDAVEQEELLKTPAQTPGFDLQPVETVAQYKSLARSQW DPSLYSTHGTGDGGSLIKKLSKECLKNQLMMENIQQAQRCLDGFPHSTTKS KKASQIQRRQPRSPRKKKPLKKPSKKSKRRRYSSTSSSSTESSEESESNSNSSC SNSPRRRATCT (SEQ ID NO: 872)
ORF2/3	MSWRPPVHNPNGIQRNLWEAFFRMHAAACGCGDLVGHITVLAGRYGAPPR PPAPGAPRPPLIRQLALPAPPADPQQANPQWPGGDGGEDGAGGPAAGGAVA DAEYQEDELNALFDAVEQEEISNRPKGASMVSPTRQPRARRRLRYRDDNLA VLARRSRSRSPPRSPRGVATAAPPPAVPRAAKNRKATPTRHAPTHQDAEQP AHKSPCPWPCINKVYMFPPEKPMPIHGYHGWETEYQACKAFDRPPRNYLSD KPIYPWLPRSQPEFKVSFKLGCQ (SEQ ID NO: 873)
ORF2t/3	MSWRPPVHNPNGIQRNLWEAFFRMHAAACGCGDLVGHITVLAGRISNRPK GASMVSPTRQPRARRRLRYRDDNLAVLARRSRSRSPPRSPRGVATAAPPPA VPRAAKNRKATPTRHAPTHQDAEQPAHKSPCPWPCINKVYMFPPEKPMPIH GYHGWETEYQACKAFDRPPRNYLSDKPIYPWLPRSQPEFKVSFKLGCQ (SEQ ID NO: 874)

ORF1	MAWWGRWRRWRWRPRRWRRRRRRRVPRRRAQRPVRRRRARRVRRRRW GRRRWRRGYRRRLRLRRRRRRKKIRLTQWNPAKVRRCTIKGVLPMILCGA GRSGFNYGLHSDDYTVQKPLGQNPHGGGMSTVTFSLQVLFDQYQRFMNR WSYSNDQLDLARYFGCTFYFYRHPEIDFVAQYDNVPPMKMDENTAPNTHP SFLLQNKRKIKIPSFKTKPFGRKRVKVTVGPPKLFEDKWYSQHDLCKVPLVS WRLTACDFRFPFCSPLTDNPCYTFQVLHENYYPVIGTSSLENGTNYNNTAIT TLETWLYGKCTHYQTFATDTRLNPQRPVSSSNANETYTPSGSKESIIWGQSD WANFKKNTDSNYGYCSYCPSNGTNGTVDKIKKYRDQRFRWLTEMPVPNTC HIHATFARGTIKYWEYHLGWYSNIFIGNLRHNLAFRPAYIDITYNPITDKGEG NIIWFQYLTKPTTEYIETQAKCTITNIPLYAAFYGYEDYLQRTLGPYQDVETL GIICVKCPYTDPPLVHKDKSKTNWGYVFYDAHFGNGKTPEGLGQVHPYWM QRWRPYVQFQKDTMHKISRTGPFSYRDDTPSITLTAEYKFRFNWGGDSIFPQ IIKNPCPDTGVRPSTGRDRRSVQVVSPLTMGPQFIFHSWDWRRGFFNQKTLK RMLEKPVNDGEYPTGPKVPRWFPPLDNQEQEGVSDTETTTSQSSQEEAAQE ALQEVQEASLQQHLLQQYREQRRIGKQLQLVMLQLTKTQSNLHINPRVLGH A (SEQ ID NO: 875)
ORF1/1	MAWWGRWRRWRWRPRRWRRRRRRRVPRRRAQRPVRRRRARRIIKNPCPD TGVRPSTGRDRRSVQVVSPLTMGPQFIFHSWDWRRGFFNQKTLKRMLEKPV NDGEYPTGPKVPRWFPPLDNQEQEGVSDTETTTSQSSQEEAAQEALQEVQE ASLQQHLLQQYREQRRIGKQLQLVMLQLTKTQSNLHINPRVLGHA (SEQ ID NO: 876)
ORF1/2	MAWWGRWRRWRWRPRRWRRRRRRRVPRRRAQRPVRRRRARRNIQQAQR CLDGFPHSTTKSKKASQIQRRQPRSPRKKKPLKKPSKKSKRRRYSSTSSSSTE SSEESESNSNSSCSNSPRRRATCT (SEQ ID NO: 877)

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

Name	TTV-CT30F
Genus/Clade	*Alphatorquevirus*, Clade 1
Accession Number	AB064597.1

Full Sequence: 3570 bp

1 10 20 30 40 50

ATTTTGTGCAGCCCGCCAATTCTCGTTCAAACAGGCCAATCAGGAGGCTC
TACGTACACTTCCTGGGGTGTGTCTTCGAAGAGTATATAAGCAGAGGCGG
TGACGAATGGTAGAGTTTTTCCTGGCCCGTCCGCGGCGAGAGCGCGAGCG
GAGCGAGCGATCGAGCGTCCCGTGGGCGGGTGCCGTAGGTGAGTTTACAC
ACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTATGGGCAA
GATTCTTAAAAAATTCCCCCGATCCCTCTGTCGCCAGGACATAAAAACAT
GCCGTGGAGACCGCCGGTGCATAGTGTCCAGGGGCGAGAGGATCAGTGGT
TCGCGAGCTTTTTTCACGGCCACGCTTCATTTTGCGGTTGCGGTGACGCT
GTTGGCCATCTTAATAGCATTGCTCCTCGCTTTCCTCGCGCCGGTCCACC
AAGGCCCCCTCCGGGGCTAGAGCAGCCTAACCCCCCGCAGCAGGGCCCGG
CCGGGCCCGGAGGGCCGCCCGCCATCTTGGCGCTGCCGGCTCCGCCCGCG
GAGCCTGACGACCCGCAGCCACGGCGTGGTGGTGGGGACGGTGGCGCCGC
CGCTGGCGCCGCAGGCGACCGTGGAGACCGAGACTACGACGAAGAAGAGC
TAGACGAGCTTTTCCGCGCCGCCGCCGAAGACGATTTGTAAGTAGGAGAT
GGCGCCGGCCTTACAGGCGCAGGAGGAGACGCGGGCGACGCAGACGCAGA
CGCAGACGCAGACATAAGCCCACCCTAGTACTCAGACAGTGGCAACCTGA
CGTTATCAGACACTGTAAGATAACAGGACGGATGCCCCTCATTATCTGTG
GAAAGGGGTCCACCCAGTTCAACTACATCACCCACGCGGACGACATCACC
CCCAGGGGAGCCTCCTACGGGGGCAACTTCACAAACATGACTTTCTCCCT
GGAGGCAATATACGAACAGTTTCTGTACCACAGAAACAGGTGGTCAGCCT
CCAACCACGACCTCGAACTCTGCAGATACAAGGGTACCACCCTAAAACTG
TACAGGCACCCAGATGTAGACTACATAGTCACCTACAGCAGAACGGGACC
CTTTGAGATCAGCCACATGACCTACCTCAGCACTCACCCCCTTCTCATGC
TGCTAAACAAACACCACATAGTGGTGCCCAGCCTAAAGACTAAGCCCAGG
GGCAGAAAGGCCATAAAAGTCAGAATAAGACCCCCCAAACTCATGAACAA
CAAGTGGTACTTCACCAGAGACTTCTGTAACATAGGCCTCTTCCAGCTCT
GGGCCACAGGCTTAGAACTCAGAAACCCCTGGCTCAGAATGAGCACCCTG
AGCCCCTGCATAGGCTTCAATGTCCTTAAAAACAGCATTTACACAAACCT
CAGCAACCTACCTCAGCACAGAGAAGACAGACTTAACATTATTAACAACA
CATTACACCCACATGACATAACAGGACCAAACAATAAAAAATGGCAGTAC
ACATATACCAAACTCATGGCCCCCATTTACTATTCAGCAAACAGGGCCAG
CACCTATGACTTACTACGAGAGTATGGCCTCTACAGTCCATACTACCTAA
ACCCCACAAGGATAAACCTTGACTGGATGACCCCCTACACACACGTCAGG
TACAATCCACTAGTAGACAAGGGCTTCGGAAACAGAATATACATACAGTG
GTGCTCAGAGGCAGATGTAAGCTACAACAGGACTAAATCCAAGTGTCTCT
TACAAGACATGCCCCTGTTTTTCATGTGCTATGGCTACATAGACTGGGCA
ATTAAAAACACAGGGGTCTCCTCACTAGCGAGAGACGCCAGAATCTGCAT
CAGGTGTCCCTACACAGAGCCACAGCTGGTGGGCTCCACAGAAGACATAG
GGTTCGTACCCATCACAGAGACCTTCATGAGGGGCGACATGCCGGTACTT
GCACCATACATACCGTTGAGCTGGTTTTGCAAGTGGTATCCCAACATAGC
TCACCAGAAGGAAGTACTTGAGGCAATCATTTCCTGCAGCCCCTTCATGC
CCCGTGACCAGGGCATGAACGGTTGGGATATTACAATAGGTTACAAAATG
GACTTCTTATGGGGCGGTTCCCCTCTCCCCTCACAGCCAATCGACGACCC
CTGCCAGCAGGGAACCCACCCGATTCCCGACCCCGATAAGCACCCTCGCC
TCCTACAAGTGTCGAACCCGAAACTGCTCGGACCGAGGACAGTGTTCCAC
AAGTGGGACATCAGACGTGGGCAGTTTAGCAAAAGAAGTATTAAAAGAGT
GTCAGAATACTCATCGGATGATGAATCTCTTGCGCCAGGTCTCCCATCAA
AGCGAAACAAGCTCGACTCGGCCTTCAGAGGAGAAAACCCAGAGCAAAAA
GAATGCTATTCTCTCCTCAAAGCACTCGAGGAAGAAGAGACCCCAGAAGA
AGAAGAACCAGCACCCCAAGAAAAAGCCCAGAAAGAGGAGCTACTCCACC
AGCTCCAGCTCCAGAGACGCCACCAGCGAGTCCTCAGACGAGGGCTCAAG
CTCGTCTTTACAGACATCCTCCGACTCCGCCAGGGAGTCCACTGGAACCC
CGAGCTCACATAGAGCCCCCACCTTACATACCAGACCTACTTTTTCCCAA
TACTGGTAAAAAAAAAAAAATTCTCTCCCTTCGACTGGGAAACGGAGGCCC
AGCTAGCAGGGATATTCAAGCGTCCTATGCGCTTCTATCCCTCAGACACC
CCTCACTACCCGTGGTTACCCCCCAAGCGCGATATCCCGAAAATATGTAA

CATAAACTTCAAAATAAAGCTGCAAGAGTGAGTGATTCGAGGCCCTCCTC
TGTTCACTTAGCGGTGTCTACCTCTTAAAGTCACCAAGCACTCCGAGCGT
CAGCGAGGAGTGCGACCCTCCACCAAGGGGCAACTTCCTCGGGGTCCGGC
GCTACGCGCTTCGCGCTGCGCCGGACGCCTCGGACCCCCCCCCGACCCGA
ATCGCTCGCGCGATTCGGACCTGCGGCCTCGGGGGGGGTCGGGGGCTTTA
CTAAACAGACTCCGAGTTGCCACTGGACTCAGGAGCTGTGAATCAGTAAC
GAAAGTGAGTGGGGCCAGACTTCGCCATAGGGCCTTTAACTTGGGGTCGT
CTGTCGGTGGCTTCCGGGTCCGCCTGGGCGCCGCCATTTTAGCTTTAGAC
GCCATTTTAGGCCCTCGCGGGCACCCGTAGGCGCGTTTTAATGACGTCAC
GGCAGCCATTTTGTCGTGACGTTTGAGACACGTGATGGGGGCGTGCCTAA
ACCCGGAAGCATCCCTGGTCACGTGACTCTGACGTCACGGCGGCCATTTT
GTGCTGTCCGCCATCTTGTGACTTCCTTCCGCTTTTTCAAAAAAAAAGAG
GAAGTATGACAGTAGCGGCGGGGGGGCGGCCGCGTTCGCGCGCCGCCCAC
CAGGGGGTGCTGCGCGCCCCCCCCCGCGCATGCGCGGGGCCCCCCCCCGG
GGGGGCTCCGCCCCCCCGGCCCCCCCCCGTGCTAAACCCACCGCGCATGC
GCGACCACGCCCCCGCCGCC (SEQ ID NO: 1)

Annotations:

Putative Domain	*Base range*
TATA Box	84 – 90
Cap Site	107 – 114
Transcriptional Start Site	114
5' UTR Conserved Domain	177 – 247
ORF2	299 – 691
ORF2/2	299 – 687 ; 2137 – 2659
ORF2/3	299 – 687 ; 2339 – 2831
ORF2t/3	299 – 348 ; 2339 – 2831
ORF1	571 – 2613
ORF1/1	571 – 687 ; 2137 – 2613
ORF1/2	571 – 687 ; 2339 – 2659
Three open-reading frame region	2325 – 2610
Poly(A) Signal	2813 – 2818
GC-rich region	3415 – 3570

Table 2. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 1)

TTV-CT30F (*Alphatorquevirus* Clade 1)	
ORF2	MPWRPPVHSVQGREDQWFASFFHGHASFCGCGDAVGHLNSIAPRFPRAGPPRPPPG LEQPNPPQQGPAGPGGPPAILALPAPPAEPDDPQPRRGGGDGGAAAGAAGDRGDRD YDEEELDELFRAAAEDDL (SEQ ID NO: 2)
ORF2/2	MPWRPPVHSVQGREDQWFASFFHGHASFCGCGDAVGHLNSIAPRFPRAGPPRPPPG

	LEQPNPPQQGPAGPGGPPAILALPAPPAEPDDPQPRRGGGDGGAAAGAAGDRGDRD YDEEELDELFRAAAEDDFQSTTPASREPTRFPTPISTLASYKCRTRNCSDRGQCSTSG TSDVGSLAKEVLKECQNTHRMMNLLRQVSHQSETSSTRPSEEKTQSKKNAILSSKH SRKKRPQKKKNQHPKKKPRKRSYSTSSSSRDATSESSDEGSSSSLQTSSDSARESTGT PSSHRAPTLHTRPTFSQYW (SEQ ID NO: 3)
ORF2/3	MPWRPPVHSVQGREDQWFASFFHGHASFCGCGDAVGHLNSIAPRFPRAGPPRPPPG LEQPNPPQQGPAGPGGPPAILALPAPPAEPDDPQPRRGGGDGGAAAGAAGDRGDRD YDEEELDELFRAAAEDDLSPIKAKQARLGLQRRKPRAKRMLFSPQSTRGRRDPRRR RTSTPRKSPERGATPPAPAPETPPASPQTRAQARLYRHPPTPPGSPLEPRAHIEPPPYIP DLLFPNTGKKKKFSPFDWETEAQLAGIFKRPMRFYPSDTPHYPWLPPKRDIPKICNIN FKIKLQE (SEQ ID NO: 4)
ORF2t/3	MPWRPPVHSVQGREDQWSPIKAKQARLGLQRRKPRAKRMLFSPQSTRGRRDPRRR RTSTPRKSPERGATPPAPAPETPPASPQTRAQARLYRHPPTPPGSPLEPRAHIEPPPYIP DLLFPNTGKKKKFSPFDWETEAQLAGIFKRPMRFYPSDTPHYPWLPPKRDIPKICNIN FKIKLQE (SEQ ID NO: 5)
ORF1	TAWWWGRWRRRWRRRRPWRPRLRRRRARRAFPRRRRRRFVSRRWRRPYRRRRR RGRRRRRRRRRHKPTLVLRQWQPDVIRHCKITGRMPLIICGKGSTQFNYITHADDIT PRGASYGGNFTNMTFSLEAIYEQFLYHRNRWSASNHDLELCRYKGTTLKLYRHPD VDYIVTYSRTGPFEISHMTYLSTHPLLMLLNKHHIVVPSLKTKPRGRKAIKVRIRPPK LMNNKWYFTRDFCNIGLFQLWATGLELRNPWLRMSTLSPCIGFNVLKNSIYTNLSN LPQHREDRLNIIINNTLHPHDITGPNNKKWQYTYTKLMAPIYYSANRASTYDLLREY GLYSPYYLNPTRINLDWMTPYTHVRYNPLVDKGFGNRIYIQWCSEADVSYNRTKSK CLLQDMPLFFMCYGYIDWAIKNTGVSSLARDARICIRCPYTEPQLVGSTEDIGFVPIT ETFMRGDMPVLAPYIPLSWFCKWYPNIAHQKEVLEAIISCSPFMPRDQGMNGWDITI GYKMDFLWGGSPLPSQPIDDPCQQGTHPIPDPDKHPRLLQVSNPKLLGPRTVFHKW DIRRGQFSKRSIKRVSEYSSDDESLAPGLPSKRNKLDSAFRGENPEQKECYSLLKALE EEETPEEEEPAPQEKAQKEELLHQLQLQRRHQRVLRRGLKLVFTDILRLRQGVHWN PELT (SEQ ID NO: 6)
ORF1/1	TAWWWGRWRRRWRRRRPWRPRLRRRRARRAFPRRRRRRFPIDDPCQQGTHPIPDP DKHPRLLQVSNPKLLGPRTVFHKWDIRRGQFSKRSIKRVSEYSSDDESLAPGLPSKR NKLDSAFRGENPEQKECYSLLKALEEEETPEEEEPAPQEKAQKEELLHQLQLQRRH QRVLRRGLKLVFTDILRLRQGVHWNPELT (SEQ ID NO: 7)
ORF1/2	TAWWWGRWRRRWRRRRPWRPRLRRRRARRAFPRRRRRRFVSHQSETSSTRPSEE

	KTQSKKNAILSSKHSRKKRPQKKKNQHPKKKPRKRSYSTSSSSRDATSESSDEGSSS SLQTSSDSARESTGTPSSHRAPTLHTRPTFSQYW (SEQ ID NO: 8)

Table 3. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus*, Clade 2)

Name	TTV-P13-1
Genus/Clade	*Alphatorquevirus*, Clade 2
Accession Number	KT163896.1

Full Sequence: 3451 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
AATTTTGCTAAACAGACTCCGAGGTGCTCTTGGACACTGAGTGGGCGTAC
AGCAACGAAAGTGAGTGGGGCCAGACTTCGCCATAAGGCCTTTATCTTCG
GGTCTACATCATAATATAAAGATGTGCACTTCCGAATGGCTGAGTTTTTC
ACGCCATTCCGCAGCGGTGGAGCAGCGCAGCCACGACCCCCGCGTCCCGA
GGGCGGGTGCCGGAGGTGAGTTTACACACCGCAGTCAAGGGGCAATTCGG
GCTCGGGACTGGCCGGGCCCGGGCAAGGCTCTTAAAGCGAAACCATGTTC
CTCGGCAGGCCCTACCGCCACAGAAAGCGGCACCAGGCCGGCAAGAAAGG
GCCACTGCCACTGCCAAATCTGCAACCTGCACAGGAGAAACGGGCTGGTG
GTCCGTCCTTGATGGCCTCCGGACGCAGGGGATGGATGCCCCCGGACCTG
ACGGTCCAGGAGAGGGAGGATGCCTGGTGGACCAGCTTCTGCGCTAGCCA
CCGCAGCTTTTGTAGCTGCGACGATCCTGTGGGCCATATTAATACTCTCG
CCCGCGATAATAGTCCTCTGGCCCAGACTCCTACTACAACTTCAGGCCAG
GGGCCGCCGCCGCCGCCTACGCCTCCGCGGACGCCGGGGCCGCGCCCTGG
GTCTGCTCCGGACCAGGGGGGAAGGATCAGGGCCTCCTGGACCTACCCCC
TAGCCCCCGGAGGTCCCGGTAGCACGCCATGGCCTACTGGTGGGGCCGGA
GACGCCGGTGGCGCCGCTGGAGGAGGCGCCGGCGTCCTCTCCGCCGCCGC
CGGCGGTGGCGGAGAAGGCGACGCTGGCCCAGAAGGCGCCGGTGGAGGCG
AAGGAGACGACGTGCGAGACCTGCTCGCCGCTATCGAAGGAGACGTGGGC
GCAGACGGGTAAGGAGACGCCGTCGCCCCCAGAAACTAGTACTGACTCAG
TGGAATCCCCAGACTGTGAGAAAGTGTGTTATTAGGGGGTTTCTGCCCCT
GTTCTTCTGCGGACAGGGGGCCTACCACAGAAACTTTACAGACCACTATG
ACGATGTGTTCCCCAAGGGACCCAGCGGAGGTGGGCACGGGAGCATGGTG
TTCAACCTGTCCTTTCTGTACCAAGAGTTTAAGAAGCACCACAATAAGTG
GTCGCGCAGCAACCTGGACTTTGACTTAGTGAGATACAAGGGCACAGTGA
TAAAGCTGTACAGACACCAGGACTTTGACTACATAGTGTGGATAAGCAGG
ACCCCTCCCTTCCAGGAGAGCCTGCTCACAGTAATGACCCACCAGCCCAG
CGTCATGCTGCAGGCAAAAAAGTGCATAATAGTAAAGAGCTACAGGACCC
ACCCGGGGGGCAAACCCTATGTAACTGCAAAAGTTAGGCCCCCCAGACTC
CTAACTGACAAGTGGTACTTCCAGTCAGACTTCTGCAACGTTCCGCTTTT
TAGCCTACAGTTTGCCCTTGCGGAACTGCGGTTTCCGATCTGCTCACCAC
AAACTGACACCAATTGCATTAACTTCCTGGTGTTAGATGACATCTACTAC
AAGTTTCTAGATAATAAGCCTAAACAGAGTTCAGACCCTAATGACGAAAA
CAGAATAAAATTCTGGCACGGCCTATGGTCCACTATGAGATATTTAAACA
CCACCTACATAAACACACTGTTTCCAGGCACAGACAGTCTAGTGGCCGCC
AAAGATACTGACAATAGTGTAAATAAATACCCCAGCACAGCCACTAAACA
GCCCTACAAAGACAGTCAGTACATGCAAAATATATGGAATACATCAAAAA
TACATGCCTTATATACGTGGGTAGCAGAGACAAACTACAAAAGACTGCAG
GCCTACTACACACAGACCTACGGAGGCTACCAGAGACAATTTTTCACAGG
AAAACAGTACTGGGACTACAGAGTAGGCATGTTTAGTCCAGCCTTCCTGA
GTCCCAGCAGACTAAATCCCCAGAACCCAGGGGCATACACAGAGGTCTCC
```

TACAACCCCTGGACAGACGAGGGCACGGGCAACGTAGTGTGCCTGCAGTA
TCTGACTAAAGAGACCTCAGACTACAAACCAGGTGGTGGGAGCAAGTTCT
GCATAGAAGGTGTGCCTCTATGGGCAGCGCTGGTGGGATACGTAGACATG
TGTAAAAAAGAGGGCAAGGACCCGGGCATCAGACTAAACTGTCTCCTGTT
AGTCAAGTGTCCCTATACAAAGCCTCAGCTGTATGACAAAAAAAACCCCG
AGAAACTGTTTGTACCTTACTCCTATAACTTTGGGCACGGCAAGATGCCG
GGGGGAGACAAATACATACCCATAGAGTTCAAAGACAGGTGGTACCCCTG
CCTGCTCCACCAAGAGGAGTGGATAGAGGACATTGTCAGGTCGGGACCCT
TCGTTCCAAAAGACATGCCCAGCAGCGTCACCTGCATGATGAGGTACAGC
TCTCTTTTTAACTGGGGCGGTAATATAATCCAAGAACAGGCCGTGGAAGA
CCCCTGTAAGAAAGGCACCTTCGTCGTTCCCGGAACCAGTGGCATCGCTC
GCATACTACAAGTCAGCAACCCGGCCAAGCAGACCCCCACGACAACCTGG
CACTCGTGGGACTGGAGACGATCCCTCTTTACAGAGACGGGTCTTAAAAG
AATGCGCGAACAACAACCATATGATGAACTGTCTTATACGGGCCCTAAAA
AGCCAAAACTGTCCCTTCCCGCAGGGCCCGCCGTCCCCGGTGCCGCCGTC
GCCTCCTCCTGGTGGGAAACAAAACAGGTCACCTCGCCAGACGTCAGCGA
GACGGAGACCGAAGCAGAAGCCCACCAAGAGGAAGAGACGGAGCCGGAGG
AGGGAGTCCAGCTCCAGCAGCTGTGGGAGCAGCAACTCCTGCAAAAGCGA
CAGCTGGGAGTCGTGTTCCAGCAACTCCTCCGACTCAGACAGGGGGCGGA
GATCCACCCGGGCCTCGTATAATTCCTGGGCCCCAGAACCCGTACCTGCT
TTTCCCGGAGCAGGCCCCTCCAAAAGTGCCTATTTTTGACCCCTTTGGTC
AGAAAACAGAGCTAGAGCTGTGCGGCTGCTTCGACAGGCCGCCCAGGAAC
AACCCCTACGACCACCCCTTCTACCCCTGGCTGCCCAAAGAGCCTCCCTC
CTACTACCAGGGCTACAAAGTGTCTTTCAAACTAGGGTTCCACCCAGACA
AGCATGTGTGAACCCCGCCAATAAACCACTGCTGCTACACTGATTCTTAG
GCCGTGGGAGTCTCACTGGTCGGTGTCTACCTCTTAAGGTCACTAAGCAC
TCCGAGCGTTAGCGAGGAGTGCGACCCTACCCCCTGGGCCCACTTCTTCG
GAGCCGCGCGCTACGCCTTCGGCTGCGCGCGGCACCTCAGACCCCCGCTC
GTGCTGACACGCTTGCGCGTGTCAGACCACTTCGGGCTCGCGGGGGTCGG
G (SEQ ID NO: 9)

Annotations:

Putative Domain	*Base range*
TATA Box	112 – 119
Initiator Element	128-148
Transcriptional Start Site	148
5’ UTR Conserved Domain	204 – 273
ORF2	412-912
ORF2/2	412 – 908 ; 2490 – 3039
ORF2/3	412 – 908 ; 2725 – 3208
ORF1	729 – 2972
ORF1/1	729 – 908 ; 2490 – 2972
ORF1/2	729 – 908 ; 2725 – 3039
Three open-reading frame region	2699 – 2969
Poly(A) Signal	3220 - 3225
GC-rich region	3302 – 3541

Table 4. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 2)

TTV-P13-1 (*Alphatorquevirus* Clade 2)	
ORF2	MASGRRGWMPPDLTVQEREDAWWTSFCASHRSFCSCDDPVGHINTLARDN SPLAQTPTTTSGQGPPPPPTPPRTPGPRPGSAPDQGGRIRASWTYPLAPGGPG STPWPTGGAGDAGGAAGGGAGVLSAAAGGGGEGDAGPEGAGGGEGDDV RDLLAAIEGDVGADG (SEQ ID NO: 10)
ORF2/2	MASGRRGWMPPDLTVQEREDAWWTSFCASHRSFCSCDDPVGHINTLARDN SPLAQTPTTTSGQGPPPPPTPPRTPGPRPGSAPDQGGRIRASWTYPLAPGGPG STPWPTGGAGDAGGAAGGGAGVLSAAAGGGGEGDAGPEGAGGGEGDDV RDLLAAIEGDVGADGPWKTPVRKAPSSFPEPVASLAYYKSATRPSRPPRQPG TRGTGDDPSLQRRVLKECANNNHMMNCLIRALKSQNCPFPQGPPSPVPPSPP PGGKQNRSPRQTSARRRPKQKPTKRKRRSRRRESSSSSCGSSNSCKSDSWES CSSNSSDSDRGRRSTRASYNSWAPEPVPAFPGAGPSKSAYF (SEQ ID NO: 11)
ORF2/3	MASGRRGWMPPDLTVQEREDAWWTSFCASHRSFCSCDDPVGHINTLARDN SPLAQTPTTTSGQGPPPPPTPPRTPGPRPGSAPDQGGRIRASWTYPLAPGGPG STPWPTGGAGDAGGAAGGGAGVLSAAAGGGGEGDAGPEGAGGGEGDDV RDLLAAIEGDVGADGARRPRCRRRLLLVGNKTGHLARRQRDGDRSRSPPRG RDGAGGGSPAPAAVGAATPAKATAGSRVPATPPTQTGGGDPPGPRIIPGPQN PYLLFPEQAPPKVPIFDPFGQKTELELCGCFDRPPRNNPYDHPFYPWLPKEPP SYYQGYKVSFKLGFHPDKHV (SEQ ID NO: 12)
ORF1	MAYWWGRRRRWRRWRRRRRPLRRRRRWRRRRRWPRRRRWRRRRRRARP ARRYRRRRGRRRVRRRRRPQKLVLTQWNPQTVRKCVIRGFLPLFFCGQGAY HRNFTDHYDDVFPKGPSGGGHGSMVFNLSFLYQEFKKHHNKWSRSNLDFD LVRYKGTVIKLYRHQDFDYIVWISRTPPFQESLLTVMTHQPSVMLQAKKCII VKSYRTHPGGKPYVTAKVRPPRLLTDKWYFQSDFCNVPLFSLQFALAELRF PICSPQTDTNCINFLVLDDIYYKFLDNKPKQSSDPNDENRIKFWHGLWSTMR YLNTTYINTLFPGTDSLVAAKDTDNSVNKYPSTATKQPYKDSQYMQNIWNT SKIHALYTWVAETNYKRLQAYYTQTYGGYQRQFFTGKQYWDYRVGMFSP AFLSPSRLNPQNPGAYTEVSYNPWTDEGTGNVVCLQYLTKETSDYKPGGGS KFCIEGVPLWAALVGYVDMCKKEGKDPGIRLNCLLLVKCPYTKPQLYDKK

	NPEKLFVPYSYNFGHGKMPGGDKYIPIEFKDRWYPCLLHQEEWIEDIVRSGP FVPKDMPSSVTCMMRYSSLFNWGGNIIQEQAVEDPCKKGTFVVPGTSGIARI LQVSNPAKQTPTTTWHSWDWRRSLFTETGLKRMREQQPYDELSYTGPKKP KLSLPAGPAVPGAAVASSWWETKQVTSPDVSETETEAEAHQEEETEPEEGV QLQQLWEQQLLQKRQLGVVFQQLLRLRQGAEIHPGLV (SEQ ID NO: 13)
ORF1/1	MAYWWGRRRRWRRWRRRRRPLRRRRRWRRRRRWPRRRRWRRRRRRARP ARRYRRRRGRRRAVEDPCKKGTFVVPGTSGIARILQVSNPAKQTPTTTWHS WDWRRSLFTETGLKRMREQQPYDELSYTGPKKPKLSLPAGPAVPGAAVASS WWETKQVTSPDVSETETEAEAHQEEETEPEEGVQLQQLWEQQLLQKRQLG VVFQQLLRLRQGAEIHPGLV (SEQ ID NO: 14)
ORF1/2	MAYWWGRRRRWRRWRRRRRPLRRRRRWRRRRRWPRRRRWRRRRRRARP ARRYRRRRGRRRGPPSPVPPSPPPGGKQNRSPRQTSARRRPKQKPTKRKRRS RRRESSSSSCGSSNSCKSDSWESCSSNSSDSDRGRRSTRASYNSWAPEPVPAF PGAGPSKSAYF (SEQ ID NO: 15)

Table 5. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus,* Clade 3)

Name	TTV-tth8
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 6. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 3)

TTV-tth8 (*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 7. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus*, Clade 4)

Name	TTV-HD20a
Genus/Clade	*Alphatorquevirus*, Clade 4
Accession Number	FR751492.1

Full Sequence: 3878 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
AAATACGTCACTAACCACGTGACTCCCACAGGCCAACCACAGTCTATGTC
GTGCACTTCCTGGGCATGGTCTACGTGATAATATAAAGCGGTGCACTTCC
GAATGGCTGAGTTTTCCACGCCCGTCCGCAGCGAGATCGCGACGTAGGAG
CGATCGAGCGTCCCGAGGGCGGGTGCCGGAGGTGAGTTTACACACCGCAG
TCAAGGGGCAATTCGGGCTCGGGAGGCCGGGCCATGGGCAAGGCTCTTAA
AAAGCTATGTTTCTCGGTAAAATCTACAGGAAGAAAAGGAAACTGCTTCT
GCAGGCTGTGCGTGCTCCGCAGACGCCATCTTCCATGAGCCGCTGCTGGT
GTCCCCCTCGGGGTGATGTCTCCTCCCGCGAGTCTCGATGGTACGAGGCG
GTTCGAGGAAGCCACGATGCTTTTTGTGGCTGTAGTGATCCTATTCTTCA
TCTTTCTCGTCTGGCTGCACGTTTTAACCATCAGGGACCTCCGACGCCCC
CCACGGACGACCGTGCGCCGCAGAATACCCCAGTGAGACGCCTGCTGCCT
CTCCCCAGCTACCCCGGCGAGGGTCCCCAGGCTAGATGGCCTGGTGGGGA
TGGAGGCGCCGCTGGTGGCGACCGAAGAGAAGGTGGAGATGGCGGCGCGC
GCGCCGCCGAAGACGAGTACCAGCCCGAAGACCTAGACGAGCTTTTCGGC
GCTATCGAACAAGAACAGTAAGGAGGAGGCGAAGGGGGAGGCGGAGGGGC
TACCGGCGCCGTTACAGACTGAGACGCTATGCCAGACGCAGGTTCCGACG
CAAAAAGATAGTACTGACTCAGTGGAACCCCCAGACTACCAGAAAATGTA
TAATAAGGGGCATGATGCCAGTACTGTGGGCCGGCATGGGTACGGGGGGC
AGAAACTATGCAGTGAGGTCAGATGACTATGTGGTGAACAAAGGGTTCGG
GGGCTCCTTCGCCACGGAGACCTTCTCCCTGAAGGTTCTCTATGACCAGT
TTCAAAGGGGCTTCAACAGGTGGTCCCACACTAACGAGGACCTAGACCTG
GCCCGCTACAGGGGCTGCAGGTGGACTTTTTACAGACATAAAGACACAGA
CTTTATAGTGTACTTTACAAACAATCCTCCCATGAAGACCAACCAGTTCT
CCGCGCCCCTGACGACCCCCGGCATGCTCATGCGCAGTAAATACAAAGTC
CTCATTCCCAGCTTCCAGACCAGACCCAAGGGTCGCAAAACAGTAACCGT
TAAAATAAGACCCCCCAAACTATTTCAAGACAAGTGGTACACCCAGCAGG
ACCTGTGTTCAGTTCCTCTTGTCCAACTGAACGTGACCGCAGCTGATTTC
ACACATCCGTTCGGCTCACCACTAACTGAAACTCCTTGCGTAGAGTTCCA
GGTGCTGGGTGACTTGTACAATACATGTCTCAATATCGACCTTCCGCAAT
TTAGTGAATTAGGAGAAATAACTAGTGCCTACTCAAAACCAAACTCAAAT
AACCTAAAAGAATTATACAAAGAATTGTTCACAAAAGCCACATCAGGACA
CTACTGGCAGACATTCATAACCAACAGCATGGTCAGAGCACACATAGATG
CAGACAAAGCTAAAGAAGCACAAAGAGCATCCACCACACCCTCATACAAC
AATGACCCCTTCCCCACAATACCTGTTAAATCAGAGTTTGCACAGTGGAA
AAAGAAATTCACAGACACTAGAGACAGCCCCTTTCTTTTTGCCACTTACC
ATCCCGAAGCTATAAAAGACACAATTATGAAAATGAGAGAGAACAACTTT
AAGCTAGAGACAGGACCCAATGACAAGTATGGAGACTACACAGCACAGTA
CCAAGGAAACACACACATGCTAGACTACTACCTTGGCTTTTACAGCCCCA
TATTCCTCTCAGATGGAAGGTCTAACGTAGAATTCTTCACTGCCTACAGA
GACATAGTATACAATCCCTTCTTAGACAAGGCCCAGGGCAACATGGTGTG
GTTTCAGTACCACACAAAGACAGACAACAAGTTTAAAAAACCAGAGTGCC
ACTGGGAAATCAAAGACATGCCCCTGTGGGCCCTCCTAAACGGATATGTA
GACTACTTAGAGACTCAAATACAGTATGGTGACCTCAGTAAAGAAGGGAA
AGTCCTCATCAGGTGTCCCTACACCAAGCCAGCACTAGTAGACCCCAGAG
ACGACACTGCAGGATATGTAGTCTACAACAGAAACTTTGGCAGAGGCAAG
TGGATAGACGGAGGGGGCTACATCCCCCTGCACGAGAGGACAAAATGGTA
CGTGATGCTCAGATACCAGACGGACGTCTTCCATGACATAGTGACCTGTG
GGCCCTGGCAGTACAGAGACGACAACAAAAACAGCCAGCTAGTGGCCAAA
TACCGCTTCAGCTTTATATGGGGAGGTAACACTGTCCACTCTCAGGTCAT
CAGAAACCCGTGCAAAGACAACCAAGTATCCGGTCCCCGTCGACAGCCTA
GGGATATACAAGTCGTTGACCCGCAACGCATCACGCCGCCGTGGGTCCTC
```

CACAGCTTCGACCAGCGAAGAGGCCTCTTTACTGAAACAGCTCTCAGGCG
CCTGCTCCAGGAACCACTACCTGGCGAGTATGCTGTTAGCACCCTCAGGA
CACCCCTCCTCTTTCTACCCTCAGAATACCAGCGAGAAGACGGCGCTGCA
GAAAGCGCCTCAGGTTCACCGGCCAAAAGACCCCGTATCTGGTCAGAAGA
GAGTCAGACGGAGACGATCTCCTCGGAGGAGAACCCGGCGGAGACGACGA
GGGAGCTCCTCCAGCGAAAGCTCCGAGAGCAGCGAGCACTCCAGTTCCAA
CTCCAGCACTTCGCGGTCCAACTCGCCAAGACCCAGGCGAATCTCCACGT
AAACCCCCTGTTATCTTTCCCGCAATGAATAAGGTCTTTCTGTTTCCCCC
AGAGGGTCCCAAGCCCATCCTGGGCAAAGAGGCCTGGCAGGACGAGTACG
AGACCTGCAGGGTCTGGAACAGACCTGCCAGAACCCACCACACAGACACC
CCCTTCTATCCCTGGGCCCCCCACAAGTTCCATGTAAGCTTCAAACTTGG
CTTCCAATAAAATTACTAGGCCGTGGAACTCTCACTGGTCGGTGTCTACC
TCTTAAGGTCACTAAGCACTCCGAGCGTCAGCGAGGAGTGCGACCCTCTA
CCCTGGTGCAACGCCCTCGGCGGCCGCGCGCTACGCCTTCGGCTGCGCGC
GGCACCTCGGACCCCCGCTCGTGCTGACGCGCTCGCGCGCGTCAGACCAC
TTCGGGCTCGCGGGGGTCGGGAATTTTGCTAAACAGACTCCGAGTTGCCA
TTGGACACTGTAGCTGTGAATCAGTAACGAAAGTGAGTGGGGCCAGACTT
CGCCATAGGGCCTTTATCTTCTTGCCATTGGTCCGTGTAGGGGGTCGCCA
TAGGCTTCGACCTCCCTTTTAGGCCTTCCGGACTACAAAAATGGCGGATT
CAGTGACGTCACGGCCGCCATTTTAAGTAGGTGCCGTCCAGGACTGCAGT
TCCGGGTCAGAGTGCATCCTCGGCGGAACCTGCACAAAATGGCGGTCAAT
ATCTTCCGGGTCAAAGGTCACACCTACGTCATAAGTCACGTGACTGGGTC
CTGCTACGTCATATGCGGAAGTAGGCCCCGCCACGTGACTCGTCACGTGG
GCGCTGCGTCACGGCGGCCATTTTGTATCACAAAATGGCGGACTTCCTTC
CTCTTTTTTAAAAATAACGGCCCAGCGGCGGCGCGCGCTTCGCGCGCG
CGCCGGGGGGCTCCGCCCCCCCCCGCGCATGCGCGGGGCCCCCCCCCGCG
GGGGGCTCCGCCCCCCGGTCCCCCCCCG (SEQ ID NO: 24)

Annotations:

Putative Domain	*Base range*
TATA Box	82 – 87
Initiator Element	95-115
Transcriptional Start Site	115
5’ UTR Conserved Domain	170 - 238
ORF2	335 - 721
ORF2/2	335 – 717 ; 2446 – 2902
ORF2/3	335 – 717 ; 2675 – 3109
ORF1	586 – 2928
ORF1/1	586 – 717 ; 2446 – 2928
ORF1/2	586 – 717 ; 2675 – 2902
Three open-reading frame region	2640 – 2899
Poly(A) Signal	3106 - 3114
GC-rich region	3768 – 3878

Table 8. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 4)

TTV-HD20a (*Alphatorquevirus* Clade 4)	
ORF2	MSRCWCPPRGDVSSRESRWYEAVRGSHDAFCGCSDPILHLSRLAARFNHQG PPTPPTDDRAPQNTPVRRLLPLPSYPGEGPQARWPGGDGGAAGGDRREGGD GGARAAEDEYQPEDLDELFGAIEQEQ (SEQ ID NO: 25)
ORF2/2	MSRCWCPPRGDVSSRESRWYEAVRGSHDAFCGCSDPILHLSRLAARFNHQG PPTPPTDDRAPQNTPVRRLLPLPSYPGEGPQARWPGGDGGAAGGDRREGGD GGARAAEDEYQPEDLDELFGAIEQEQSSETRAKTTKYPVPVDSLGIYKSLTR NASRRRGSSTASTSEEASLLKQLSGACSRNHYLASMLLAPSGHPSSFYPQNT SEKTALQKAPQVHRPKDPVSGQKRVRRRRSPRRRTRRRRRGSSSSESSESSE HSSSNSSTSRSNSPRPRRIST (SEQ ID NO: 26)
ORF2/3	MSRCWCPPRGDVSSRESRWYEAVRGSHDAFCGCSDPILHLSRLAARFNHQG PPTPPTDDRAPQNTPVRRLLPLPSYPGEGPQARWPGGDGGAAGGDRREGGD GGARAAEDEYQPEDLDELFGAIEQEQIPARRRRCRKRLRFTGQKTPYLVRRE SDGDDLLGGEPGGDDEGAPPAKAPRAASTPVPTPALRGPTRQDPGESPRKPP VIFPAMNKVFLFPPEGPKPILGKEAWQDEYETCRVWNRPARTHHTDTPFYP WAPHKFHVSFKLGFQ (SEQ ID NO: 27)
ORF1	MAWWGWRRRWWRPKRRWRWRRARRRRRVPARRPRRAFRRYRTRTVRR RRRGRRRGYRRRYRLRRYARRRFRRKKIVLTQWNPQTTRKCIIRGMMPVL WAGMGTGGRNYAVRSDDYVVNKGFGGSFATETFSLKVLYDQFQRGFNRW SHTNEDLDLARYRGCRWTFYRHKDTDFIVYFTNNPPMKTNQFSAPLTTPGM LMRSKYKVLIPSFQTRPKGRKTVTVKIRPPKLFQDKWYTQQDLCSVPLVQL NVTAADFTHPFGSPLTETPCVEFQVLGDLYNTCLNIDLPQFSELGEITSAYSK PNSNNLKELYKELFTKATSGHYWQTFITNSMVRAHIDADKAKEAQRASTTP SYNNDPFPTIPVKSEFAQWKKKFTDTRDSPFLFATYHPEAIKDTIMKMRENN FKLETGPNDKYGDYTAQYQGNTHMLDYYLGFYSPIFLSDGRSNVEFFTAYR DIVYNPFLDKAQGNMVWFQYHTKTDNKFKKPECHWEIKDMPLWALLNGY VDYLETQIQYGDLSKEGKVLIRCPYTKPALVDPRDDTAGYVVYNRNFGRGK WIDGGGYIPLHERTKWYVMLRYQTDVFHDIVTCGPWQYRDDNKNSQLVA KYRFSFIWGGNTVHSQVIRNPCKDNQVSGPRRQPRDIQVVDPQRITPPWVLH

	SFDQRRGLFTETALRRLLQEPLPGEYAVSTLRTPLLFLPSEYQREDGAAESAS GSPAKRPRIWSEESQTETISSEENPAETTRELLQRKLREQRALQFQLQHFAVQ LAKTQANLHVNPLLSFPQ (SEQ ID NO: 28)
ORF1/1	MAWWGWRRRWWRPKRRWRWRRARRRRRVPARRPRRAFRRYRTRTVIRN PCKDNQVSGPRRQPRDIQVVDPQRITPPWVLHSFDQRRGLFTETALRRLLQE PLPGEYAVSTLRTPLLFLPSEYQREDGAAESASGSPAKRPRIWSEESQTETISS EENPAETTRELLQRKLREQRALQFQLQHFAVQLAKTQANLHVNPLLSFPQ (SEQ ID NO: 29)
ORF1/2	MAWWGWRRRWWRPKRRWRWRRARRRRRVPARRPRRAFRRYRTRTNTSE KTALQKAPQVHRPKDPVSGQKRVRRRRSPRRRTRRRRRGSSSSESSESSEHS SSNSSTSRSNSPRPRRIST (SEQ ID NO: 30)

Table 9. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus,* Clade 5)

Name	TTV-16 (TUS01)
Genus/Clade	*Alphatorquevirus*, Clade 5
Accession Number	AB017613.1

Full Sequence: 3818 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
AAGTCCGCCACTAACCACGTGACTCCCGCAGGCCAACCCAGTACTATGTC
GTCCACTTCCTGGGACGAGTCTACGTCCTGATATAAGTAAGTGCACTTCC
GAATGGCTGAGTTTTCCACGCCCGTCCGCAGCGAGAACGCCACGGAGGGG
AGTCCGCGCGTCCCGAGGGCGGGTGCCGGAGGTGAGTTTACACACCGCAG
TCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCCCCGGGCAAGGCTCTT
AAAAAATGCACTTTCGCAGAGTGCGAGCGAAAAGGAAACTGCTACTGCAA
GCTGTGCGAGCTCCACCGAAGGCACCTGCCATGAGCTTCACCACACCTAC
TATTAATGCCGGGATCCGAGAGCAGCAATGGTTCGAGTCCACCCTTAGAT
CCCACCACTCGTTCTGTGGCTGTGGTGATCCCGTGCTTCATTTTACTAAC
CTTGCTACTCGCTTTAACTATCTGCCTGCTACCTCTTCGCCTCTGGACCC
TCCCGGCCCAGCGCCGCGAGGCCGCCCGGCGCTCCGCCGCCTCCCGGCAC
TCCCTTCAGCCCCCGCGACCCCTTCTAGAGAACTAGCATGGCCTACTGGT
TCAGAAGGTGGGGCTGGAGGCCGAGGCGCCGGTGGAGAAGGTGGCGCCGC
CGTCGAAGGAGACTACCGAGAAGAAGAACTAGACGAGCTGTTCGCGGCCT
TGGAAGAAGACGCAAACCAAGGGTAAGGAGGCGCCGCAGAACTCGCAGAC
GTACCTACAGACGGGGGTGGAGACGCAGGAGGTACATAAGACGGGGGCGA
CGCAAAAAGAAACTCATACTGACTCAGTGGAACCCGGCAATAGTTAAGAG
GTGCAACATTAAGGGCGGACTTCCAATAATTATATGCGGAGAGCCCAGGG
CAGCCTTTAACTATGGCTACCACATGGAGGACTACACTCCTCAACCTTTC
```

CCCTTCGGAGGGGGAATGAGCACAGTGACTTTCTCTCTGAAAGCCTTGTA
TGACCAGTACCTAAAACACCAAAACAGGTGGACTTTCTCAAACGACCAGC
TAGACCTCGCCAGATACAGGGGCTGTAAACTAAGGTTCTACAGAAGCCCC
GTCTGTGACTTTATAGTACACTACAACCTAATACCTCCACTAAAAATGAA
CCAGTTCACAAGTCCCAACACGCACCCGGGACTACTCATGCTCAGCAAAC
ACAAGATAATAATTCCCAGCTTTCAAACAAGACCTGGGGGCAGACGCTTT
GTTAAAATAAGACTTAATCCCCCCAAACTATTTGAAGACAAGTGGTACAC
TCAGCAAGACCTGTGCAAGGTTCCGCTCGTTAGTATTACAGCAACTGCGG
CTGACTTGCGGTATCCGTTCTGCTCACCACAAACGAACAACCCTTGCACC
ACCTTCCAGGTACTGCGCAAGAACTACAATACAGTTATAGGAACTTCCGT
AAAAGACCAAGAGTCCACACAAGACTTTGAAAATTGGCTTTATAAAACAG
ACTCACACTATCAAACATTTGCCACAGAGGCTCAACTAGGCAGAATTCCT
GCATTTAATCCTGATGGCACTAAAAACACTAAACAGCAGTCGTGGCAAGA
TAACTGGAGCAAAAAAAATTCACCATGGACAGGTAACTCAGGTACATACC
CACAAACAACCAGTGAAATGTACAAAATTCCATATGACAGTAACTTCGGC
TTTCCCACATACAGAGCCCAAAAAGACTACATTTTAGAAAGAAGACAGTG
CAACTTTAACTATGAAGTTAATAATCCAGTTAGCAAAAAAGTATGGCCAC
AACCTAGTACAACAACACCCACAGTAGACTACTATGAATACCACTGTGGA
TGGTTCAGCAACATATTCATAGGCCCCAACAGATACAACCTACAGTTTCA
AACAGCATATGTAGACACCACATACAACCCACTAATGGACAAGGGCAAAG
GCAACAAAATATGGTTTCAATATCTGTCTAAAAAGGGCACAGACTACAAT
GAAAAACAATGCTACTGCACCCTAGAAGACATGCCCCTATGGGCAATATG
CTTTGGATACACTGACTATGTAGAGACTCAACTAGGACCCAATGTGGACC
ATGAAACAGCAGGCTTAATAATTATGATCTGTCCATACACTCAACCACCT
ATGTATGACAAAAACAGACCTAACTGGGGATACGTAGTCTATGACACAAA
CTTTGGCAATGGAAAAATGCCCTCAGGAAGTGGCCAAGTCCCAGTATACT
GGCAATGCCGATGGAGGCCCATGCTGTGGTTCCAACAACAAGTACTCAAT
GACATCTCAAAGACTGGACCGTACGCCTACAGAGACGAATATAAAAATGT
ACAACTGACTCTCTACTACAACTTTATTTTTAACTGGGGGGGCGACATGT
ATTACCCACAGGTCGTTAAAAACCCCTGTGGAGACTCCGGAATCGTTCCC
GGTTCCGGTAGATTCACTCGAGAAGTACAAGTCGTTAGCCCGCTTTCCAT
GGGACCGGCCTACATCTTCCACTACTTCGACTCCAGACGCGGGTTCTTTA
GTGAAAAAGCTCTTAAAAGAATGCAACAACAACAAGAATTTGATGAATCT
TTTACATTCAAACCTAAGAGACCCAAACTTTCTACAGCAGCCGCAGAAAT
CCTCCAGCTCGAAGAAGACTCGACTTCAGGGGAAGGAAAATCGCCACTAC
AGCAAGAAGAGAAAGAAGTCGAAGTCCTCCAAACGCCGACAGTACAGCTC
CAGCTCCAGCGAAACATCCAGGAGCAGCTCGCAATCAAGCAGCAGCTCCA
ATTCCTCTTGCTCCAACTCCTCAAAACCCAATCCAATTTGCATTTAAACC
CACAATTTTTAAGCCCTTCATAAAATATGACATGTTTGGGGACCCCCTTC
CTCACCCCCCAACAGCCGAAGAGTGGGAAACAGAGTACCAGTGCTGTAAG
GCCTTTAACAGACCACCTAGAACCAACCTAAAAGACACCCCCTTCTACCC
CTGGGTACCTAAACCTAAACCTCAATTCCGTGTATCTTTTAAACTTGGTT
TTCAATAAACAAGGCCGTGGGAGTTTCACTTGTCGGTGTCAACCTCTTAA
GGTCACTAAGCACTCCGAGCGTAAGCGAGGAGTGCGACCCTCCCCCCTGG
GGCAACTCCCTCGAAGTCCGGCGCTACGCGCTTCGCGCTGCGCCGGACAT
CTCGGACCCCCCCTCCACCCGAAACGCTTGCGCGTTTCGGACCTTCGGCG
TCGGGGGGGTCGGGGGCTTTACTAAACAGACTCCGAGGTGCCATTGGACA
CTGAGGGGATGAACAGCAACGAAAGTGAGTGGGGCCAGACTTCGCCATAA
GGCCTTTATCTTCTTGCCATTTGTCAGTATAGAGGGTCGCCATAGGCTTC
GGCCTCCATTTTAACCTCTAAAAACTACCAAAATGGCCGTTCCAGTGACG
TCACAGCCGCCATTTTAAGTAGCTGACGTCAAGGATTGACGTGAAGGTTA
AAGGTCATCCTCGGCGGAAGCTACACAAAATGGTGGACAACATCTTCCGG
GTCAAAGGTCGTGCACACGTCATAAGTCACGTGGTGGGGACCCGCTGTAA
CCCGGAAGTAGGCCCCGTCACGTGATTTGTCACGTGTGTACACGTCACAA
CCGCCATTTTGTTTTACAAAATGGCTGACTTCCTTCCTCTTTTTTAAAAA
AAACGGCCGTGCGGCGGCGCGCGCGCTTCGCGCGCGCGCCGGGGGCTGCC
GCCCCCCCCCGCGCATGCGCGCGGGGCCCCCCCCCGCGGGGGGCTCCGCC

CCCCGGCCCCCCCCCCCG (SEQ ID NO: 31)

Annotations:

Putative Domain	*Base range*
TATA Box	82 – 86
Initiator Element	100-115
Transcriptional Start Site	115
5' UTR Conserved Domain	170 - 240
ORF2	331- 726
ORF2/2	331 – 722 ; 2412 – 2847
ORF2/3	331 – 722 ; 2638 – 3058
ORF2t/3	331 – 380 ; 2638 - 3058
ORF1	588 – 2873
ORF1/1	588 – 722 ; 2412 – 2873
ORF1/2	588 – 722 ; 2638 – 2847
Three open-reading frame region	2699 – 2969
Poly(A) Signal	3220 - 3225
GC-rich region	3302 – 3541

Table 10. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 5)

TTV-16-TUS01 (*Alphatorquevirus* Clade 5)	
ORF2	MSFTTPTINAGIREQQWFESTLRSHHSFCGCGDPVLHFTNLATRFNYLPATSS PLDPPGPAPRGRPALRRLPALPSAPATPSRELAWPTGSEGGAGGRGAGGEGG AAVEGDYREEELDELFAALEEDANQG (SEQ ID NO: 32)
ORF2/2	MSFTTPTINAGIREQQWFESTLRSHHSFCGCGDPVLHFTNLATRFNYLPATSS PLDPPGPAPRGRPALRRLPALPSAPATPSRELAWPTGSEGGAGGRGAGGEGG AAVEGDYREEELDELFAALEEDANQGSLKTPVETPESFPVPVDSLEKYKSLA RFPWDRPTSSTTSTPDAGSLVKKLLKECNNNKNLMNLLHSNLRDPNFLQQP QKSSSSKKTRLQGKENRHYSKKRKKSKSSKRRQYSSSSSETSRSSSQSSSSSN SSCSNSSKPNPICI (SEQ ID NO: 33)
ORF2/3	MSFTTPTINAGIREQQWFESTLRSHHSFCGCGDPVLHFTNLATRFNYLPATSS PLDPPGPAPRGRPALRRLPALPSAPATPSRELAWPTGSEGGAGGRGAGGEGG AAVEGDYREEELDELFAALEEDANQGSRRNPPARRRLDFRGRKIATTARRE

	RSRSPPNADSTAPAPAKHPGAARNQAAAPIPLAPTPQNPIQFAFKPTIFKPFIK YDMFGDPLPHPPTAEEWETEYQCCKAFNRPPRTNLKDTPFYPWVPKPKPQF RVSFKLGFQ (SEQ ID NO: 34)
ORF2t/3	MSFTTPTINAGIREQQCSRRNPPARRRLDFRGRKIATTARRERSRSPPNADST APAPAKHPGAARNQAAAPIPLAPTPQNPIQFAFKPTIFKPFIKYDMFGDPLPH PPTAEEWETEYQCCKAFNRPPRTNLKDTPFYPWVPKPKPQFRVSFKLGFQ (SEQ ID NO: 35)
ORF1	MAYWFRRWGWRPRRRWRRWRRRRRRLPRRRTRRAVRGLGRRRKPRVRR RRRTRRRTYRRGWRRRRYIRRGRRKKKLILTQWNPAIVKRCNIKGGLPIIICG EPRAAFNYGYHMEDYTPQPFPFGGGMSTVTFSLKALYDQYLKHQNRWTFS NDQLDLARYRGCKLRFYRSPVCDFIVHYNLIPPLKMNQFTSPNTHPGLLMLS KHKIIIPSFQTRPGGRRFVKIRLNPPKLFEDKWYTQQDLCKVPLVSITATAAD LRYPFCSPQTNNPCTTFQVLRKNYNTVIGTSVKDQESTQDFENWLYKTDSH YQTFATEAQLGRIPAFNPDGTKNTKQQSWQDNWSKKNSPWTGNSGTYPQT TSEMYKIPYDSNFGFPTYRAQKDYILERRQCNFNYEVNNPVSKKVWPQPST TTPTVDYYEYHCGWFSNIFIGPNRYNLQFQTAYVDTTYNPLMDKGKGNKIW FQYLSKKGTDYNEKQCYCTLEDMPLWAICFGYTDYVETQLGPNVDHETAG LIIMICPYTQPPMYDKNRPNWGYVVYDTNFGNGKMPSGSGQVPVYWQCR WRPMLWFQQQVLNDISKTGPYAYRDEYKNVQLTLYYNFIFNWGGDMYYP QVVKNPCGDSGIVPGSGRFTREVQVVSPLSMGPAYIFHYFDSRRGFFSEKAL KRMQQQQEFDESFTFKPKRPKLSTAAAEILQLEEDSTSGEGKSPLQQEEKEV EVLQTPTVQLQLQRNIQEQLAIKQQLQFLLLQLLKTQSNLHLNPQFLSPS (SEQ ID NO: 36)
ORF1/1	MAYWFRRWGWRPRRRWRRWRRRRRRLPRRRTRRAVRGLGRRRKPRVVK NPCGDSGIVPGSGRFTREVQVVSPLSMGPAYIFHYFDSRRGFFSEKALKRMQ QQQEFDESFTFKPKRPKLSTAAAEILQLEEDSTSGEGKSPLQQEEKEVEVLQT PTVQLQLQRNIQEQLAIKQQLQFLLLQLLKTQSNLHLNPQFLSPS (SEQ ID NO: 37)

ORF1/2	MAYWFRRWGWRPRRRWRRWRRRRRRLPRRRTRRAVRGLGRRRKPRQPQ KSSSSKKTRLQGKENRHYSKKRKKSKSSKRRQYSSSSSETSRSSSQSSSSSNS SCSNSSKPNPICI (SEQ ID NO: 38)

Table 11. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus,* Clade 6)

Name	TTV-TJN02
Genus/Clade	*Alphatorquevirus*, Clade 6
Accession Number	AB028669.1

Full Sequence: 3794 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
CCCGAAGTCCGTCACTAACCACGTGACTCCTGTCGCCCAATCAGAGTGTA
TGTCGTGCATTTCCTGGGCATGGTCTACATCCTGATATAACTAAGTGCAC
TTCCGAATGGCTGAGTTTTCCACGCCCGTCCGCAGCGAGGGAGCGACGGA
GGAGCTCCCGAGCGTCCCGAGGGCGGGTGCCGGAGGTGAGTTTACACACC
GCAGTCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTATGGGCAAGGC
TCTTAGGGTCTTCATTCTTAATATGTTTCTTGGCAGAGTTTACCGCCACA
AGAAAAGGAAAGTGCTACTGTCCACACTGCGAGCTCCACAGGCGTCTCGC
AGGGCTATGAGTTGGCGACCCCCGGTACACGATGCACCCGGCATCGAGCG
CAATTGGTACGAGGCCTGTTTCAGAGCCCACGCTGGAGCTTGTGGCTGTG
GCAATTTTATTATGCACCTTAATCTTTTGGCTGGGCGTTATGGTTTTACT
CCGGGGTCAGCGCCGCCAGGTGGTCCTCCTCCGGGCACCCCGCAGATAAG
GAGAGCCAGGCCTAGTCCCGCCGCACCAGAGCAGCCCGCTGCCCTACCAT
GGCATGGGGATGGTGGAGATGGCGGCGCCGCTGGCCCGCCAGACGCTGGA
GGAGACGCCGTCGCCGGCGCCCCGTACGGAGAACAAGAGCTCGCCGACCT
GCTCGACGCTATAGAAGACGACGAACAGTAAGAACCAGGCGAAGGCGGTG
GGGGCGCAGACGGTACAGACGGGGCTGGAGACGCAGGACTTATGTGAGAA
AGGGGCGACACAGAAAAAAGAAAAAGAGACTGATACTGAGACAGTGGCAA
CCAGCCACAAGACGCAGATGTACCATAACTGGGTACCTGCCCATAGTGTT
CTGCGGCCACACTAGGGGCAATAAAAACTATGCACTACACTCTGACGACT
ACACCCCCCAAGGACAACCATTTGGAGGGGCTCTAAGCACTACCTCATTC
TCTTTAAAAGTACTATTTGACCAGCATCAGAGAGGACTAAACAAGTGGTC
TTTTCCAAACGACCAACTAGACCTCGCCAGATATAGAGGCTGCAAATTTA
TATTTTATAGAACAAAACAAACTGACTGGGTGGGCCAGTATGACATATCA
GAACCCTACAAGCTAGACAAATACAGCTGCCCCAACTATCACCCTGGAAA
CATGATTAAGGCAAAGCACAAATTTTTAATACCAAGCTATGACACTAATC
CTAGAGGCAGACAAAAAATTATAGTTAAAATTCCCCCCCCAGACCTCTTT
GTAGACAAGTGGTACACTCAAGAGGATCTGTGTTCCGTTAATCTTGTGTC
ACTTGCGGTTTCTGCGGCTTCCTTTCTCCACCCATTCGGCTCACCACAAA
CTGACAACCCTTGCTACACCTTCCAGGTGTTGAAAGAGTTCTACTATCAG
GCAATAGGCTTCTCTGCAAGCACACAAGCAATGACATCAGTATTAGACAC
GCTATACACACAAAACAGTTATTGGGAATCTAATCTAACTCAGTTTTATG
TACTTAATGCAAAAAAAGGCAGTGATACAACACAGCCTTTAACTAGCAAT
ATGCCAACTCGTGAAGAGTTTATGGCAAAAAAAAATACCAATTACAACTG
GTATACATACAAGGCCGCGTCAGTAAAAAATAAACTACATCAAATGAGAC
```

AAACCTATTTTGAGGAGTTAACCTCTAAGGGGCCACAAACAACAAAAAGT
GAGGAAGGCTACAGTCAGCACTGGACCACCCCCTCCACAAACGCCTACGA
ATATCACTTAGGAATGTTTAGTGCAATATTTCTAGCCCCAGACAGGCCAG
TACCTAGATTTCCATGCGCCTACCAAGATGTAACTTACAACCCCTTAATG
GACAAAGGGGTGGGAAACCACATTTGGTTTCAGTACAACACAAAGGCAGA
CACTCAGCTAATAGTCACAGGAGGGTCCTGCAAAGCACACATACAAGACA
TACCACTGTGGGCGGCCTTCTATGGATACAGTGACTTTATAGAGTCAGAA
CTAGGCCCCTTTGTAGATGCAGAGACGGTAGGCTTAGTGTGTGTAATATG
CCCTTATACAAAACCCCCCATGTACAACAAGACAAACCCCGCCATGGGCT
ACGTGTTCTATGACAGAAACTTTGGTGACGGAAAATGGACTGACGGACGG
GGCAAAATAGAGCCCTACTGGCAAGTTAGGTGGAGGCCCGAAATGCTTTT
CCAAGAAACTGTAATGGCAGACCTAGTTCAGACTGGGCCCTTTAGCTACA
AAGACGAACTTAAAAACAGCACCCTAGTGTGCAAGTACAAATTCTATTTC
ACCTGGGGAGGTAACATGATGTTCCAACAGACGATCAAAAACCCGTGCAA
GACGGACGGACAACCCACCGACTCCAGTAGACACCCTAGAGGAATACAAG
TGGCGGACCCGGAACAAATGGGACCCCGCTGGGTGTTCCACTCCTTTGAC
TGGCGAAGGGGCTATCTTAGCGAGAAAGCTCTCAAACGCCTGCAAGAAAA
ACCTCTTGACTATGACGAATATTTTACACAACCAAAAAGACCTAGAATCT
TTCCTCCAACAGAATCAGCAGAGGGAGAGTTCCGAGAGCCCGAAAAAGGC
TCGTATTCAGAGGAAGAAAGGTCGCAAGCCTCTGCCGAAGAGCAGACGCA
GGAGGCGACAGTACTCCTCCTCAAGCGACGACTCAGAGAGCAACAGCAGC
TCCAGCAGCAGCTCCAATTCCTCACCCGAGAAATGTTCAAAACGCAAGCG
GGTCTCCACCTAAACCCTATGTTATTAAACCAGCGATAAACCAAGTGTAC
CTGTTTCCAGAGAGGGCCCCAAAACCCCCTCCTAGCAGCCAAGACTGGCA
GCAGGAGTACGAGGCCTGCGCAGCCTGGGACAGGCCCCCTAGATACAATC
TGTCCTCTCCTCCTTTCTACCCCAGCTGCCCTTCAAAATTCTGTGTAAAA
TTCAGCCTTGGCTTTAAATAAATGGCAACTTTACTGTGCAAGGCCGTGGG
AGTTTCACTGGTCGGTGTCTACCTCTAAAGGTCACTAAGCACTCCGAGCG
TTAGCGAGGAGTGCGACCCTTCCCCCTGACTCAACTTCTTCGGAGCCGCG
CGCTACGCCTTCGGCTGCGCGCGGCACCTCAGACCCCCGCTCGTGCTGAC
ACGCTCGCGCGTGTCAGACCACTTCGGGCTCGCGGGGGTCGGGAATTTTG
CTAAACAGACTCCGAGTTGCTCTTGGACACTGAGGGGGCATATCAGTAAC
GAAAGTGAGTGGGGCCAGACTTCGCCATAAGGCCTTTATCTTCTTGCCAT
TGGATAGTATCGAGGGTTGCCATAGGCTTCGACCTCCATTTTAGGCCTTC
CGGACTACAAAAATGGCCGTTTTAGTGACGTCACGGCCGCCATTTTAAGT
AAGGCGGAAGCAGCTCGGCGTACACAAAATGGCGGCGGAGCACTTCCGGC
TTGCCCAAAATGGTGGGCAACTTCTTCCGGGTCAAAGGTCACAGCTACGT
CACAAGTCACGTGGGGAGGGTTGGCGTTTAACCCGGAAGCCAATCCTCTT
ACGTGGCCTGTCACGTGACTTGTACGTCACGACCACCATTTTGTTTTACA
AAATGGCCGACTTCCTTCCTCTTTTTTAAAAATAACGGTTCGGCGGCGGC
GCGCGCGCTACGCGCGCGCGCCGGGGGGCTGCCGCCCCCCCCCCGCGCAT
GCGCGGGGCCCCCCCCCGCGGGGGGCTCCGCCCCCGGCCCCCC (SEQ ID NO: 39)

Annotations:

Putative Domain	*Base range*
TATA Box	89 – 90
Cap Site	107 – 114
Transcriptional Start Site	114
5’ UTR Conserved Domain	174 – 244
ORF2	357 – 731
ORF2/2	357 – 727 ; 2381 – 2813

ORF2/3	357 – 727 ; 2619 – 3021
ORF2t/3	357 – 406 ; 2619 – 3021
ORF1	599 – 2839
ORF1/1	599 – 727 ; 2381 – 2839
ORF1/2	599 – 727 ; 2619 – 2813
Three open-reading frame region	2596 – 2810
Poly(A) Signal	3017 – 3022
GC-rich region	3691 – 3794

Table 12. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 6)

TTV-TJN02 (*Alphatorquevirus* Clade 6)	
ORF2	MSWRPPVHDAPGIERNWYEACFRAHAGACGCGNFIMHLNLLAGRYGFTPGSAPPG GPPPGTPQIRRARPSPAAPEQPAALPWHGDGGDGGAAGPPDAGGDAVAGAPYGEQ ELADLLDAIEDDEQ (SEQ ID NO: 40)
ORF2/2	MSWRPPVHDAPGIERNWYEACFRAHAGACGCGNFIMHLNLLAGRYGFTPGSAPPG GPPPGTPQIRRARPSPAAPEQPAALPWHGDGGDGGAAGPPDAGGDAVAGAPYGEQ ELADLLDAIEDDEQRSKTRARRTDNPPTPVDTLEEYKWRTRNKWDPAGCSTPLTGE GAILARKLSNACKKNLLTMTNILHNQKDLESFLQQNQQRESSESPKKARIQRKKGR KPLPKSRRRRRQYSSSSDDSESNSSSSSSSNSSPEKCSKRKRVST (SEQ ID NO: 41)
ORF2/3	MSWRPPVHDAPGIERNWYEACFRAHAGACGCGNFIMHLNLLAGRYGFTPGSAPPG GPPPGTPQIRRARPSPAAPEQPAALPWHGDGGDGGAAGPPDAGGDAVAGAPYGEQ ELADLLDAIEDDEHRGRVPRARKRLVFRGRKVASLCRRADAGGDSTPPQATTQRAT AAPAAAPIPHPRNVQNASGSPPKPYVIKPAINQVYLFPERAPKPPPSSQDWQQEYEA CAAWDRPPRYNLSSPPFYPSCPSKFCVKFSLGFK (SEQ ID NO: 42)
ORF2t/3	MSWRPPVHDAPGIERNCRGRVPRARKRLVFRGRKVASLCRRADAGGDSTPPQATT QRATAAPAAAPIPHPRNVQNASGSPPKPYVIKPAINQVYLFPERAPKPPPSSQDWQQ EYEACAAWDRPPRYNLSSPPFYPSCPSKFCVKFSLGFK (SEQ ID NO: 43)
ORF1	MAWGWWRWRRRWPARRWRRRRRRRPVRRTRARRPARRYRRRRTVRTRRRRWG RRRYRRGWRRRTYVRKGRHRKKKKRLILRQWQPATRRRCTITGYLPIVFCGHTRG NKNYALHSDDYTPQGQPFGGALSTTSFSLKVLFDQHQRGLNKWSFPNDQLDLARY

	RGCKFIFYRTKQTDWVGQYDISEPYKLDKYSCPNYHPGNMIKAKHKFLIPSYDTNP RGRQKIIVKIPPPDLFVDKWYTQEDLCSVNLVSLAVSAASFLHPFGSPQTDNPCYTF QVLKEFYYQAIGFSASTQAMTSVLDTLYTQNSYWESNLTQFYVLNAKKGSDTTQPL TSNMPTREEFMAKKNTNYNWYTYKAASVKNKLHQMRQTYFEELTSKGPQTTKSE EGYSQHWTTPSTNAYEYHLGMFSAIFLAPDRPVPRFPCAYQDVTYNPLMDKGVGN HIWFQYNTKADTQLIVTGGSCKAHIQDIPLWAAFYGYSDFIESELGPFVDAETVGLV CVICPYTKPPMYNKTNPAMGYVFYDRNFGDGKWTDGRGKIEPYWQVRWRPEMLF QETVMADLVQTGPFSYKDELKNSTLVCKYKFYFTWGGNMMFQQTIKNPCKTDGQ PTDSSRHPRGIQVADPEQMGPRWVFHSFDWRRGYLSEKALKRLQEKPLDYDEYFT QPKRPRIFPPTESAEGEFREPEKGSYSEEERSQASAEEQTQEATVLLLKRRLREQQQL QQQLQFLTREMFKTQAGLHLNPMLLNQR (SEQ ID NO: 44)
ORF1/1	MAWGWWRWRRRWPARRWRRRRRRRPVRRTRARRPARRYRRRRTTIKNPCKTDG QPTDSSRHPRGIQVADPEQMGPRWVFHSFDWRRGYLSEKALKRLQEKPLDYDEYF TQPKRPRIFPPTESAEGEFREPEKGSYSEEERSQASAEEQTQEATVLLLKRRLREQQQ LQQQLQFLTREMFKTQAGLHLNPMLLNQR (SEQ ID NO: 45)
ORF1/2	MAWGWWRWRRRWPARRWRRRRRRRPVRRTRARRPARRYRRRRTQRESSESPKK ARIQRKKGRKPLPKSRRRRRQYSSSSDDSESNSSSSSSSNSSPEKCSKRKRVST (SEQ ID NO: 46)

Table 13. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus,* Clade 7)

Name	TTV-HD16d
Genus/Clade	*Alphatorquevirus*, Clade 7
Accession Number	FR751479.1

Full Sequence: 3866 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
AAGTCCGTCACTAACCACGTGACTCCCGCAGGCCAATCAGAGTCTATGTC
GTGCACTTCCTGGGCATGGTCTACGTTCTCATATAACTAACTGCACTTCC
GAATGGCTGAGTTTTCCACGCCCGTCCGCAGCGGCAGCACCACGGAGGGT
GATCCCCGCGTCCCGAGGGCGGGTGCCGAAGGTGAGTTTACACACCGCAG
TCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTATGGGCAAGGCTCTT
AGGGCTTTCATTGTTAAAAATGTTTCTCGGCAGGCCTTACAGGAGAAAGA
AAAGGGCGCTGTCACTGCCTGGCGTGCGAGCTGCACAGGCGAAACAACCT
GGTGATATGAGCTGGAGCCGTCCAGTACATAATGCCGCCGGGATCGAAAG
GCAGTGGTTCGAATCCACCTTTAGATCCCACGCTAGTTGCTGTGGCTGCG
GCAATTTTGTTAATCATATTAATGTACTGGCTGCTCGCTACGGCTTTACT
GGGGGGCCGACGCCGCCAGGTGGTCCTGGGCCGCGTCCACAACTGAGGCC
```

CGCGCTTCCCGCGCCGGACCCCGACCCCCAGGCGCCCAACCGTGAGCCAT
GGCGTGGAGCTGGTGGTGGCAACGATGGAGAAGGCGCCGCTGGAAACCCA
GGAGGCGCCGCTGGAGACGTCTACGATGGAGAAGACCTAGACGCGCTGTT
CGCCGCCGTCGTCGAGGACGTAGAGTAAGGAGGCGGAGGTGGGCGCGTAG
ACGGGGGCGACGCAGACGGTACGCCACCAGACGAAAGAGACGTTATAGGG
GTCGCCGCTTTAAAAAGAAACTAGTACTGACTCAGTGGCACCCTAATACC
ATGAGACGCTGCTTAATCAAGGGCATAGTCCCCCTGGTAATATGCGGCCA
CACCAGGTGGAACTACAACTACGCCCTCCATAGCAAGGACTACACAGAGG
AGGGTCGCTACCCTCACGGGGGGGCCCTCAGCACCACTACGTGGTCCCTT
AAGGTGCTGTATGACGAGCACCTCAAACACCACGACTTCTGGGGCTATCC
CAACAACCAGCTAGACCTGGCCAGGTACAAGGGGGCCAAGTTCACCTTCT
ACAGACACAAAAAGACTGACTTTATAATATTCTTTAACAGAAAGCCTCCC
TTTAAGCTAAACAAGTACAGCTGTGCCTCCTATCACCCAGGCATGCTGAT
GCAGCAGAGACACAAGATCCTGCTACCCAGCTACGAAACTAAACCCAAGG
GCAGGCCAAAGATAACAGTTAGAATAAAGCCCCCCACTCTGTTAGAGGAC
AAGTGGTACACCCAGCAGGACCTGTGCGACGTTAACCTGTTGCAACTTGT
GGTCACTGCGGCTGACTTTCGACATCCACTCTGCTCACCACAAACGAACA
CTCCAACCACAACCTTCCAGGTGTTGAAAGACATCTATTATGACACTATG
AGCATATCTGAACCCACAGACTCCTACACTAGTGTTAACAATAAAAGTAC
AACACAAACTTTTACTAACTACTCAAACACCTTAGAAAACATTCTGTACA
CACGAGCCTCCTACTGGAACTCGTTCCACGCCACTGAATACCTAAACCCC
AACATCATATACAAAAACGGTGAAAAACTATTCAAAGAACATGAAGACTT
AATAACCTGGATGACCCAAACTAACAATACCGGGTTTCTAACTAAAAACA
ACACAGCTTTTGGCAACAACAGCTACAGGCCCAATGCAGACAAAATTAAA
AAAGCCAGAAAGACATACTGGAACGCCCTAATAGGCACCAACGACCTGGC
CACTAATATAGGCCAGGCCAGAGCAGAAAGGTTCGAGTACCACCTAGGCT
GGTACTCCCCCATATTTCTCAGCAGACACAGGAGCAACATGAACTTTGCC
AGGGCCTACCAAGACGTCACATACAACCCCAACTGTGACAGGGGAGTTAA
CAACAGGGTGTGGGTTCAGCCTCTAACTAAACCCACCACAGAGTTCGACG
AGAAAAGGTGTAAGTGCGTAGTGCAGCACCTGCCTCTGTGGGCGGCTCTG
TACTGCTACCAAGACTTTGTAGAGGAGGAGCTGGGGTCCTCCTCAGAGAT
ATTAAATTCATGCCTACTGGTATTACAGTGCCCTTACACCTTTCCCCCAA
TGTATGACAAAAAGCTACCAGACAAGGGATTCGTGTTTTATGACTCCCTT
TTTGGAGACGGCAAAATGTCTGACGGACGCGGACAGGTGGACATTTTCTG
GCAACAGCGATGGTACCCTCGCTTAGCCACTCAGATGCAAGTCATGCACG
ACATCACCATGACGGGCCCCTTCTCCTACCGAGACGAGCTAGTTAGCACC
CAACTGACTGCCAAGTACACCTTTGACTTTATGTGGGGCGGAAATATGAT
CTCCACACAGATCATCAAGAACCCCTGCAAAGACAGTGGACTGGAACCCG
CCTACCCCGGTAGACAGCGTCGCGACTTACAAATTGTTGACCCATACTCC
ATGGGCCCCCAATTCTCGTTCCACAACTGGGACTACAGACATGGCCTTTT
TGGCCAAGACGCTATCGACAGAGTGTCTAAACAACCAAAAGATGATGCAG
ACTATCCTAACCCATACAAAAGGCCTAGATATTTTCCACCCACAGACCAA
GCCGCCCAAGAGCAAGAAAAAGACTTCAGTTTCCTCAAAACAGCACCGTC
GAACTCAGAAGAGAGCGATCAAGAAGTCCTCCAAGAAACGCAAGTACTCC
GATTCCAGCCAGAGCAGCACAAGCAACTCCACCTGCAGCTCGCAGAGCGG
CAGCGAATCGGAGAGCAACTCCGATACCTACTCCAACAGATGTTCAAAAC
TCAGGCCAATCTCCACCTAAACCCATATACATTTACCCAGCTGTAAAGCA
GGTGTTTATGTTTGACCCCCCGGGCCCTAAGGCTATCTCGGGCGCCAAGG
CCTGGGAGGACGAGTTCCTCACCGCAAAAGTGTGGAACCGCCCGGTACGC
AAGTACTACTCAGACACCCCCTACTACCCCTGGGCCCCCAAACCCCAGTA
CTCTGTCAGTTTCAAACTCGGCTGGAAATAAAAAAAGCCTGCTCCACTGT
ACTAGGCCGTGGGAGTTTCACTCGTCGGTGTCTACCTCTTAAGGTCACCA
AGCACTCCGAGCGTCAGCGAGGAGTGCGACCCTTGGGGGTGGGTGCAACG
CCCTCGGCGGCCGCGCGCTACGCCTTCGGCTGCGCGCGGCACCTCGGACC
CCCGCTCGTGCTGACGCGCTTGCGCGCGTCAGACCACTTCGGGCTCGCGG
GGGTCGGAAATTTTGCTAAACAGACTCCGAGTTGCCATTGGACACTGGAG
CCGTGAATCAGTAACGAAAGTGAGTGGGGCCAGACTTCGCCATAAGGCCT

TTATCTTTTTGCCATTTGTCCGTGGGGAAGGGTCGCTGCAAGCGCGGACC
CCGTTTTCACCCCTTCCGGACTACAAAAATAGCGCATTAGTGACGTCACG
GCCGCCATTTTAAGTAAGGCGGAAGCAACTCCACTTTCTCACAAAATGGC
GGCGGAGCACTTCCGGCTTGCCCAAAATGGCCGCCAAAAACATCCGGGTC
AAAGTTCGCCGCTACGTCATAAGTCACGTGACTGGGGAGGTACTTAAACA
CGGAAGTATCCTCAACCACGTAACTGGTCACGTGGTGCGCACGTCACGGC
AACCATTTTGTTTTACAAAATGGCGCATTTCCTTCCTCTTTTTTAAAAAT
TAACCGTTGGCGGCGGCGCGCGCTACGCGCGCGCGCCGGGGAGCTCTG
CCCCCCCCCGCGCATGCGCGCGGGTCCCCCCCCCGCGGGGGGCTCCGCCC
CCCGGTCCCCCCCCCG (SEQ ID NO: 47)

Annotations:

Putative Domain	*Base range*
TATA Box	82 – 86
Initiator Element	94-115
Transcriptional Start Site	115
5' UTR Conserved Domain	170 - 240
ORF2	357 - 728
ORF2/2	357 – 724 ; 2411 – 2870
ORF2/3	357 – 724 ; 2646 – 3081
ORF1	599 – 2896
ORF1/1	599 – 724 ; 2411 – 2896
ORF1/2	599 – 724 ; 2646 – 2870
Three open-reading frame region	2629 – 2867
Poly(A) Signal	3076 - 3086
GC-rich region	3759 – 3866

Table 14. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 7)

TTV-HD16d (*Alphatorquevirus* Clade 7)	
ORF2	MSWSRPVHNAAGIERQWFESTFRSHASCCGCGNFVNHINVLAARYGFTGGP TPPGGPGPRPQLRPALPAPDPDPQAPNREPWRGAGGGNDGEGAAGNPGGA AGDVYDGEDLDALFAAVVEDVE (SEQ ID NO: 48)
ORF2/2	MSWSRPVHNAAGIERQWFESTFRSHASCCGCGNFVNHINVLAARYGFTGGP TPPGGPGPRPQLRPALPAPDPDPQAPNREPWRGAGGGNDGEGAAGNPGGA AGDVYDGEDLDALFAAVVEDVESSRTPAKTVDWNPPTPVDSVATYKLLTH

	TPWAPNSRSTTGTTDMAFLAKTLSTECLNNQKMMQTILTHTKGLDIFHPQT KPPKSKKKTSVSSKQHRRTQKRAIKKSSKKRKYSDSSQSSTSNSTCSSQSGSE SESNSDTYSNRCSKLRPIST (SEQ ID NO: 49)
ORF2/3	MSWSRPVHNAAGIERQWFESTFRSHASCCGCGNFVNHINVLAARYGFTGGP TPPGGPGPRPQLRPALPAPDPDPQAPNREPWRGAGGGNDGEGAAGNPGGA AGDVYDGEDLDALFAAVVEDVEPSRPRARKRLQFPQNSTVELRRERSRSPP RNASTPIPARAAQATPPAARRAAANRRATPIPTPTDVQNSGQSPPKPIYIYPA VKQVFMFDPPGPKAISGAKAWEDEFLTAKVWNRPVRKYYSDTPYYPWAPK PQYSVSFKLGWK (SEQ ID NO: 50)
ORF1	MAWSWWWQRWRRRRWKPRRRRWRRLRWRRPRRAVRRRRRGRRVRRRR WARRRGRRRRYATRRKRRYRGRRFKKKLVLTQWHPNTMRRCLIKGIVPLV ICGHTRWNYNYALHSKDYTEEGRYPHGGALSTTTWSLKVLYDEHLKHHDF WGYPNNQLDLARYKGAKFTFYRHKKTDFIIFFNRKPPFKLNKYSCASYHPG MLMQQRHKILLPSYETKPKGRPKITVRIKPPTLLEDKWYTQQDLCDVNLLQ LVVTAADFRHPLCSPQTNTPTTTFQVLKDIYYDTMSISEPTDSYTSVNNKSTT QTFTNYSNTLENILYTRASYWNSFHATEYLNPNIIYKNGEKLFKEHEDLITW MTQTNNTGFLTKNNTAFGNNSYRPNADKIKKARKTYWNALIGTNDLATNI GQARAERFEYHLGWYSPIFLSRHRSNMNFARAYQDVTYNPNCDRGVNNRV WVQPLTKPTTEFDEKRCKCVVQHLPLWAALYCYQDFVEEELGSSSEILNSC LLVLQCPYTFPPMYDKKLPDKGFVFYDSLFGDGKMSDGRGQVDIFWQQRW YPRLATQMQVMHDITMTGPFSYRDELVSTQLTAKYTFDFMWGGNMISTQII KNPCKDSGLEPAYPGRQRRDLQIVDPYSMGPQFSFHNWDYRHGLFGQDAID RVSKQPKDDADYPNPYKRPRYFPPTDQAAQEQEKDFSFLKTAPSNSEESDQE VLQETQVLRFQPEQHKQLHLQLAERQRIGEQLRYLLQQMFKTQANLHLNPY TFTQL (SEQ ID NO: 51)
ORF1/1	MAWSWWWQRWRRRRWKPRRRRWRRLRWRRPRRAVRRRRRGRRIIKNPC KDSGLEPAYPGRQRRDLQIVDPYSMGPQFSFHNWDYRHGLFGQDAIDRVSK QPKDDADYPNPYKRPRYFPPTDQAAQEQEKDFSFLKTAPSNSEESDQEVLQE TQVLRFQPEQHKQLHLQLAERQRIGEQLRYLLQQMFKTQANLHLNPYTFTQ L (SEQ ID NO: 52)

ORF1/2	MAWSWWWQRWRRRRWKPRRRRWRRLRWRRPRRAVRRRRRGRRTKPPK SKKKTSVSSKQHRRTQKRAIKKSSKKRKYSDSSQSSTSNSTCSSQSGSESESN SDTYSNRCSKLRPIST (SEQ ID NO: 53)

Table 15. Exemplary *Anellovirus* nucleic acid sequence (*Betatorquevirus*)

Name	TTMV-LY2
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 16. Exemplary *Anellovirus* amino acid sequences (*Betatorquevirus*)

TTMV-LY2 (*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 17. Exemplary *Anellovirus* nucleic acid sequence (*Gammatorquevirus*)

Name	TTMDV-MD1-073
Genus/Clade	*Gammatorquevirus*
Accession Number	AB290918.1

Full Sequence: 3242 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
AGGTGGAGACTCTTAAGCTATATAACCAAGTGGGGTGGCGAATGGCTGAG
TTTACCCCGCTAGACGGTGCAGGGACCGGATCGAGCGCAGCGAGGAGGTC
CCCGGCTGCCCGTGGGCGGGAGCCCGAGGTGAGTGAAACCACCGAGGTCT
AGGGGCAATTCGGGCTAGGGCAGTCTAGCGGAACGGGCAAGAAACTTAAA
AATATTTCTTTTACAGATGCAAAACCTATCAGCCAAAGACTTCTACAAAC
CATGCAGATACAACTGTGAAACTAAAAACCAAATGTGGATGTCTGGCATT
GCTGACTCCCATGACAGTTGGTGTGACTGTGATACTCCTTTTGCTCACCT
CCTGGCTAGTATTTTTCCTCCTGGTCACACAGATCGCACACGAACCATCC
AAGAAATACTTACCAGAGATTTTAGGAAAACATGCCTTTCTGGTGGGGCC
GACGCAACAAATTCTGGTATGGCCGAAACTATAGAAGAAAAAAGAGAAGA
TTTCCAAAAAGAAGAAAAAGAAGATTTTACAGAAGAACAAAATATAGAAG
ACCTGCTCGCCGCCGTCGCAGACGCAGAAGGAAGGTAAGAAGAAAAAAAA
AAACTCTTATAGTAAGACAATGGCAGCCAGACTCTATTGTACTCTGTAAA
ATTAAAGGGTATGACTCTATAATATGGGGAGCTGAAGGCACACAGTTTCA
ATGTTCTACACATGAAATGTATGAATATACAAGACAAAAGTACCCTGGGG
GAGGAGGATTTGGTGTACAACTTTACAGCTTAGAGTATTTGTATGACCAA
TGGAAACTTAGAAATAATATATGGACTAAAACAAATCAACTCAAAGATTT
GTGTAGATACTTAAAATGTGTTATGACCTTTTACAGACACCAACACATAG
ATTTTGTAATTGTATATGAAAGACAACCCCCATTTGAAATAGATAAACTA
ACATACATGAAATATCATCCATATATGTTATTACAAAGAAAGCATAAAAT
AATTTTACCTAGTCAAACAACTAATCCTAGAGGTAAATTAAAAAAAAAGA
AAACTATTAAACCTCCCAAACAAATGCTCAGCAAATGGTTTTTTCAACAA
CAATTTGCTAAATATGATCTACTACTTATTGCTGCAGCAGCATGTAGTTT
AAGATACCCTAGAATAGGCTGCTGCAATGAAAATAGAATGATAACCTTAT
ACTGTTTAAATACTAAATTTTATCAAGATACAGAATGGGGAACTACAAAA
CAGGCCCCCCACTACTTTAAACCATATGCAACAATTAATAAATCCATGAT
ATTTGTCTCTAACTATGGAGGTAAAAAAACAGAATATAACATAGGCCAAT
GGATAGAAACAGATATACCTGGAGAAGGTAATCTAGCAAGATACTACAGA
TCAATAAGTAAAGAAGGAGGTTACTTTTCACCTAAAATACTGCAAGCATA
TCAAACAAAAGTAAAGTCTGTAGACTACAAACCTTTACCAATTGTTTTAG
GTAGATATAACCCAGCAATAGATGATGGAAAAGGCAACAAAATTTACTTA
CAAACTATAATGAATGGCCATTGGGGCCTACCTCAAAAAACACCAGATTA
TATAATAGAAGAGGTCCCTCTTTGGCTAGGCTTCTGGGGATACTATAACT
ACTTAAAACAAACAAGAACTGAAGCTATATTTCCACTACACATGTTTGTA
GTGCAAAGCAAATACATTCAAACACAACAAACAGAAACACCTAACAATTT
TTGGGCATTTATAGACAACAGCTTTATACAGGGCAAAAACCCATGGGACT
CAGTTATTACTTACTCAGAACAAAAGCTATGGTTTCCTACAGTTGCATGG
CAACTAAAAACCATAAATGCTATTTGTGAAAGTGGACCATATGTACCTAA
ACTAGACAATCAAACATATAGTACCTGGGAACTAGCAACTCATTACTCAT
TTCACTTTAAATGGGGTGGTCCACAGATATCAGACCAACCAGTTGAAGAC
CCAGGAAACAAAAACAAATATGATGTGCCCGATACAATCAAAGAAGCATT
ACAAATTGTTAACCCAGCAAAAAACATTGCTGCCACGATGTTCCATGACT
GGGACTACAGACGGGGTTGCATTACATCAACAGCTATTAAAAGAATGCAA
CAAAACCTCCCAACTGATTCATCTCTCGAATCTGATTCAGACTCAGAACC
AGCACCCAAGAAAAAAAGACTACTACCAGTCCTCCACGACCCACAAAAGA
AAACGGAAAAGATCAACCAATGTCTCCTCTCTCTCTGCGAAGAAAGTACA
TGCCAGGAGCAGGAAACGGAGGAAAACATCCTCAAGCTCATCCAGCAGCA
GCAGCAGCAGCAGCAGAAACTCAAGCACAACCTCTTAGTACTAATCAAGG
ACTTAAAAGTGAAACAAAGATTATTACAACTACAAACGGGGGTACTAGAA
TAACCCTTACCAGATTTAAACCAGGATTTGAGCAAGAAACTGAAAAAGAG
TTAGCACAAGCATTTAACAGACCCCCTAGACTGTTCAAAGAAGATAAACC
CTTTTACCCCTGGCTACCCAGATTTACACCCCTTGTAAACTTTCACCTTA
ATTTTAAAGGCTAGGCCTACACTGCTCACTTAGTGGTGTATGTTTATTAA
```

AGTTTGCACCCCAGAAAAATTGTAAAATAAAAAAAAAAAAAAAAAATAAA
AAATTGCAAAAATTCGGCGCTCGCGCGCGCTGCGCGCGCGAGCGCCGTCA
CGCGCCGGCGCTCGCGCGCCGCGCGTATGTGCTAACACACCACGCACCTA
GATTGGGGTGCGCGCGTAGCGCGCGCACCCCAATGCGCCCCGCCCTCGTT
CCGACCCGCTTGCGCGGGTCGGACCACTTCGGGCTCGGGGGGGCGCGCCT
GCGGCGCTTATTTACTAAACAGACTCCGAGTCGCCATTGGGCCCCCCCTA
AGCTCCGCCCCCCTCATGAATATTCATAAAGGAAACCACAAAATTAGAAT
TGCCGACCACAAACTGCCATATGCTAATTAGTTCCCCTTTTACACAGTAA
AAAGGGGAAGTGGGGGGGCAGAGCCCCCCCACACCCCCCGCGGGGGGGGC
AGAGCCCCCCCCCGCACCCCCCCTACGTCACAGGCCACGCCCCCGCCGCCA
TCTTGGGTGCGGCAGGGCGGGGACTAAAATGGCGGGACCCAATCATTTTA
TACTTTCACTTTCCAATTAAAACCCGCCACGTCACACAAAAG (SEQ ID NO: 61)

Annotations:

Putative Domain	*Base range*
TATA Box	21– 25
Cap Site	42 – 49
Transcriptional Start Site	49
5’ UTR Conserved Domain	117 – 187
ORF2	283 – 588
ORF2/2	283 – 584 ; 1977 – 2388
ORF2/3	283 – 584 ; 2197 – 2614
ORF1	432 – 2453
ORF1/1	432 – 584 ; 1977 – 2453
ORF1/2	432 – 584 ; 2197 – 2388
Three open-reading frame region	2186 – 2385
Poly(A) Signal	2676 – 2681
GC-rich region	3054 – 3172

Table 18. Exemplary *Anellovirus* amino acid sequences (*Gammatorquevirus*)

TTMDV-MD1-073 (*Gammatorquevirus*)	
ORF2	MWMSGIADSHDSWCDCDTPFAHLLASIFPPGHTDRTRTIQEILTRDFRKTCLSGGAD ATNSGMAETIEEKREDFQKEEKEDFTEEQNIEDLLAAVADAEGR (SEQ ID NO: 62)
ORF2/2	MWMSGIADSHDSWCDCDTPFAHLLASIFPPGHTDRTRTIQEILTRDFRKTCLSGGAD ATNSGMAETIEEKREDFQKEEKEDFTEEQNIEDLLAAVADAEGRYQTNQLKTQETK TNMMCPIQSKKHYKLLTQQKTLLPRCSMTGTTDGVALHQQLLKECNKTSQLIHLSN LIQTQNQHPRKKDYYQSSTTHKRKRKRSTNVSSLSAKKVHARSRKRRKTSSSSSSSS SSSSRNSSTTS (SEQ ID NO: 63)
ORF2/3	MWMSGIADSHDSWCDCDTPFAHLLASIFPPGHTDRTRTIQEILTRDFRKTCLSGGAD

	ATNSGMAETIEEKREDFQKEEKEDFTEEQNIEDLLAAVADAEGRTSTQEKKTTTSPP RPTKENGKDQPMSPLSLRRKYMPGAGNGGKHPQAHPAAAAAAAETQAQPLSTNQ GLKSETKIITTTNGGTRITLTRFKPGFEQETEKELAQAFNRPPRLFKEDKPFYPWLPRF TPLVNFHLNFKG (SEQ ID NO: 64)
ORF1	MPFWWGRRNKFWYGRNYRRKKRRFPKRRKRRFYRRTKYRRPARRRRRRRRKVR RKKKTLIVRQWQPDSIVLCKIKGYDSIIWGAEGTQFQCSTHEMYEYTRQKYPGGGG FGVQLYSLEYLYDQWKLRNNIWTKTNQLKDLCRYLKCVMTFYRHQHIDFVIVYER QPPFEIDKLTYMKYHPYMLLQRKHKIILPSQTTNPRGKLKKKKTIKPPKQMLSKWFF QQQFAKYDLLLIAAAACSLRYPRIGCCNENRMITLYCLNTKFYQDTEWGTTKQAPH YFKPYATINKSMIFVSNYGGKKTEYNIGQWIETDIPGEGNLARYYRSISKEGGYFSPK ILQAYQTKVKSVDYKPLPIVLGRYNPAIDDGKGNKIYLQTIMNGHWGLPQKTPDYII EEVPLWLGFWGYYNYLKQTRTEAIFPLHMFVVQSKYIQTQQTETPNNFWAFIDNSFI QGKNPWDSVITYSEQKLWFPTVAWQLKTINAICESGPYVPKLDNQTYSTWELATH YSFHFKWGGPQISDQPVEDPGNKNKYDVPDTIKEALQIVNPAKNIAATMFHDWDY RRGCITSTAIKRMQQNLPTDSSLESDSDSEPAPKKKRLLPVLHDPQKKTEKINQCLLS LCEESTCQEQETEENILKLIQQQQQQQQKLKHNLLVLIKDLKVKQRLLQLQTGVLE (SEQ ID NO: 65)
ORF1/1	MPFWWGRRNKFWYGRNYRRKKRRFPKRRKRRFYRRTKYRRPARRRRRRRRKISD QPVEDPGNKNKYDVPDTIKEALQIVNPAKNIAATMFHDWDYRRGCITSTAIKRMQQ NLPTDSSLESDSDSEPAPKKKRLLPVLHDPQKKTEKINQCLLSLCEESTCQEQETEEN ILKLIQQQQQQQQKLKHNLLVLIKDLKVKQRLLQLQTGVLE (SEQ ID NO: 66)
ORF1/2	MPFWWGRRNKFWYGRNYRRKKRRFPKRRKRRFYRRTKYRRPARRRRRRRRKISD QPVEDPGNKNKYDVPDTIKEALQIVNPAKNIAATMFHDWDYRRGCITSTAIKRMQQ NLPTDSSLESDSDSEPAPKKKRLLPVLHDPQKKTEKINQCLLSLCEESTCQEQETEEN ILKLIQQQQQQQQKLKHNLLVLIKDLKVKQRLLQLQTGVLE (SEQ ID NO: 67)

Table B1. 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
AGCCATGTTTGCATTTGAAGTAAAAGCGGGGGGTGAAGGAACAACAGAAC
CACCACAAGGCGCCCAATTAATTGCTAACCTTCCACTCATTGCACTAAGA
TATAATCCACATGAAGACACAGGCCATGGCAATGAAATTTACCTTACATC
AACTTTTAAAGGTACATATGACAAACCTAAAGTTACTGATGCTCTATACT
TTAACAATGTACCCCTGTGGATGGGATTTTATGGCTACTGGGACTTTATA
TTACAAGAAACAAAAAACAAAGGTGTCTTTGATCAACATATGTTTGTTGT
TAAATGTCCTGCCTTAAGGCCCATATCACAAGTCACAAAACAAGTATACT
ACCCACTTGTAGACATGGACTTTTGTTCAGGGAGACTGCCATTTGATGAA
TATTTATCCAAAGACATTAAAAGTCATTGGTATCCCACTGCAGAAAGACA
AACAGTTACAATAAATAATTTTGTTACAGCAGGTCCATACATGCCTAAAT
TTGAACCCACAGACAAAGACAGTACATGGCAATTAAACTATCACTATAAA
TTTTTTTTTAAGTGGGGTGGTCCACAAGTCACAGACCCAACTGTTGAAGA
CCCATGCAGCAGAAACAAATATCCTGTCCCCGATACAATGCAACAAACAA
TACAAATTAAAAACCCTGAAAAGCTGCACCCAGCAACCCTCTTCCATGAC
TGGGACCTTAGAAGGGGCTTCATTACACAAGCAGCTATTAAAAGAATGTC
AGAAAACCTCCAAATTGATTCATCTTTCGAATCTGATGGCACAGAATCAC
CCAAAAAAAAAGAAAAGATGCACCAAAGAAATCCCAACACAAAACCAAAAG
CAAGAAGAGATCCAAGAATGTCTCCTCTCACTCTGCGAAGAGCCTACATG
CCAAGAAGAAACAGAGGACCTCCAGCTCTTCATCCAGCAGCAGCAGCAGC
AGCAGTACAAGCTCAGAAAAAACCTCTTCAAACTCCTCACTCACCTGAAA
AAAGGACAGAGAATAAGTCAACTACAAACGGGACTTTTAGAGTAATACCA
TTTAAACCAGGTTTTGAACAAGAAACAGAAAAAGAACTTGCCATAGCTTT
CTGCAGACCACCTAGAAAATATAAAAATGATCCCCCTTTTTATCCCTGGT
TACCATGGACACCCCTTGTACACTTTAACCTTAATTACAAAGGCTAGGCC
AACACTGTTCACTTAGTGGTGTATGTTTAATAAAGTTTCACCCCCAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAATTGCAAAAATTCG
GCGCTCGCGCGCGCTGCGCGCGCGCGAGCGCCGTCACGCGCCGGCGCTCG
CGCGCCGCGCGTATGTGCTAACACACCACGCACCTAGATTGGGGTGCGCG

CGCTAGCGCGCGCACCCCAATGCGCCCCGCCCTCGTTCCGACCCGCTTGC
GCGGGTCGGACCACTTCGGGCTCGGGGGGGGCGCGCCTGCGGCGCTTTTTT
ACTAAACAGACTCCGAGCCGCCATTTGGCCCCCCCTAAGCTCCGCCCCCC
TCATGAATATTCATAAAGGAAACCACATAATTAGAATTGCCGACCACAAA
CTGCCATATGCTAATTAGTTCCCCTTTTACACAGTAAAAAGGGGAAGTGG
GGGGGCATAGCCCCCCCACACCCCCCGCGGGGGGGGCAGAGCCCCCCCCC
GCACCCCCCCCCTACGTCACAATCCACGCCCCCGCCGCCATCTTGGGTGC
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 C1. 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 B2. Exemplary *Anellovirus* nucleic acid sequence (*Gammatorquevirus*)

Name Ring4.0

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
AAAAATAAAAAAAAAAAAAAAAAAAATAAAAAATTGCAAAAATTCGGCGCT
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*)

Ring 4.0 (*Gammatorquevirus*)	
ORF2	MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPP GHKDRDLTINQILLR DYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGAAEEDHTDALFAAAV ENFER (SEQ ID NO: 887)
ORF2/2	MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPP GHKDRDLTINQILLRDYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGA AEEDHTDALFAAAVENFESGVDHNSMNQKLLTLANKSSMMSPILSTKQYKL KIQKDKTPDLSSMIGTTDEALLKKDLLKECQLTSQLIQISKQLQRKTFPKRKR ELDPNSQSHNKKKRRHCHVSSLSAKKIPSKKQRHKKTSSSSSSSSRSSSSSR ETSSSSTN (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 B3. 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 C3. 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 B4. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus*) – Clade 3

Name	Ring 6.0
Genus/Clade	*Alphatorquevirus* – Clade 3
Accession Number	

Full Sequence: 3828 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
GTGCTACGTCACTAACCTACGTGTCCGTCTCCCATAGGCCGGACACCGTA
TACGTCATACACTTCCTGGGCATGGTCTACGTGATAATATAAGTGGCTGC
ACTTCCGAATGGCTGAGTTTTCCACGCCCGTCCGCAGCGAGGACGCCACG
GAGGGGGATCCGCGCGTCCCGAGGGCGGGTGCCGGAGGTGAGTTTACACA
CCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTATGGGCAAG
GCTCTTAAAAATGCACTTTTCTAGGTGCAGTAGAAAGAAAAGGACATTGT
CACTGCTACCACTGTACCATTCACAGAAAGCTAGGCCATCTGTGACAGGT
ATGTGGAGACCCCCGACTCGAAATGCGTTCAATATTCAACGTGACTGGTT
CTACAGTTGCTTTCACTCCCACGCTTCTATGTGCGGCTGTGCTGATTTTA
TTGGTCATTTCAATCATATCGCTGCTATGCTCGGCCGTCCGGAAGACCAG
AACCCTCCTCCGCCACCCGGGGCTCTGAGACCCCTACCCGCTCTCCCGGC
CTCTTCCGAGGCACCCGGTGATCGAGCGCCATGGCCTATGGGTGGTGGCG
GAGGCGACGGAGGCGCCCGTGGTGGAGGAGGAGATGGCGCCGCTGGAGAC
GCCGTCGGAGACCCCGCAGACGCCGACCTCGTCGCCGCTATCGACGCCGC
AGAACAGTAAGGAGGCGCGGCAGGGGGAGGTGGACTAGAGCACACAGGAG
ATGGCGCCGCAAGGGAAAACGCAGTCGCAAAAAAAAAGATTATTATAAGAC
AATGGCAGCCCAACTACACTCGCAGATGCAACATAGTGGGCTACATGCCT
CTACTAATATGTGGGGAAAATACTGTTGCTACAAACTATGCCACCCACTC
AGACGACAGCTACTACCCCGGACCCTTTGGGGGGGGAATGACTACAGACA
AATTTACTCTAAGAATACTGTATGATGAGTACAAAAGGTTCATGAACTAC
TGGACCTCTTCAAACGAGGACCTAGACCTATGTAGATACCTGGGATGCAC
TCTATATGTGTTTAGACACCCAGAAGTAGACTTTATAATCATTATAAATA
CCTCTCCTCCATTCCTAGACACAGAAATAACAGGGCCTAGCATACACCCA
GGTATGATGGCCCTTAACAAAAGAAGCAGATGGATACCTAGCATAAAAAA
CAGACCAGGCAGAAAGCACTATATAAAGATTAAAGTAGGAGCCCCCCGAA
TGTTCACAGATAAGTGGTACCCCCAAACAGACCTCTGTGACATGACACTC
CTAACGATCTTTGCCAGTGCGGCGGATATGCAATATCCGTTCGGCTCACC
ACTAACTGACACCATAGTTGTGTCATTCCAAGTTCTGCAATCCATGTACA
ACGACTGCCTGAGTGTACTTCCTGATAATTTTGCAGAGACATCAGGCAAA
GGCACCCAACTACATGAGAACATAATACAACATCTGCCCTACTACAACAC
CACACAAACACAAGCACAATTTAAAAGATTTATAGAAAACATGAATGCAA
CAAATGGAGACAATATATGGGCAAGCTACATAAACACAACCAAGTTCTCA
TCCGCAAACACTCCAAAGAATGACACAGGCATAGGAGGCCCTTACACTAC
ATATTCAGACTCATGGTACAAAGGCACAGTATACAATGACAAAATTAAAA
```

CCATACCAATAAAAGCAAGCAAGTTATACTACGAGCAAACCAAAAACCTC
ATTGGCATTACATTCACTGGATCCACACACAGACTCCATTACTGTGGAGG
CCTATACTCCTCCGTATGGCTATCAGCAGGTAGATCCTACTTTGAAACCA
AAGGCCCATACACAGACATAACTTACAACCCCTTTTCAGACAGAGGAGAG
GGTAACATGCTATGGATAGACTGGCTAACTAAAAATGACTCAGTGTACTC
AAAAACAAGTAGCAAGTGTCTTATAGAAAACCTGCCCCTGTGGGCCTCAG
TATACGGATATAAAGAATACTGCAGCAAGGTAACAGGAGACACAAACATA
GAACACAACTGTAGATGTGTTATCAGAAGCCCCTACACAGTACCACAACT
GTTAGACCACAACAATCCCTTCAGAGGATACGTGCCTTATAGCTTCAACT
TTGGAAATGGTAAAATGCCAGGCGGTAGCAGCCTAGTGCCCATTAGAATG
AGAGCCAAGTGGTACCCCACTCTGTTCCACCAAAAAGAAGTTCTAGAAGC
CATAGCACAGGCGGGCCCCTTCGCATACCACTCAGATATTAAAAAAGTGT
CCCTGGGCATAAAGTACAGATTTAAGTGGGTGTGGGGTGGCAACCCCGTG
TCCCAACAGGTTGTTAGAAACCCCTGCAAGACCACCCAAGGTTCCTCGGG
CAATAGAGTGCCTCGATCAATACAAGTCGTTGACCCGCGGTACAACACGC
CAGAACTCACCATACACGCGTGGGACTTCAGACATGGGTTCTTTGGCAGA
AAAGCTATTAAGAGAATGCAAGAACAACCAATACCTCATGACACTTTTTC
AGCAGGGTTCAAGCGCAGTCGCCGAGATACAGAAGCACTCCAATGCAGCC
AAGAAGAGCAACAAAAAGAAAACTTACTTTTCCCAGTCCAGCAGCTCAAG
CGAGTCCCCCCGTGGGAGACCTCGCAAGAGAGCCAAAGCGAGGAAGAAAA
CTCGCAAAAACAGGAGACCCTCTCCCAGCAACTCAGAGACCAGCTGCACA
AGCAGCGGCTCATGGGAGAGCAACTCCGATCGCTCCTCTACCAAATGCAG
AGGGTCCAACAAAATCAACACATAAACCCTATGTTATTGCCAAAGGGTCT
GGCATTAACTTCTATTTCTCACAATGTAATATAGATATGTTTGGTGACCC
CAAACCCTACAAGCCCTCCTCCAATGACTGGAAGGAGGAGTACGAGGCCG
CAAAGTACTGGGACAGACCCCCCAGACGCGACCTGAGGAGCACCCCCTTC
TACCCCTGGGCCCCCACCCCCAAACCATACAATGTCAACTTTGCCCTCAA
CTACAAATAAACGGTGGCCGTGGGAGTTTCACTTGTCGGTGTCTACCTCT
TAAGGTCACTAAGCACTCCGAGCGTAAGCGAGGAGTGCGACCCTTCACCA
AGGGCAACTCCCTCGAAGTCCGGCGCTACGCGCTTCGCGCTGCGCCGGAC
ATCTCGGACCCCCCCTCGACCCGAATCGCTTGCGCGATTCGGACCTGCGG
CCTCGGGGGGGTCGGGGGCTTTACTAAACAGACTCCGAGGTGCCATTGGA
CACTGAGGGGGTGAACAGCAACGAAAGTGAGTGGGGCCAGACTTCGCCAT
AAGGCCTTTATCTTCTTGCCATTTGTCCGCGACCGGGGGTCGCTCCTAGG
CGCGGACCCCGTTTCGGGGTCCTTCCGGGTTCATCGGCGCCGTTCCAGTG
ACGTCACGGGCGCCATGTTAAGTGGCTGTCGCCGAGGATTGACGTCACAG
TTCAAAGGTCATCCTCGGCGGTAACCGCAAACATGGCGGTCAATCTCTTC
CGGGTCAAAGGTCGTGCATACGTCATAAGTCACATGACAGGGGTCCACTT
AAACACGGAAGTAGGCCCCGACATGTGACTCGTCACGTGTGTACACGTCA
CGGCCGCCATTTTGTTTTACAAAATGGCCGACTTCCTTCCTGTTTTTTAA
AAAAAGGCGCGAAAAAACCGTCGGCGGGGGCCGCGCGCTGCGCGCGCGGG
AGGCAATGCCTCCCCCCCCCCGCGCGCATGCGCGCGGGTCCCCCCCCCTC
CGGGGGGCTCCGCCCCCCGGCCCCCCCC (SEQ ID NO: 903)

Annotations:

Putative Domain	*Base range*
TATA Box	85– 92
Cap Site	109 – 116
Transcriptional Start Site	116
5’ UTR Conserved Domain	176 – 246

ORF2	351 – 710
ORF2/2	351 – 706 ; 2360 – 2825
ORF2/3	351 – 706 ; 2556 – 3060
TAIP	373 – 528
ORF1	581 – 2884
ORF1/1	581 – 706 ; 2360 – 2884
ORF1/2	581 – 706 ; 2556 – 2825
Three open-reading frame region	2556 – 2821
Poly(A) Signal	3055 – 3061
GC-rich region	3720 – 3828

Table C4. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*) – Clade 3

Ring 6.0 (*Alphatorquevirus*)	
ORF2	MWRPPTRNAFNIQRDWFYSCFHSHASMCGCADFIGHFNHIAAMLGRPEDQ NPPPPPGALRPLPALPASSEAPGDRAPWPMGGGGGDGGARGGGGDGAAGD AVGDPADADLVAAIDAAEQ (SEQ ID NO: 904)
ORF2/2	MWRPPTRNAFNIQRDWFYSCFHSHASMCGCADFIGHFNHIAAMLGRPEDQ NPPPPPGALRPLPALPASSEAPGDRAPWPMGGGGGDGGARGGGGDGAAGD AVGDPADADLVAAIDAAEQLLETPARPPKVPRAIECLDQYKSLTRGTTRQNS PYTRGTSDMGSLAEKLLRECKNNQYLMTLFQQGSSAVAEIQKHSNAAKKS NKKKTYFSQSSSSSESPRGRPRKRAKARKKTRKNRRPSPSNSETSCTSSGSW ESNSDRSSTKCRGSNKINT (SEQ ID NO: 905)
ORF2/3	MWRPPTRNAFNIQRDWFYSCFHSHASMCGCADFIGHFNHIAAMLGRPEDQ NPPPPPGALRPLPALPASSEAPGDRAPWPMGGGGGDGGARGGGGDGAAGD AVGDPADADLVAAIDAAEQVQAQSPRYRSTPMQPRRATKRKLTFPSPAAQA SPPVGDLAREPKRGRKLAKTGDPLPATQRPAAQAAAHGRATPIAPLPNAEG PTKSTHKPYVIAKGSGINFYFSQCNIDMFGDPKPYKPSSNDWKEEYEAAKY WDRPPRRDLRSTPFYPWAPTPKPYNVNFALNYK (SEQ ID NO: 906)
TAIP	MRSIFNVTGSTVAFTPTLLCAAVLILLVISIISLLCSAVRKTRTLLRHPGL (SEQ ID NO: 907)
ORF1	MAYGWWRRRRRRPWWRRRWRRWRRRRRPRRRRPRRRYRRRRTVRRRGR GRWTRAHRRWRRKGKRSRKKKIIIRQWQPNYTRRCNIVGYMPLLICGENTV

	ATNYATHSDDSYYPGPFGGGMTTDKFTLRILYDEYKRFMNYWTSSNEDLD LCRYLGCTLYVFRHPEVDFIIIINTSPPFLDTEITGPSIHPGMMALNKRSRWIPS IKNRPGRKHYIKIKVGAPRMFTDKWYPQTDLCDMTLLTIFASAADMQYPFG SPLTDTIVVSFQVLQSMYNDCLSVLPDNFAETSGKGTQLHENIIQHLPYYNTT QTQAQFKRFIENMNATNGDNIWASYINTTKFSSANTPKNDTGIGGPYTTYSD SWYKGTVYNDKIKTIPIKASKLYYEQTKNLIGITFTGSTHRLHYCGGLYSSV WLSAGRSYFETKGPYTDITYNPFSDRGEGNMLWIDWLTKNDSVYSKTSSKC LIENLPLWASVYGYKEYCSKVTGDTNIEHNCRCVIRSPYTVPQLLDHNNPFR GYVPYSFNFGNGKMPGGSSLVPIRMRAKWYPTLFHQKEVLEAIAQAGPFAY HSDIKKVSLGIKYRFKWVWGGNPVSQQVVRNPCKTTQGSSGNRVPRSIQVV DPRYNTPELTIHAWDFRHGFFGRKAIKRMQEQPIPHDTFSAGFKRSRRDTEA LQCSQEEQQKENLLFPVQQLKRVPPWETSQESQSEEENSQKQETLSQQLRD QLHKQRLMGEQLRSLLYQMQRVQQNQHINPMLLPKGLALTSISHNVI (SEQ ID NO: 908)
ORF1/1	MAYGWWRRRRRRPWWRRRWRRWRRRRRPRRRRPRRRYRRRRTVVRNPC KTTQGSSGNRVPRSIQVVDPRYNTPELTIHAWDFRHGFFGRKAIKRMQEQPI PHDTFSAGFKRSRRDTEALQCSQEEQQKENLLFPVQQLKRVPPWETSQESQS EEENSQKQETLSQQLRDQLHKQRLMGEQLRSLLYQMQRVQQNQHINPMLL PKGLALTSISHNVI (SEQ ID NO: 909)
ORF1/2	MAYGWWRRRRRRPWWRRRWRRWRRRRRPRRRRPRRRYRRRRTGSSAVA EIQKHSNAAKKSNKKKTYFSQSSSSSESPRGRPRKRAKARKKTRKNRRPSPS NSETSCTSSGSWESNSDRSSTKCRGSNKINT (SEQ ID NO: 910)

Table B5. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus*) – Clade 7

Name	Ring7
Genus/Clade	*Alphatorquevirus* – Clade 7
Accession Number	

Full Sequence: 3815 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
AAGATCGTCACTAACCACGTGACTCCTCTCGCCCAATCAGTGTCTACGTC
```

GTCCATTTCCTGGGCATGGTCTACATCCTGATATAAAGCGATGCACTTCC
GAATGGCTGAGTTTTCCACGCCCGTCCGCGGCGAGATCGCGACGGAGGAG
CGATCGAGCGTCCCGAGGGCGGGTGCCGGAGGTGAGTTTACACACCGCAG
TCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTATGGGCAAGGCTCTT
AAAGCGTACGTCCCCCGCTATGTTTCTCGGCAGGGTGTGGAGGAAACAGA
AAAGGAAAGTGCTTCTGCTGGCTGTGCGAGCTACACAGAAAACATCTTCC
ATGAGTATCTGGCGTCCCCCCCTTGGGAATGTCTCCTACAGGGAGAGAAA
TTGGCTTCAGGCCGTCGAAACATCCCACAGTTCTTTTTGTGGCTGTGGTG
ATTTTATTCTTCATCTTACTAATTTGGCTGCACGCTTTGCTCTCCAGGGG
CCCCCGCCAGAGGGTGGTCCACCTCGGCCGAGGCCGCCGCTCCTGAGAGC
GCTGCCGGCCCCCGAGGTCCGCAGGGAGACGCGCACAGAGAACCGGGGCG
CCTCCGGTGAGCCATGGCCTGGCGATGGTGGTGGCAGAGACGATGGCGCC
GCCGCCGGTGGCCCCGCAGACGGTGGAGACGCCTACGACGCCGGAGACCT
AGACGACCTGTTCGCCGCCGTCGAAGAAGAACAACAGTAAGGAGGCGGAG
GTGGAGGGGCAGACGTGGGCGACGCACATACACCCGACGCGCGGTCAGAC
GCAGACGCAGACCCAGAAAGAGACTTGTACTGACTCAGTGGAGCCCCCAG
ACAGTCAGAAACTGCTCAATAAGGGGCATAGTGCCCATGGTAATATGCGG
ACACACAAAAGCAGGTAGAAACTATGCTATTCATAGCGAGGACTTCACCA
CACAGATACAACCCTTCGGGGGCAGTTTCAGCACGACCACCTGGTCCCTA
AAAGTGCTGTGGGACGAGCACCAGAAATTCCAGAACAGATGGTCCTACCC
AAACACACAACTAGACCTGGCCAGATACAGAGGGGTCACCTTCTGGTTCT
ACAGAGACCAGAAAACAGACTATATAGTACAGTGGAGTAGGAATCCCCCT
TTTAAACTCAATAAATACAGCAGTGCCATGTACCACCCGGGCATGATGAT
GCAGGCCAAAAGGAAACTAGTTGTACCTAGTTTCCAGACCAGACCCAAAG
GCAAGAAGAGATACAGAGTCACAATAAAACCCCCTAACATGTTTGCTGAC
AAGTGGTACACTCAAGAGGACCTGTGTCCGGTACCTCTTGTGCAAATTGT
GGTTTCTGCGGCGAGCCTGCTACATCCGTTCTGCCCACCACAAACGAACA
ACCCTTGCATCACCTTCCAGGTTTTGAAAGACATATATGATGAATGCATA
GGAGTTAACGAAACTATGAAAGATAAGTATAAGAAATTACAAACAACACT
ATACACCACTTGCACATACTATCAAACAACACAAGTACTGGCACAGCTAT
CTCCTGCCTTTCAACCTGCTATGAAACCTACTACTACACAATCAGCAGCT
ACAGCGACAACACTAGGAAACTATGTACCAGAGTTAAAGTACAACAATGG
CTCTTTTCACACAGGACAAAACGCAGTATTCGGCATGTGCTCATACAAAC
CAACAGACAGCATAATGACAAAAGCTAATGGCTGGTTTTGGCAAAACCTA
ATGGTAGACAACAACCTACATAGTTCTTATGGCAAGGCAACATTAGAATG
CATGGAGTATCACACAGGCATATACAGCTCTATATTTCTAAGTCCACAAA
GATCTTTAGAATTCCCAGCAGCATACCAAGACGTTACATACAACCCTAAC
TGTGATAGAGCAGTTGGAAACGTAGTTTGGTTTCAGTACAGCACTAAAAT
GGATACAAATTTTGATGAAACAAAATGTAAATGTGTCCTTAAAAACATTC
CACTGTGGGCGGCCTTCAATGGCTACTCAGACTTTATAATGCAAGAACTC
AGCATAAGTACAGAAATCCACAACTTTGGCATAGTGTGCTTTCAGTGCCC
GTACACTTTTCCCCCCTGTTTCAATAAAAACAAACCCCTAAAGGGGTACG
TGTTCTATGACACCACCTTTGGTAATGGAAAAATGCCAGACGGATCGGGG
CACGTACCCATCTACTGGCAGCAGAGATGGTGGATCAGACTAGCCTTCCA
GGTCCAGGTCATGCATGACTTTGTACTAACAGGCCCCTTTAGCTACAAAG
ATGACCTAGCAAACACCACACTCACAGCCAGATACAAATTTAAATTCAAA
TGGGGCGGCAATATCATCCCTGAACAGATTATCAAGAACCCGTGTCACAG
AGAGCAGTCCCTCGCTTCCTATCCCGATAGACAACGTCGCGACCTACAAG
TTGTTGACCCATCAACCATGGGCCCGATCTACACCTTCCACACATGGGAC
TGGCGACGGGGGCTTTTTGGTGCAGATGCTATCCAGAGAGTGTCACAAAA
ACCGGGAGATGCTCTCCGCTTTACAAACCCTTTCAAGAGACCCAGATATC
TTCCCCCGACAGACAGAGAAGACTACCGACAAGAAGAAGACTTCGCTTTA
CAGGAAAAAAGACGGCGCACATCCACAGAAGAAGCCCAGGACGAGGAGAG
CCCCCCGGAAAGCGCGCCGCTCCTACAGCAGCAGCAGCAGCAGCGGCAGC
TCTCAGTCCACCTCGCGGAGCAGCAGCGACTCGGAGTCCAACTCCGATAC
ATCCTCCAAGAAGTCCTCAAAACGCAAGCGGGTCTCCACCTAAACCCCCT
ATTATTAGGCCCGCCACAAACAAGGTCTATCTCTTTGAGCCCTCCAAAGG

CCTACTCCCCATAGTAGGAAAAGAGGCCTGGGAGGACGAGTACTGCACCT
GCAAGTACTGGGATCGCCCTCCCAGAACCAACCACCTAGACATCCCCACT
TATCCCTGGATGCCCACAAACTTCAAAGTCAGCTTCAAACTTGGATTTAA
ACCCTAAATAAAAATACAAGGCCGTACACTGTTCACTTGTCGGTGTCTAC
CTCTATAAGTCACTAAGCACTCCGAGCGCAGCGAGGAGTGCGACCCTCAG
CGGTGGGTGCAACGCCCTCGGCGGCCGCGCGCTACGCCTTCGGCTGCGCG
CGGCACCTCGGACCCCCGCTCGTGCTGACACGCTCGCGCGTGTCAGACCA
CTTCGGGCTCGCGGGGGTCGGGAATTTTGCTAAACAGACTCCGAGTTGCT
CTTGGACACTGTAGCTGTGAATCAGTAACGAAAGTGAGTGGGGCCAGACT
TCGCCATAAGGCCTTTATCTTCTTGCCATTGGTCCGTCTCGGGGGTCGCC
ATAGGCTTCGGGCTCGGTTTTAGGCCTTCCGGACTACCAAAATGGCGGAT
TCCGTGACGTCATGGCCGCCATTTTAAGTAAGGCGGAACAGGCTGTCACC
CCGTGTCAAAGTTCAGGGGTCAGCCTTCCGCTTTACACAAAATGGAGGTC
AATATCTTCCGGGTCAAAGGTCGCTACCGCGTCATAAGTCACGTGGGGAA
GGCTGCTGTGAATCCGGAAGTAGCTGACCCACGTGACTTGTCACGTGACT
AGCACGTCACGGCAGCCATTTTGAATCACAAAATGGCCGACTTCCTTCCT
CTTTTTTAAAAATAACGGCCCGGCGGCGGCGCGCGCTTCGCGCCGCTC
CGCCCCCCCCGCGCATGCGCGGGACCCCCCCCCGCGGGGGGCTCCGCCCC
CCGGTCCCCCCCCCG (SEQ ID NO: 911)

Annotations:

Putative Domain	*Base range*
TATA Box	82– 87
Cap Site	103 – 110
Transcriptional Start Site	110
5’ UTR Conserved Domain	170 – 240
ORF2	351 – 740
ORF2/2	351 – 737 ; 2378 – 2843
ORF2/3	351 – 737 ; 2526 – 3057
TAIP	379 - 543
ORF1	614 – 2911
ORF1/1	614 – 737 ; 2378 – 2911
ORF1/2	614 – 737 ; 2526 – 2843
Three open-reading frame region	2526 – 2840
Poly(A) Signal	3056 – 3062
GC-rich region	3716 – 3815

Table C5. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*) – Clade 7

Ring7.0 (*Alphatorquevirus*)	
ORF2	MSIWRPPLGNVSYRERNWLQAVETSHSSFCGCGDFILHLTNLAARFALQGPP

	PEGGPPRPRPPLLRA LPAPEVRRETRTENRGASGEPWPGDGGGRDDGAAAGGPADGGDAYDAGD LDDLFAAVEEEQQ (SEQ ID NO: 912)
ORF2/2	MSIWRPPLGNVSYRERNWLQAVETSHSSFCGCGDFILHLTNLAARFALQGPP PEGGPPRPRPPLLRA LPAPEVRRETRTENRGASGEPWPGDGGGRDDGAAAGGPADGGDAYDAGD LDDLFAAVEEEQQLSRTRVTESSPSLPIPIDNVATYKLLTHQPWARSTPSTHG TGDGGFLVQMLSRECHKNREMLSALQTLSRDPDIFPRQTEKTTDKKKTSLY RKKDGAHPQKKPRTRRAPRKARRSYSSSSSSGSSQSTSRSSSDSESNSDTSSK KSSKRKRVST (SEQ ID NO: 913)
ORF2/3	MSIWRPPLGNVSYRERNWLQAVETSHSSFCGCGDFILHLTNLAARFALQGPP PEGGPPRPRPPLLRA LPAPEVRRETRTENRGASGEPWPGDGGGRDDGAAAGGPADGGDAYDAGD LDDLFAAVEEEQQCYPESVTKTGRCSPLYKPFQETQISSPDRQRRLPTRRRLR FTGKKTAHIHRRSPGRGEPPGKRAAPTAAAAAAAALSPPRGAAATRSPTPIH PPRSPQNASGSPPKPPIIRPATNKVYLFEPSKGLLPIVGKEAWEDEYCTCKYW DRPPRTNHLDIPTYPWMPTNFKVSFKLGFKP (SEQ ID NO: 914)
TAIP	MSPTGREIGFRPSKHPTVLFVAVVILFFILLIWLHALLSRGPRQRVVHLGRGR RS (SEQ ID NO: 915)
ORF1	MAWRWWWQRRWRRRRWPRRRWRRLRRRRPRRPVRRRRRRTTVRRRRW RGRRGRRTYTRRAVRRRRRPRKRLVLTQWSPQTVRNCSIRGIVPMVICGHT KAGRNYAIHSEDFTTQIQPFGGSFSTTTWSLKVLWDEHQKFQNRWSYPNTQ LDLARYRGVTFWFYRDQKTDYIVQWSRNPPFKLNKYSSAMYHPGMMMQA KRKLVVPSFQTRPKGKKRYRVTIKPPNMFADKWYTQEDLCPVPLVQIVVSA ASLLHPFCPPQTNNPCITFQVLKDIYDECIGVNETMKDKYKKLQTTLYTTCT YYQTTQVLAQLSPAFQPAMKPTTTQSAATATTLGNYVPELKYNNGSFHTGQ NAVFGMCSYKPTDSIMTKANGWFWQNLMVDNNLHSSYGKATLECMEYHT GIYSSIFLSPQRSLEFPAAYQDVTYNPNCDRAVGNVVWFQYSTKMDTNFDE TKCKCVLKNIPLWAAFNGYSDFIMQELSISTEIHNFGIVCFQCPYTFPPCFNK NKPLKGYVFYDTTFGNGKMPDGSGHVPIYWQQRWWIRLAFQVQVMHDFV LTGPFSYKDDLANTTLTARYKFKFKWGGNIIPEQIIKNPCHREQSLASYPDRQ

	RRDLQVVDPSTMGPIYTFHTWDWRRGLFGADAIQRVSQKPGDALRFTNPFK RPRYLPPTDREDYRQEEDFALQEKRRRTSTEEAQDEESPPESAPLLQQQQQQ RQLSVHLAEQQRLGVQLRYILQEVLKTQAGLHLNPLLLGPPQTRSISLSPPKA YSP (SEQ ID NO: 916)
ORF1/1	MAWRWWWQRRWRRRRWPRRRWRRLRRRRPRRPVRRRRRRRTTIIKNPCHR EQSLASYPDRQRRDLQVVDPSTMGPIYTFHTWDWRRGLFGADAIQRVSQKP GDALRFTNPFKRPRYLPPTDREDYRQEEDFALQEKRRRTSTEEAQDEESPPE SAPLLQQQQQQRQLSVHLAEQQRLGVQLRYILQEVLKTQAGLHLNPLLLGP PQTRSISLSPPKAYSP (SEQ ID NO: 917)
ORF1/2	MAWRWWWQRRWRRRRWPRRRWRRLRRRRPRRPVRRRRRRRTTMLSRECH KNREMLSALQTLSRDPDIFPRQTEKTTDKKKTSLYRKKDGAHPQKKPRTRR APRKARRSYSSSSSSGSSQSTSRSSSDSESNSDTSSKKSSKRKRVST (SEQ ID NO: 918)

In some embodiments, an anellosome 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 anellosome 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 anellosome comprises an *Anellovirus* genome, e.g., as identified according to the method described in Example 9. In some embodiments, an anellosome comprises an *Anellovirus* sequence, or a portion thereof, as described in Example 13.

In some embodiments, an anellosome comprises a genetic element comprising a consensus *Anellovirus* motif, e.g., as shown in Table 19. In some embodiments, an anellosome comprises a genetic element comprising a consensus *Anellovirus* ORF1 motif, e.g., as shown in Table 19. In some embodiments, an anellosome comprises a genetic element comprising a consensus *Anellovirus* ORF1/1 motif, e.g., as shown in Table 19. In some embodiments, an anellosome comprises a genetic element comprising a consensus *Anellovirus* ORF1/2 motif, e.g., as shown in Table 19. In some embodiments, an anellosome comprises a genetic element comprising a consensus *Anellovirus* ORF2/2 motif, e.g., as shown in Table 19. In some embodiments, an anellosome comprises a genetic element comprising a consensus *Anellovirus* ORF2/3 motif, e.g., as shown in Table 19. In some embodiments, an anellosome comprises a genetic element comprising a consensus *Anellovirus* ORF2t/3 motif, e.g., as shown in Table 19. In some embodiments, X, as shown in Table 19, indicates any amino acid. In some embodiments, Z,

as shown in Table 19, indicates glutamic acid or glutamine. In some embodiments, B, as shown in Table 19, indicates aspartic acid or asparagine. In some embodiments, J, as shown in Table 19, indicates leucine or isoleucine.

Table 19. Consensus motifs in open reading frames (ORFs) of *Anelloviruses*

Consensus Threshold	Open Reading Frame	Position	Motif	SEQ ID NO:
50	ORF1	79	LIJRQWQPXXIRRCXIXGYXPLIXC	68
50	ORF1	111	NYXXHXD	69
50	ORF1	135	FSLXXLYDZ	70
50	ORF1	149	NXWTXSNXDLDLCRYXGC	71
50	ORF1	194	TXPSXHPGXMXLXKHK	72
50	ORF1	212	IPSLXTRPXG	73
50	ORF1	228	RIXPPXLFXDKWYFQXDL	74
50	ORF1	250	LLXIXATA	75
50	ORF1	260	LXXPFXSPXTD	76
50	ORF1	448	YNPXXDKGXGNXIW	77
50	ORF1	519	CPYTZPXL	78
50	ORF1	542	XFGXGXMP	79
50	ORF1	569	HQXEVXEX	80
50	ORF1	600	KYXFXFXWGGNP	81
50	ORF1	653	HSWDXRRG	82
50	ORF1	666	AIKRXQQ	83
50	ORF1	750	XQZQXXLR	84
50	ORF1/1	73	PRXJQXXDP	85
50	ORF1/1	91	HSWDXRRG	86
50	ORF1/1	105	AIKRXQQ	87
50	ORF1/1	187	QZQXXLR	88
50	ORF1/2	97	KXKRRRR	89
50	ORF2/2	158	PIXSLXXYKXXTR	90

50	ORF2/2	189	LAXQLLKECXKN	91
50	ORF2/3	39	HLNXLA	92
50	ORF2/3	272	DRPPR	93
50	ORF2/3	281	DXPFYPWXP	94
50	ORF2/3	300	VXFKLXF	95
50	ORF2t/3	4	WXPPVHBVXGIERXW	96
50	ORF2t/3	37	AKRKLX	97
50	ORF2t/3	140	PSSXDWXXEY	98
50	ORF2t/3	156	DRPPR	99
50	ORF2t/3	167	PFYPW	100
50	ORF2t/3	183	NVXFKLXF	101
50	ORF1	84	JXXXXWQPXXXXXCXIXGXXXJWQP	102
50	ORF1	149	NXWXXXNXXXXLXRY	103
50	ORF1	448	YNPXXDXG	104

ORF1 molecules

In some embodiments, the anellosome 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, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10), 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, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10). 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 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF1 protein sequence as shown in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10. 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 encapsidate a nucleic acid molecule (e.g., a genetic element as described herein). In some embodiments, a plurality of ORF1 molecules may form a multimer, e.g., to produce a proteinaceous exterior. In some embodiments, the multimer may be a homomultimer. In other embodiments, the multimer may be a heteromultimer (e.g., comprising a plurality of distinct ORF1 molecules). It is also contemplated that an ORF1 molecule may have replicase activity.

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.

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.

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).

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). 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 (e.g., as shown in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20-37, or D1-D10).

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

Exemplary $YNPX^2DXGX^2N$ (SEQ ID NO: 829) motif-flanking secondary structures are described in Example 47 and Figure 48. 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. 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, 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 and Figure 47. 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 Figure 47.

Exemplary ORF1 Sequences

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 20-37, or D1-D10). In some embodiments, an anellosome 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 20-37, or D1-D10. In some embodiments, an anellosome 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 20-37, or D1-D10.

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 20-37, or D1-D10), 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 20-37, or D1-D10). 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.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 20 (e.g., amino acids 1 – 66 of Table 20). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 21. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 20 (e.g., amino acids 67 – 277 of Table 20). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 21. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 20 (e.g., amino acids 278 – 347 of Table 20). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 21. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 20 (e.g., amino acids 348 – 513 of Table 20). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 21. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or

100% sequence identity to the CTD amino acid sequence of Table 20 (e.g., amino acids 513 – 680 of Table 20). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table 21.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 22 (e.g., amino acids 1 – 69 of Table 22). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 23. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 22 (e.g., amino acids 70 – 279 of Table 22). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 23. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 22 (e.g., amino acids 280 – 411 of Table 22). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 23. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 22 (e.g., amino acids 412 – 578 of Table 22). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 23. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table 22 (e.g., amino acids 579 – 747 of Table 22). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table 23.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%

sequence identity to the Arg-rich region amino acid sequence of Table 24 (e.g., amino acids 1 – 68 of Table 24). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 25. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 24 (e.g., amino acids 69 – 280 of Table 24). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 25. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 24 (e.g., amino acids 281 – 413 of Table 24). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 25. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 24 (e.g., amino acids 414 – 479 of Table 24). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 25. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table 24 (e.g., amino acids 580 – 743 of Table 24). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table 25.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 26 (e.g., amino acids 1 – 74 of Table 26). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 27. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino

acid sequence of Table 26 (e.g., amino acids 75 – 284 of Table 26). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 27. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 26 (e.g., amino acids 285 – 445 of Table 26). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 27. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 26 (e.g., amino acids 446 – 611 of Table 26). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 27. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table 26 (e.g., amino acids 612 – 780 of Table 26). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table 27.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 28 (e.g., amino acids 1 – 75 of Table 28). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 29. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 28 (e.g., amino acids 75 – 284 of Table 28). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 29. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 28 (e.g., amino acids 285 – 432 of

Table 28). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 29. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 28 (e.g., amino acids 433 – 599 of Table 28). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 29. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table 28 (e.g., amino acids 600 – 780 of Table 28). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table 29.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 30 (e.g., amino acids 1 – 77 of Table 30). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 31. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 30 (e.g., amino acids 78 – 286 of Table 30). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 31. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 30 (e.g., amino acids 287 – 416 of Table 30). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 31. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 30 (e.g., amino acids 417 – 585 of Table 30). In embodiments, the one or more

Anellovirus ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 31. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table 30 (e.g., amino acids 586 – 746 of Table 30). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table 31.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 32 (e.g., amino acids 1 – 74 of Table 32). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 33. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 32 (e.g., amino acids 75 – 286 of Table 32). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 33. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 32 (e.g., amino acids 287 – 428 of Table 32). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 33. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 32 (e.g., amino acids 429 – 595 of Table 32). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 33. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table 32 (e.g., amino acids 596 – 765 of Table 32). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid

sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table 33.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 34 (e.g., amino acids 1 – 38 of Table 34). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 35. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 34 (e.g., amino acids 39 – 246 of Table 34). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 35. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 34 (e.g., amino acids 247 – 374 of Table 34). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 35. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 34 (e.g., amino acids 375 – 537 of Table 34). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 35. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table 34 (e.g., amino acids 538 – 666 of Table 34). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table 35.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 36 (e.g., amino acids 1 – 57 of Table 36). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid

sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table 37. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 36 (e.g., amino acids 58 – 259 of Table 36). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table 37. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 36 (e.g., amino acids 260 – 351 of Table 36). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table 37. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 36 (e.g., amino acids 352 – 510 of Table 36). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table 37. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table 36 (e.g., amino acids 511 – 673 of Table 36). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table 37.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D1 (e.g., amino acids 1 – 66 of Table D1). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D2. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D1 (e.g., amino acids 67 – 277 of Table D1). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%,

80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D2. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table D1 (e.g., amino acids 278 – 347 of Table D1). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table D2. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table D1 (e.g., amino acids 348 – 513 of Table D1). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table D2. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table D1 (e.g., amino acids 513 – 680 of Table D1). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table D2.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D3 (e.g., amino acids 1 – 66 of Table D3). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D4. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D3 (e.g., amino acids 67 – 277 of Table D3). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D4. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table D3 (e.g., amino acids 278 – 347 of Table D3). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%

sequence identity to the HVR amino acid sequence of Table D4. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table D3 (e.g., amino acids 348 – 513 of Table D3). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table D4. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table D3 (e.g., amino acids 513 – 680 of Table D3). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table D4.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D5 (e.g., amino acids 1 – 66 of Table D5). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D6. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D5 (e.g., amino acids 67 – 277 of Table D5). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D6. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table D5 (e.g., amino acids 278 – 347 of Table D5). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table D6. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table D5 (e.g., amino acids 348 – 513 of Table D5). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid

sequence of Table D6. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table D5 (e.g., amino acids 513 – 680 of Table D5). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table D6.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D7 (e.g., amino acids 1 – 57 of Table D7). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D8. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D7 (e.g., amino acids 58 – 259 of Table D7). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D8. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table D7 (e.g., amino acids 260 – 351 of Table D7). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table D8. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table D7 (e.g., amino acids 352 – 510 of Table D7). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table D8. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table D7 (e.g., amino acids 511 – 673 of Table D7). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table D8.

In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D9 (e.g., amino acids 1 – 57 of Table D9). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the Arg-rich region amino acid sequence of Table D10. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D9 (e.g., amino acids 58 – 259 of Table D9). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the jelly-roll region amino acid sequence of Table D10. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table D9 (e.g., amino acids 260 – 351 of Table D9). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the HVR amino acid sequence of Table D10. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table D9 (e.g., amino acids 352 – 510 of Table D9). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the N22 domain amino acid sequence of Table D10. In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD amino acid sequence of Table D9 (e.g., amino acids 511 – 673 of Table D9). In embodiments, the one or more *Anellovirus* ORF1 subsequences comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the CTD region amino acid sequence of Table D10.

Table 20. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 1)

Name	CT30F
Genus/Clade	*Alphatorquevirus*, Clade 1
Strain Accesion Number	AB064597.1

Protein Accession Number ANQ39351.1

Full Sequence: 680 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
TAWWWGRWRRRWRRRRPWRPRLRRRRARRAFPRRRRRRFVSRRWRRPYRR
RRRRGRRRRRRRRRHKPTLVLRQWQPDVIRHCKITGRMPLIICGKGSTQF
NYITHADDITPRGASYGGNFTNMTFSLEAIYEQFLYHRNRWSASNHDLEL
CRYKGTTLKLYRHPDVDYIVTYSRTGPFEISHMTYLSTHPLLMLLNKHHI
VVPSLKTKPRGRKAIKVRIRPPKLMNNKWYFTRDFCNIGLFQLWATGLEL
RNPWLRMSTLSPCIGFNVLKNSIYTNLSNLPQHREDRLNIINNTLHPHDI
TGPNNKKWQYTYTKLMAPIYYSANRASTYDLLREYGLYSPYYLNPTRINL
DWMTPYTHVRYNPLVDKGFGNRIYIQWCSEADVSYNRTKSKCLLQDMPLF
FMCYGYIDWAIKNTGVSSLARDARICIRCPYTEPQLVGSTEDIGFVPITE
TFMRGDMPVLAPYIPLSWFCKWYPNIAHQKEVLEAIISCSPFMPRDQGMN
GWDITIGYKMDFLWGGSPLPSQPIDDPCQQGTHPIPDPDKHPRLLQVSNP
KLLGPRTVFHKWDIRRGQFSKRSIKRVSEYSSDDESLAPGLPSKRNKLDS
AFRGENPEQKECYSLLKALEEEETPEEEEPAPQEKAQKEELLHQLQLQRR
HQRVLRRGLKLVFTDILRLRQGVHWNPELT (SEQ ID NO: 173)
```

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 66
Jelly-roll domain	67 - 277
Hypervariable Region	278 - 347
N22	348 – 513
C-terminal Domain	513 - 680

Table 21. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 1)

TTV-CT30F-ORF1 (*Alphatorquevirus* Clade 1)	
Arg-Rich Region	TAWWWGRWRRRWRRRRPWRPRLRRRRARRAFPRRRRRRFVSRRWR RPYRRRRRRGRRRRRRRRRHK (SEQ ID NO: 174)
Jelly-roll Domain	PTLVLRQWQPDVIRHCKITGRMPLIICGKGSTQFNYITHADDITPRGASY GGNFTNMTFSLEAIYEQFLYHRNRWSASNHDLELCRYKGTTLKLYRHP DVDYIVTYSRTGPFEISHMTYLSTHPLLMLLNKHHIVVPSLKTKPRGRK AIKVRIRPPKLMNNKWYFTRDFCNIGLFQLWATGLELRNPWLRMSTLS PCIGFNVLKNSIYTNL (SEQ ID NO: 175)
Hypervariable	SNLPQHREDRLNIINNTLHPHDITGPNNKKWQYTYTKLMAPIYYSANR

domain	ASTYDLLREYGLYSPYYLNPTR (SEQ ID NO: 176)
N22	INLDWMTPYTHVRYNPLVDKGFGNRIYIQWCSEADVSYNRTKSKCLL QDMPLFFMCYGYIDWAIKNTGVSSLARDARICIRCPYTEPQLVGSTEDI GFVPITETFMRGDMPVLAPYIPLSWFCKWYPNIAHQKEVLEAIISCSPFM PRDQGMNGWDITIGYKMDFL (SEQ ID NO: 177)
C-terminal domain	WGGSPLPSQPIDDPCQQGTHPIPDPDKHPRLLQVSNPKLLGPRTVFHKW DIRRGQFSKRSIKRVSEYSSDDESLAPGLPSKRNKLDSAFRGENPEQKEC YSLLKALEEEETPEEEEPAPQEKAQKEELLHQLQLQRRHQRVLRRGLK LVFTDILRLRQGVHWNPELT (SEQ ID NO: 178)

Table 22. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus,* Clade 2)

Name	TTV-P13-1
Genus/Clade	*Alphatorquevirus*, Clade 2
Accession Number	KT163896.1
Protein Accession Number	ANQ39351.1

Full Sequence: 747 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MAYWWGRRRRWRRWRRRRRPLRRRRRWRRRRRWPRRRRWRRRRRRARPAR
RYRRRRGRRRVRRRRRPQKLVLTQWNPQTVRKCVIRGFLPLFFCGQGAYH
RNFTDHYDDVFPKGPSGGGHGSMVFNLSFLYQEFKKHHNKWSRSNLDFDL
VRYKGTVIKLYRHQDFDYIVWISRTPPFQESLLTVMTHQPSVMLQAKKCI
IVKSYRTHPGGKPYVTAKVRPPRLLTDKWYFQSDFCNVPLFSLQFALAEL
RFPICSPQTDTNCINFLVLDDIYYKFLDNKPKQSSDPNDENRIKFWHGLW
STMRYLNTTYINTLFPGTDSLVAAKDTDNSVNKYPSTATKQPYKDSQYMQ
NIWNTSKIHALYTWVAETNYKRLQAYYTQTYGGYQRQFFTGKQYWDYRVG
MFSPAFLSPSRLNPQNPGAYTEVSYNPWTDEGTGNVVCLQYLTKETSDYK
PGGGSKFCIEGVPLWAALVGYVDMCKKEGKDPGIRLNCLLLVKCPYTKPQ
LYDKKNPEKLFVPYSYNFGHGKMPGGDKYIPIEFKDRWYPCLLHQEEWIE
DIVRSGPFVPKDMPSSVTCMMRYSSLFNWGGNIIQEQAVEDPCKKGTFVV
PGTSGIARILQVSNPAKQTPTTTWHSWDWRRSLFTETGLKRMREQQPYDE
LSYTGPKKPKLSLPAGPAVPGAAVASSWWETKQVTSPDVSETETEAEAHQ
EEETEPEEGVQLQQLWEQQLLQKRQLGVVFQQLLRLRQGAEIHPGLV
(SEQ ID NO: 179)
```

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 69

Jelly-roll domain	70 - 279
Hypervariable Region	280 - 411
N22	412 – 578
C-terminal Domain	579 - 747

Table 23. Exemplary *Anellovirus* ORF1 amino *acid subsequence* (*Alphatorquevirus,* Clade 2)

TTV-P13-1-ORF1 (*Alphatorquevirus* Clade 2)	
Arg-Rich Region	MAYWWGRRRRWRRWRRRRRPLRRRRRWRRRRRWPRRRRWRRRRR RARPARRYRRRRGRRRVRRRRRPQK (SEQ ID NO: 180)
Jelly-roll Domain	LVLTQWNPQTVRKCVIRGFLPLFFCGQGAYHRNFTDHYDDVFPKGPSG GGHGSMVFNLSFLYQEFKKHHNKWSRSNLDFDLVRYKGTVIKLYRHQ DFDYIVWISRTPPFQESLLTVMTHQPSVMLQAKKCIIVKSYRTHPGGKP YVTAKVRPPRLLTDKWYFQSDFCNVPLFSLQFALAELRFPICSPQTDTN CINFLVLDDIYYKFLDN (SEQ ID NO: 181)
Hypervariable domain	KPKQSSDPNDENRIKFWHGLWSTMRYLNTTYINTLFPGTDSLVAAKDT DNSVNKYPSTATKQPYKDSQYMQNIWNTSKIHALYTWVAETNYKRLQ AYYTQTYGGYQRQFFTGKQYWDYRVGMFSPAFLSPSR (SEQ ID NO: 182)
N22	LNPQNPGAYTEVSYNPWTDEGTGNVVCLQYLTKETSDYKPGGGSKFCI EGVPLWAALVGYVDMCKKEGKDPGIRLNCLLLVKCPYTKPQLYDKK NPEKLFVPYSYNFGHGKMPGGDKYIPIEFKDRWYPCLLHQEEWIEDIVR SGPFVPKDMPSSVTCMMRYSSLFN (SEQ ID NO: 183)
C-terminal domain	WGGNIIQEQAVEDPCKKGTFVVPGTSGIARILQVSNPAKQTPTTTWHS WDWRRSLFTETGLKRMREQQPYDELSYTGPKKPKLSLPAGPAVPGAA VASSWWETKQVTSPDVSETETEAEAHQEEETEPEEGVQLQQLWEQQL LQKRQLGVVFQQLLRLRQGAEIHPGLV (SEQ ID NO: 184)

Table 24. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 3)

Name	TTV-tth8
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 25. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 3)

TTV-tth8-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 26. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 4)

Name	TTV-HD20a
Genus/Clade	*Alphatorquevirus*, Clade 4
Accession Number	FR751492.1
Protein Accession Number	NA

Full Sequence: 780 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MAWWGWRRRWWRPKRRWRWRRARRRRRVPARRPRRAFRRYRTRTVRRRRR
GRRRGYRRRYRLRRYARRRFRRKKIVLTQWNPQTTRKCIIRGMMPVLWAG
MGTGGRNYAVRSDDYVVNKGFGGSFATETFSLKVLYDQFQRGFNRWSHTN
EDLDLARYRGCRWTFYRHKDTDFIVYFTNNPPMKTNQFSAPLTTPGMLMR
SKYKVLIPSFQTRPKGRKTVTVKIRPPKLFQDKWYTQQDLCSVPLVQLNV
TAADFTHPFGSPLTETPCVEFQVLGDLYNTCLNIDLPQFSELGEITSAYS
```

KPNSNNLKELYKELFTKATSGHYWQTFITNSMVRAHIDADKAKEAQRAST
TPSYNNDPFPTIPVKSEFAQWKKKFTDTRDSPFLFATYHPEAIKDTIMKM
RENNFKLETGPNDKYGDYTAQYQGNTHMLDYYLGFYSPIFLSDGRSNVEF
FTAYRDIVYNPFLDKAQGNMVWFQYHTKTDNKFKKPECHWEIKDMPLWAL
LNGYVDYLETQIQYGDLSKEGKVLIRCPYTKPALVDPRDDTAGYVVYNRN
FGRGKWIDGGGYIPLHERTKWYVMLRYQTDVFHDIVTCGPWQYRDDNKNS
QLVAKYRFSFIWGGNTVHSQVIRNPCKDNQVSGPRRQPRDIQVVDPQRIT
PPWVLHSFDQRRGLFTETALRRLLQEPLPGEYAVSTLRTPLLFLPSEYQR
EDGAAESASGSPAKRPRIWSEESQTETISSEENPAETTRELLQRKLREQR
ALQFQLQHFAVQLAKTQANLHVNPLLSFPQ (SEQ ID NO: 191)

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 74
Jelly-roll domain	75 - 284
Hypervariable Region	285 - 445
N22	446 – 611
C-terminal Domain	612 - 780

Table 27. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 4)

TTV- HD20a -ORF1 (*Alphatorquevirus* Clade 4)	
Arg-Rich Region	MAWWGWRRRWWRPKRRWRWRRARRRRRVPARRPRRAFRRYRTRT VRRRRRGRRRGYRRRYRLRRYARRRFRRKK (SEQ ID NO: 192)
Jelly-roll Domain	IVLTQWNPQTTRKCIIRGMMPVLWAGMGTGGRNYAVRSDDYVVNKG FGGSFATETFSLKVLYDQFQRGFNRWSHTNEDLDLARYRGCRWTFYR HKDTDFIVYFTNNPPMKTNQFSAPLTTPGMLMRSKYKVLIPSFQTRPKG RKTVTVKIRPPKLFQDKWYTQQDLCSVPLVQLNVTAADFTHPFGSPLT ETPCVEFQVLGDLYNTCLNI (SEQ ID NO: 193)
Hypervariable domain	DLPQFSELGEITSAYSKPNSNNLKELYKELFTKATSGHYWQTFITNSMV RAHIDADKAKEAQRASTTPSYNNDPFPTIPVKSEFAQWKKKFTDTRDSP FLFATYHPEAIKDTIMKMRENNFKLETGPNDKYGDYTAQYQGNTHML DYYLGFYSPIFLSDGR (SEQ ID NO: 194)
N22	SNVEFFTAYRDIVYNPFLDKAQGNMVWFQYHTKTDNKFKKPECHWEI KDMPLWALLNGYVDYLETQIQYGDLSKEGKVLIRCPYTKPALVDPRD DTAGYVVYNRNFGRGKWIDGGGYIPLHERTKWYVMLRYQTDVFHDI VTCGPWQYRDDNKNSQLVAKYRFSFI (SEQ ID NO: 195)

C-terminal domain	WGGNTVHSQVIRNPCKDNQVSGPRRQPRDIQVVDPQRITPPWVLHSFD QRRGLFTETALRRLLQEPLPGEYAVSTLRTPLLFLPSEYQREDGAAESA SGSPAKRPRIWSEESQTETISSEENPAETTRELLQRKLREQRALQFQLQH FAVQLAKTQANLHVNPLLSFPQ (SEQ ID NO: 196)

Table 28. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 5)

Name	TTV-16 (TUS01)
Genus/Clade	*Alphatorquevirus*, Clade 5
Accession Number	AB017613.1
Protein Accession Number	BAA82454.1

Full Sequence: 761 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MAYWFRRWGWRPRRRWRRWRRRRRRLPRRRTRRAVRGLGRRRKPRVRRRR
RTRRRTYRRGWRRRRYIRRGRRKKKLILTQWNPAIVKRCNIKGGLPIIIC
GEPRAAFNYGYHMEDYTPQPFPFGGGMSTVTFSLKALYDQYLKHQNRWTF
SNDQLDLARYRGCKLRFYRSPVCDFIVHYNLIPPLKMNQFTSPNTHPGLL
MLSKHKIIIPSFQTRPGGRRFVKIRLNPPKLFEDKWYTQQDLCKVPLVSI
TATAADLRYPFCSPQTNNPCTTFQVLRKNYNTVIGTSVKDQESTQDFENW
LYKTDSHYQTFATEAQLGRIPAFNPDGTKNTKQQSWQDNWSKKNSPWTGN
SGTYPQTTSEMYKIPYDSNFGFPTYRAQKDYILERRQCNFNYEVNNPVSK
KVWPQPSTTTPTVDYYEYHCGWFSNIFIGPNRYNLQFQTAYVDTTYNPLM
DKGKGNKIWFQYLSKKGTDYNEKQCYCTLEDMPLWAICFGYTDYVETQLG
PNVDHETAGLIIMICPYTQPPMYDKNRPNWGYVVYDTNFGNGKMPSGSGQ
VPVYWQCRWRPMLWFQQQVLNDISKTGPYAYRDEYKNVQLTLYYNFIFNW
GGDMYYPQVVKNPCGDSGIVPGSGRFTREVQVVSPLSMGPAYIFHYFDSR
RGFFSEKALKRMQQQQEFDESFTFKPKRPKLSTAAAEILQLEEDSTSGEG
KSPLQQEEKEVEVLQTPTVQLQLQRNIQEQLAIKQQLQFLLLQLLKTQSN
LHLNPQFLSPS (SEQ ID NO: 197)
```

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 75
Jelly-roll domain	75 - 284
Hypervariable Region	285 - 432
N22	433 – 599
C-terminal Domain	600 - 780

Table 29. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 5)

TTV-16(TUS01)-ORF1 (*Alphatorquevirus* Clade 5)	
Arg-Rich Region	MAYWFRRWGWRPRRRWRRWRRRRRRLPRRRTRRAVRGLGRRRKPR VRRRRRTRRRTYRRGWRRRRYIRRGRRKKK (SEQ ID NO: 198)
Jelly-roll Domain	LILTQWNPAIVKRCNIKGGLPIIICGEPRAAFNYGYHMEDYTPQPFPFGG GMSTVTFSLKALYDQYLKHQNRWTFSNDQLDLARYRGCKLRFYRSPV CDFIVHYNLIPPLKMNQFTSPNTHPGLLMLSKHKIIIPSFQTRPGGRRFV KIRLNPPKLFEDKWYTQQDLCKVPLVSITATAADLRYPFCSPQTNNPCT TFQVLRKNYNTVI (SEQ ID NO: 199)
Hypervariable domain	GTSVKDQESTQDFENWLYKTDSHYQTFATEAQLGRIPAFNPDGTKNTK QQSWQDNWSKKNSPWTGNSGTYPQTTSEMYKIPYDSNFGFPTYRAQK DYILERRQCNFNYEVNNPVSKKVWPQPSTTTPTVDYYEYHCGWFSNIFI GPNR (SEQ ID NO: 200)
N22	YNLQFQTAYVDTTYNPLMDKGKGNKIWFQYLSKKGTDYNEKQCYCT LEDMPLWAICFGYTDYVETQLGPNVDHETAGLIIMICPYTQPPMYDKN RPNWGYVVYDTNFGNGKMPSGSGQVPVYWQCRWRPMLWFQQQVLN DISKTGPYAYRDEYKNVQLTLYYNFIFN (SEQ ID NO: 201)
C-terminal domain	WGGDMYYPQVVKNPCGDSGIVPGSGRFTREVQVVSPLSMGPAYIFHY FDSRRGFFSEKALKRMQQQQEFDESFTFKPKRPKLSTAAAEILQLEEDS TSGEGKSPLQQEEKEVEVLQTPTVQLQLQRNIQEQLAIKQQLQFLLLQL LKTQSNLHLNPQFLSPS (SEQ ID NO: 202)

Table 30. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 6)

Name	TTV-TJN02
Genus/Clade	*Alphatorquevirus*, Clade 6
Accession Number	AB028669.1
Protein Accession Number	BAA94878.1

Full Sequence: 746 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MAWGWWRWRRRWPARRWRRRRRRRPVRRTRARRPARRYRRRRTVRTRRRR
WGRRRYRRGWRRRTYVRKGRHRKKKKRLILRQWQPATRRRCTITGYLPIV
FCGHTRGNKNYALHSDDYTPQGQPFGGALSTTSFSLKVLFDQHQRGLNKW
SFPNDQLDLARYRGCKFIFYRTKQTDWVGQYDISEPYKLDKYSCPNYHPG
```

NMIKAKHKFLIPSYDTNPRGRQKIIVKIPPPDLFVDKWYTQEDLCSVNLV
SLAVSAASFLHPFGSPQTDNPCYTFQVLKEFYYQAIGFSASTQAMTSVLD
TLYTQNSYWESNLTQFYVLNAKKGSDTTQPLTSNMPTREEFMAKKNTNYN
WYTYKAASVKNKLHQMRQTYFEELTSKGPQTTKSEEGYSQHWTTPSTNAY
EYHLGMFSAIFLAPDRPVPRFPCAYQDVTYNPLMDKGVGNHIWFQYNTKA
DTQLIVTGGSCKAHIQDIPLWAAFYGYSDFIESELGPFVDAETVGLVCVI
CPYTKPPMYNKTNPAMGYVFYDRNFGDGKWTDGRGKIEPYWQVRWRPEML
FQETVMADLVQTGPFSYKDELKNSTLVCKYKFYFTWGGNMMFQQTIKNPC
KTDGQPTDSSRHPRGIQVADPEQMGPRWVFHSFDWRRGYLSEKALKRLQE
KPLDYDEYFTQPKRPRIFPPTESAEGEFREPEKGSYSEEERSQASAEEQT
QEATVLLLKRRLREQQQLQQQLQFLTREMFKTQAGLHLNPMLLNQR
(SEQ ID NO: 203)

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 77
Jelly-roll domain	78 - 286
Hypervariable Region	287 - 416
N22	417 – 585
C-terminal Domain	586 - 746

Table 31. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 6)

TTV-TJN02-ORF1 (*Alphatorquevirus* Clade 6)	
Arg-Rich Region	MAWGWWRWRRRWPARRWRRRRRRRPVRRTRARRPARRYRRRRTVR TRRRRWGRRRYRRGWRRRTYVRKGRHRKKKKR (SEQ ID NO: 204)
Jelly-roll Domain	LILRQWQPATRRRCTITGYLPIVFCGHTRGNKNYALHSDDYTPQGQPFG GALSTTSFSLKVLFDQHQRGLNKWSFPNDQLDLARYRGCKFIFYRTKQ TDWVGQYDISEPYKLDKYSCPNYHPGNMIKAKHKFLIPSYDTNPRGRQ KIIVKIPPPDLFVDKWYTQEDLCSVNLVSLAVSAASFLHPFGSPQTDNPC YTFQVLKEFYYQAI (SEQ ID NO: 205)
Hypervariable domain	GFSASTQAMTSVLDTLYTQNSYWESNLTQFYVLNAKKGSDTTQPLTSN MPTREEFMAKKNTNYNWYTYKAASVKNKLHQMRQTYFEELTSKGPQ TTKSEEGYSQHWTTPSTNAYEYHLGMFSAIFLAPDR (SEQ ID NO: 206)
N22	PVPRFPCAYQDVTYNPLMDKGVGNHIWFQYNTKADTQLIVTGGSCKA HIQDIPLWAAFYGYSDFIESELGPFVDAETVGLVCVICPYTKPPMYNKT NPAMGYVFYDRNFGDGKWTDGRGKIEPYWQVRWRPEMLFQETVMA DLVQTGPFSYKDELKNSTLVCKYKFYFT (SEQ ID NO: 207)

C-terminal domain	WGGNMMFQQTIKNPCKTDGQPTDSSRHPRGIQVADPEQMGPRWVFHS FDWRRGYLSEKALKRLQEKPLDYDEYFTQPKRPRIFPPTESAEGEFREP EKGSYSEEERSQASAEEQTQEATVLLLKRRLREQQQLQQQLQFLTREM FKTQAGLHLNPMLLNQR (SEQ ID NO: 208)

Table 32. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 7)

Name	TTV-HD16d
Genus/Clade	*Alphatorquevirus*, Clade 7
Accession Number	FR751479.1
Protein Accession Number	NA

Full Sequence: 765 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MAWSWWWQRWRRRRWKPRRRRWRRLRWRRPRRAVRRRRRGRRVRRRRWAR
RRGRRRRYATRRKRRYRGRRFKKKLVLTQWHPNTMRRCLIKGIVPLVICG
HTRWNYNYALHSKDYTEEGRYPHGGALSTTTWSLKVLYDEHLKHHDFWGY
PNNQLDLARYKGAKFTFYRHKKTDFIIFFNRKPPFKLNKYSCASYHPGML
MQQRHKILLPSYETKPKGRPKITVRIKPPTLLEDKWYTQQDLCDVNLLQL
VVTAADFRHPLCSPQTNTPTTTFQVLKDIYYDTMSISEPTDSYTSVNNKS
TTQTFTNYSNTLENILYTRASYWNSFHATEYLNPNIIYKNGEKLFKEHED
LITWMTQTNNTGFLTKNNTAFGNNSYRPNADKIKKARKTYWNALIGTNDL
ATNIGQARAERFEYHLGWYSPIFLSRHRSNMNFARAYQDVTYNPNCDRGV
NNRVWVQPLTKPTTEFDEKRCKCVVQHLPLWAALYCYQDFVEEELGSSSE
ILNSCLLVLQCPYTFPPMYDKKLPDKGFVFYDSLFGDGKMSDGRGQVDIF
WQQRWYPRLATQMQVMHDITMTGPFSYRDELVSTQLTAKYTFDFMWGGNM
ISTQIIKNPCKDSGLEPAYPGRQRRDLQIVDPYSMGPQFSFHNWDYRHGL
FGQDAIDRVSKQPKDDADYPNPYKRPRYFPPTDQAAQEQEKDFSFLKTAP
SNSEESDQEVLQETQVLRFQPEQHKQLHLQLAERQRIGEQLRYLLQQMFK
TQANLHLNPYTFTQL (SEQ ID NO: 209)
```

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 74
Jelly-roll domain	75 - 286
Hypervariable Region	287 - 428
N22	429 – 595
C-terminal Domain	596 - 765

Table 33. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 7)

TTV- HD16d -ORF1 (*Alphatorquevirus* Clade 7)	
Arg-Rich Region	MAWSWWWQRWRRRRWKPRRRRWRRLRWRRPRRAVRRRRRGRRVR RRRWARRRGRRRRYATRRKRRYRGRRFKKK (SEQ ID NO: 210)
Jelly-roll Domain	LVLTQWHPNTMRRCLIKGIVPLVICGHTRWNYNYALHSKDYTEEGRYP HGGALSTTTWSLKVLYDEHLKHHDFWGYPNNQLDLARYKGAKFTFY RHKKTDFIIFFNRKPPFKLNKYSCASYHPGMLMQQRHKILLPSYETKPK GRPKITVRIKPPTLLEDKWYTQQDLCDVNLLQLVVTAADFRHPLCSPQ TNTPTTTFQVLKDIYYDTMSI (SEQ ID NO: 211)
Hypervariable domain	SEPTDSYTSVNNKSTTQTFTNYSNTLENILYTRASYWNSFHATEYLNPN IIYKNGEKLFKEHEDLITWMTQTNNTGFLTKNNTAFGNNSYRPNADKI KKARKTYWNALIGTNDLATNIGQARAERFEYHLGWYSPIFLSRHR (SEQ ID NO: 212)
N22	SNMNFARAYQDVTYNPNCDRGVNNRVWVQPLTKPTTEFDEKRCKCV VQHLPLWAALYCYQDFVEEELGSSSEILNSCLLVLQCPYTFPPMYDKK LPDKGFVFYDSLFGDGKMSDGRGQVDIFWQQRWYPRLATQMQVMHD ITMTGPFSYRDELVSTQLTAKYTFDFM (SEQ ID NO: 213)
C-terminal domain	WGGNMISTQIIKNPCKDSGLEPAYPGRQRRDLQIVDPYSMGPQFSFHN WDYRHGLFGQDAIDRVSKQPKDDADYPNPYKRPRYFPPTDQAAQEQE KDFSFLKTAPSNSEESDQEVLQETQVLRFQPEQHKQLHLQLAERQRIGE QLRYLLQQMFKTQANLHLNPYTFTQL (SEQ ID NO: 214)

Table 34. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Betatorquevirus*)

Name	TTMV-LY2
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 35. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Betatorquevirus*)

TTV- HD16d -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 36. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Gammatorquevirus*)

Name	TTMDV-MD1-073
Genus/Clade	*Gammatorquevirus*
Accession Number	AB290918.1
Protein Accession Number	BAG49427.1

Full Sequence: 673 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MPFWWGRRNKFWYGRNYRRKKRRFPKRRKRRFYRRTKYRRPARRRRRRRR
KVRRKKKTLIVRQWQPDSIVLCKIKGYDSIIWGAEGTQFQCSTHEMYEYT
RQKYPGGGGFGVQLYSLEYLYDQWKLRNNIWTKTNQLKDLCRYLKCVMTF
YRHQHIDFVIVYERQPPFEIDKLTYMKYHPYMLLQRKHKIILPSQTTNPR
GKLKKKKTIKPPKQMLSKWFFQQQFAKYDLLLIAAAACSLRYPRIGCCNE
NRMITLYCLNTKFYQDTEWGTTKQAPHYFKPYATINKSMIFVSNYGGKKT
EYNIGQWIETDIPGEGNLARYYRSISKEGGYFSPKILQAYQTKVKSVDYK
PLPIVLGRYNPAIDDGKGNKIYLQTIMNGHWGLPQKTPDYIIEEVPLWLG
FWGYYNYLKQTRTEAIFPLHMFVVQSKYIQTQQTETPNNFWAFIDNSFIQ
GKNPWDSVITYSEQKLWFPTVAWQLKTINAICESGPYVPKLDNQTYSTWE
LATHYSFHFKWGGPQISDQPVEDPGNKNKYDVPDTIKEALQIVNPAKNIA
ATMFHDWDYRRGCITSTAIKRMQQNLPTDSSLESDSDSEPAPKKKRLLPV
LHDPQKKTEKINQCLLSLCEESTCQEQETEENILKLIQQQQQQQQKLKHN
LLVLIKDLKVKQRLLQLQTGVLE(SEQ ID NO: 221)
```

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 57
Jelly-roll domain	58 – 259
Hypervariable Region	260 - 351
N22	352 – 510
C-terminal Domain	511 – 673

Table 37. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Gammatorquevirus*)

TTV- HD16d -ORF1 (*Gammatorquevirus*)

Arg-Rich Region	MPFWWGRRNKFWYGRNYRRKKRRFPKRRKRRFYRRTKYRRPARRRR RRRRKVRRKKK (SEQ ID NO: 222)
Jelly-roll Domain	TLIVRQWQPDSIVLCKIKGYDSIIWGAEGTQFQCSTHEMYEYTRQKYPG GGGFGVQLYSLEYLYDQWKLRNNIWTKTNQLKDLCRYLKCVMTFYR HQHIDFVIVYERQPPFEIDKLTYMKYHPYMLLQRKHKIILPSQTTNPRG KLKKKKTIKPPKQMLSKWFFQQQFAKYDLLLIAAAACSLRYPRIGCCN ENRMITLYCL (SEQ ID NO: 223)
Hypervariable domain	NTKFYQDTEWGTTKQAPHYFKPYATINKSMIFVSNYGGKKTEYNIGQ WIETDIPGEGNLARYYRSISKEGGYFSPKILQAYQTKVKSVDYKP (SEQ ID NO: 224)
N22	LPIVLGRYNPAIDDGKGNKIYLQTIMNGHWGLPQKTPDYIIEEVPLWLG FWGYYNYLKQTRTEAIFPLHMFVVQSKYIQTQQTETPNNFWAFIDNSFI QGKNPWDSVITYSEQKLWFPTVAWQLKTINAICESGPYVPKLDNQTYS TWELATHYSFHFK (SEQ ID NO: 225)
C-terminal domain	WGGPQISDQPVEDPGNKNKYDVPDTIKEALQIVNPAKNIAATMFHDW DYRRGCITSTAIKRMQQNLPTDSSLESDSDSEPAPKKKRLLPVLHDPQK KTEKINQCLLSLCEESTCQEQETEENILKLIQQQQQQQQKLKHNLLVLI KDLKVKQRLLQLQTGVLE (SEQ ID NO: 226)

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 D3. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Gammatorquevirus*)

Name	Ring 4.0
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 D4. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Gammatorquevirus*)

Ring4.0 (*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)

Table D7. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*) – Clade 3

Name	Ring6.0
Genus/Clade	*Alphatorquevirus* Clade 3
Accession Number	
Protein Accession Number	

Full Sequence: 767 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MAYGWWRRRRRRPWWRRRWRRWRRRRRPRRRRPRRRYRRRRTVRRRGRGR
WTRAHRRWRRKGKRSRKKKIIIRQWQPNYTRRCNIVGYMPLLICGENTVA
TNYATHSDDSYYPGPFGGGMTTDKFTLRILYDEYKRFMNYWTSSNEDLDL
CRYLGCTLYVFRHPEVDFIIIINTSPPFLDTEITGPSIHPGMMALNKRSR
WIPSIKNRPGRKHYIKIKVGAPRMFTDKWYPQTDLCDMTLLTIFASAADM
QYPFGSPLTDTIVVSFQVLQSMYNDCLSVLPDNFAETSGKGTQLHENIIQ
HLPYYNTTQTQAQFKRFIENMNATNGDNIWASYINTTKFSSANTPKNDTG
IGGPYTTYSDSWYKGTVYNDKIKTIPIKASKLYYEQTKNLIGITFTGSTH
RLHYCGGLYSSVWLSAGRSYFETKGPYTDITYNPFSDRGEGNMLWIDWLT
KNDSVYSKTSSKCLIENLPLWASVYGYKEYCSKVTGDTNIEHNCRCVIRS
PYTVPQLLDHNNPFRGYVPYSFNFGNGKMPGGSSLVPIRMRAKWYPTLFH
QKEVLEAIAQAGPFAYHSDIKKVSLGIKYRFKWVWGGNPVSQQVVRNPCK
TTQGSSGNRVPRSIQVVDPRYNTPELTIHAWDFRHGFFGRKAIKRMQEQP
IPHDTFSAGFKRSRRDTEALQCSQEEQQKENLLFPVQQLKRVPPWETSQE
SQSEEENSQKQETLSQQLRDQLHKQRLMGEQLRSLLYQMQRVQQNQHINP
        MLLPKGLALTSISHNVI (SEQ ID NO: 937)
```

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 69
Jelly-roll domain	70 - 269
Hypervariable Region	270 - 424
N22	425 – 584
C-terminal Domain	585 – 767

Table D8. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*) – Clade 3

Ring6.0 (*Alphatorquevirus*)	
Arg-Rich Region	MAYGWWRRRRRRPWWRRRWRRWRRRRRPRRRRPRRRYRRRRTVRR RGRGRWTRAHRRWRRKGKRSRKKK (SEQ ID NO: 938)
Jelly-roll Domain	IIIRQWQPNYTRRCNIVGYMPLLICGENTVATNYATHSDDSYYPGPFGG GMTTDKFTLRILYDEYKRFMNYWTSSNEDLDLCRYLGCTLYVFRHPE VDFIIIIINTSPPFLDTEITGPSIHPGMMALNKRSRWIPSIKNRPGRKHYIKI KVGAPRMFTDKWYPQTDLCDMTLLTIFASAADMQYPFGSPLTDTIVVS FQVL (SEQ ID NO: 939)
Hypervariable domain	QSMYNDCLSVLPDNFAETSGKGTQLHENIIQHLPYYNTTQTQAQFKRFI ENMNATNGDNIWASYINTTKFSSANTPKNDTGIGGPYTTYSDSWYKGT VYNDKIKTIPIKASKLYYEQTKNLIGITFTGSTHRLHYCGGLYSSVWLSA GRSYFETK (SEQ ID NO: 940)
N22	GPYTDITYNPFSDRGEGNMLWIDWLTKNDSVYSKTSSKCLIENLPLWA SVYGYKEYCSKVTGDTNIEHNCRCVIRSPYTVPQLLDHNNPFRGYVPY SFNFGNGKMPGGSSLVPIRMRAKWYPTLFHQKEVLEAIAQAGPFAYHS DIKKVSLGIKYRFKWV (SEQ ID NO: 941)
C-terminal domain	WGGNPVSQQVVRNPCKTTQGSSGNRVPRSIQVVDPRYNTPELTIHAWD FRHGFFGRKAIKRMQEQPIPHDTFSAGFKRSRRDTEALQCSQEEQQKEN LLFPVQQLKRVPPWETSQESQSEEENSQKQETLSQQLRDQLHKQRLMG EQLRSLLYQMQRVQQNQHINPMLLPKGLALTSISHNVI (SEQ ID NO: 942)

Table D9. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*) – Clade 7

Name	Ring 7.0
Genus/Clade	*Alphatorquevirus* – Clade 7
Accession Number	
Protein Accession Number	

Full Sequence: 766 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MAWRWWWQRRWRRRRWPRRRWRRLRRRRPRRPVRRRRRRTTVRRRRWRGR
RGRRTYTRRAVRRRRRPRKRLVLTQWSPQTVRNCSIRGIVPMVICGHTKA
GRNYAIHSEDFTTQIQPFGGSFSTTTWSLKVLWDEHQKFQNRWSYPNTQL
DLARYRGVTFWFYRDQKTDYIVQWSRNPPFKLNKYSSAMYHPGMMMQAKR
KLVVPSFQTRPKGKKRYRVTIKPPNMFADKWYTQEDLCPVPLVQIVVSAA
SLLHPFCPPQTNNPCITFQVLKDIYDECIGVNETMKDKYKKLQTTLYTTC
TYYQTTQVLAQLSPAFQPAMKPTTTQSAATATTLGNYVPELKYNNGSFHT
GQNAVFGMCSYKPTDSIMTKANGWFWQNLMVDNNLHSSYGKATLECMEYH
TGIYSSIFLSPQRSLEFPAAYQDVTYNPNCDRAVGNVVWFQYSTKMDTNF
DETKCKCVLKNIPLWAAFNGYSDFIMQELSISTEIHNFGIVCFQCPYTFP
PCFNKNKPLKGYVFYDTTFGNGKMPDGSGHVPIYWQQRWWIRLAFQVQVM
HDFVLTGPFSYKDDLANTTLTARYKFKFKWGGNIIPEQIIKNPCHREQSL
ASYPDRQRRDLQVVDPSTMGPIYTFHTWDWRRGLFGADAIQRVSQKPGDA
LRFTNPFKRPRYLPPTDREDYRQEEDFALQEKRRRTSTEEAQDEESPPES
APLLQQQQQQRQLSVHLAEQQRLGVQLRYILQEVLKTQAGLHLNPLLLGP
         PQTRSISLSPPKAYSP (SEQ ID NO: 943)
```

Annotations:

Putative Domain	*AA range*
Arg-Rich Region	1 – 70
Jelly-roll domain	71 - 271
Hypervariable Region	272 - 418
N22	419 – 579
C-terminal Domain	580 – 766

Table D10. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*) – Clade 7

Ring7.0 (*Alphatorquevirus*)	
Arg-Rich Region	MAWRWWWQRRWRRRRWPRRRWRRLRRRRPRRPVRRRRRRTTVRRRRWRGRRGRRTYTRRAVRRRRRPRKR (SEQ ID NO: 944)
Jelly-roll Domain	LVLTQWSPQTVRNCSIRGIVPMVICGHTKAGRNYAIHSEDFTTQIQPFGGSFSTTTWSLKVLWDEHQKFQNRWSYPNTQLDLARYRGVTFWFYRDQKTDYIVQWSRNPPFKLNKYSSAMYHPGMMMQAKRKLVVPSFQTRPKGKKRYRVTIKPPNMFADKWYTQEDLCPVPLVQIVVSAASLLHPFCPPQTNNPCITFQVL (SEQ ID NO: 945)
Hypervariable domain	KDIYDECIGVNETMKDKYKKLQTTLYTTCTYYQTTQVLAQLSPAFQPAMKPTTTQSAATATTLGNYVPELKYNNGSFHTGQNAVFGMCSYKPTDSIMTKANGWFWQNLMVDNNLHSSYGKATLECMEYHTGIYSSIFLSPQRS

	LEFP (SEQ ID NO: 946)
N22	AAYQDVTYNPNCDRAVGNVVWFQYSTKMDTNFDETKCKCVLKNIPL WAAFNGYSDFIMQELSISTEIHNFGIVCFQCPYTFPPCFNKNKPLKGYVF YDTTFGNGKMPDGSGHVPIYWQQRWWIRLAFQVQVMHDFVLTGPFS YKDDLANTTLTARYKFKFK (SEQ ID NO: 947)
C-terminal domain	WGGNIIPEQIIKNPCHREQSLASYPDRQRRDLQVVDPSTMGPIYTFHTW DWRRGLFGADAIQRVSQKPGDALRFTNPFKRPRYLPPTDREDYRQEED FALQEKRRRTSTEEAQDEESPPESAPLLQQQQQQRQLSVHLAEQQRLG VQLRYILQEVLKTQAGLHLNPLLLGPPQTRSISLSPPKAYSP (SEQ ID NO: 948)

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})TTSRNYA THSDDTIQKGPFGGGMSTTTFSLRVLYDEYQRFMNRW TYSNEDLDLARYLGCKFTFYRHPDXDFIVQYNTNPPFK DTKLTAPSIHP(X_{1-5})GMLMLSKRKILIPSLKTRPKGKHY VKVRIGPPKLFEDKWYTQSDLCDVPLVXLYATAADLQ	227

	HPFGSPQTDNPCVTFQVLGSXYNKHLSISP; wherein X = any amino acid.	
N22	SNFEFPGAYTDITYNPLTDKGVGNMVWIQYLTKPDTIX DKTQS(X_{0-3})KCLIEDLPLWAALYGYVDFCEKETGDSAII XNXGRVLIRCPYTKPPLYDKT(X_{0-4})NKGFVPYSTNFGN GKMPGGSGYVPIYWRARWYPTLFHQKEVLEDIVQSGP FAYKDEKPSTQLVMKYCFNFN; wherein X = any amino acid.	228
CTD	WGGNPISQQVVRNPCKDSG(X_{0-3})SGXGRQPRSVQVVD PKYMGPEYTFHSWDWRRGLFGEKAIKRMSEQPTDDEI FTGGXPKRPRRDPPTXQXPEE(X_{1-4})QKESSSFR(X_{2-14})PW ESSSQEXESESQEEEE(X_{0-30})EQTVQQQLRQQLREQRRL RVQLQLLFQQLLKT(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	LKQWQPSTIRKCKIKGYLPLFQCGKGRISNNYTQYKESI VPHHEPGGGGWSIQQFTLGALYEEHLKLRNWWTKSN DGLPLVRYLGCTIKLYRSEDTDYIVTYQRCYPMTATKL TYLSTQPSRMLMNKHKIIVPSKXT(X_{1-4})NKKKKPYKKIF IKPPSQMQNKWYFQQDIANTPLLQLTXTACSLDRMYL SSDSISNNITFTSLNTNFFQNPNFQ; wherein X = any amino acid.	230
N22	(X_{4-10})TPLYFECRYNPFKDKGTGNKVYLVSNN(X_{1-8})TG WDPPTDPDLIIEGFPLWLLLWGWLDWQKKLGKIQNID TDYILVIQSXYYIPP(X_{1-3})KLPYYVPLDXD(X_{0-2})FLHGRS	231

	PY(X_{3-16})PSDKQHWHPKVRFQXETINNIALTGPGTPKLP NQKSIQAHMKYKFYFK; wherein X = any amino acid.	
CTD	WGGCPAPMETITDPCKQPKYPIPNNLLQTTSLQXPTTPI ETYLYKFDERRGLLTKKAAKRIKKDXTTETTLFTDTGX XTSTTLPTXXQTETTQEEXTSEEE(X_{0-5})ETLLQQLQQLR RKQKQLRXRILQLLQLLXLL(X_{0-26})*; wherein X = any amino acid.	232

Table 37C. Gammatorquevius ORF1 domain consensus sequences

Domain	Sequence	SEQ ID NO:
Jelly-Roll	TIPLKQWQPESIRKCKIKGYGTLVLGAEGRQFYCYTNE KDEYTPPKAPGGGGFGVELFSLEYLYEQWKARNNIWT KSNXYKDLCRYTGCKITFYRHPTTDFIVXYSRQPPFEID KXTYMXXHPQXLLLRKHKKIILSKATNPKGKLKKKIKI KPPKQMLNKWFFQKQFAXYGLVQLQAAACBLRYPRL GCCNENRLITLYYLN; wherein X = any amino acid.	233
N22	LPIVVARYNPAXDTGKGNKXWLXSTLNGSXWAPPTTD KDLIIEGLPLWLALYGYWSYJKKVKKDKGILQSHMFV VKSPAIQPLXTATTQXTFYPXIDNSFIQGKXPYDEPJTX NQKKLWYPTLEHQQETINAIVESGPYVPKLDNQKNST WELXYXYTFYFK; wherein X = any amino acid.	234
CTD	WGGPQIPDQPVEDPKXQGTYPVPDTXQQTIQIXNPLKQ KPETMFHDWDYRRGIITSTALKRMQENLETDSSFXSDS EETP(X_{0-2})KKKKRLTXELPXPQEETEEIQSCLLSLCEEST	235

	CQEE(X_{1-6})ENLQQLIHQQQQQQQLKHNILKLLSDLKZ KQRLLQLQTGILE(X_{1-10})*; wherein X = any amino acid.	

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.

ORF2 molecules

In some embodiments, the anellosome 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 Element

In some embodiments, the anellosome comprises a genetic element. In some embodiments, the genetic element has one or more of the following characteristics: is substantially non-integrating with a host cell's genome, is an episomal nucleic acid, is a single stranded 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 at least about 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* nucleic acid sequence, e.g., as described herein (e.g., as described in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17), or a fragment thereof, or encodes an amino acid sequence having at least about 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* amino acid sequence (e.g., as described in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18), or a fragment thereof. In embodiments, the genetic element comprises a sequence encoding an effector (e.g., an endogenous effector or an exogenous effector, e.g., a payload), e.g., a polypeptide effector (e.g., a protein) or nucleic acid effector (e.g., a non-coding RNA, e.g., a miRNA, siRNA, mRNA, lncRNA, RNA, DNA, an antisense RNA, gRNA).

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 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 some embodiments, the substantially non-pathogenic protein comprises an amino acid sequence or a functional fragment thereof or a sequence having at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of the amino acid sequences described herein, an *Anellovirus* amino acid sequence, 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.

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. 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 shown in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17).

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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table A1 (e.g., nucleotides 165 – 235 of the nucleic acid sequence of Table A1). 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 the *Anellovirus* GC-rich nucleotide sequence of Table A1 (e.g., nucleotides 3620 – 3648 of the nucleic acid sequence of Table A1). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table A3 (e.g., nucleotides 175 – 245 of the nucleic acid sequence of Table A3). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table A5 (e.g., nucleotides 138 – 208 of the nucleic acid sequence of Table A5). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table A7 (e.g., nucleotides 174 – 244 of the nucleic acid sequence of Table A7). 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 the *Anellovirus* GC-rich nucleotide sequence of Table A7 (e.g., nucleotides 3720 – 3742 of the nucleic acid sequence of Table A7). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table A9 (e.g., nucleotides 100 – 171 of the nucleic acid sequence of Table A9).

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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table A11 (e.g., nucleotides 294 – 364 of the nucleic acid sequence of Table A11). 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 the *Anellovirus* GC-rich nucleotide sequence of Table A1 (e.g., nucleotides 3844 – 3895 of the nucleic acid sequence of Table A11).

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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table 1 (e.g., nucleotides 177 – 247 of the nucleic acid sequence of Table 1). 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 the *Anellovirus* GC-rich nucleotide sequence of Table 1 (e.g., nucleotides 3415 – 3570 of the nucleic acid sequence of Table 1). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table 3 (e.g., nucleotides 204 – 273 of the nucleic acid sequence of Table 3). 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 the *Anellovirus* GC-rich nucleotide sequence of Table 3 (e.g., nucleotides 3302 – 3541 of the nucleic acid sequence of Table 3). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table 5 (e.g., nucleotides 170 – 240 of the nucleic acid sequence of Table 5). 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 the *Anellovirus* GC-rich nucleotide sequence of Table 5 (e.g., nucleotides 3632 – 3753 of the nucleic acid sequence of Table 5). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table 7 (e.g., nucleotides 170 – 238 of the nucleic acid sequence of Table 7). 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 the *Anellovirus* GC-rich nucleotide sequence of Table 7 (e.g., nucleotides 3768 – 3878 of the nucleic acid sequence of Table 7). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table 9 (e.g., nucleotides 170 – 240 of the nucleic acid sequence of Table 9). 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 the *Anellovirus* GC-rich nucleotide sequence of Table 9 (e.g., nucleotides 3302 – 3541 of the

nucleic acid sequence of Table 9). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table 11 (e.g., nucleotides 174 – 244 of the nucleic acid sequence of Table 11). 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 the *Anellovirus* GC-rich nucleotide sequence of Table 11 (e.g., nucleotides 3691 – 3794 of the nucleic acid sequence of Table 11). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table 13 (e.g., nucleotides 170 – 240 of the nucleic acid sequence of Table 13). 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 the *Anellovirus* GC-rich nucleotide sequence of Table 13 (e.g., nucleotides 3759 – 3866 of the nucleic acid sequence of Table 13). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table 15 (e.g., nucleotides 323 – 393 of the nucleic acid sequence of Table 15). 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 the *Anellovirus* GC-rich nucleotide sequence of Table 15 (e.g., nucleotides 2868 – 2929 of the nucleic acid sequence of Table 15). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table 17 (e.g., nucleotides 117 – 187 of the nucleic acid sequence of Table 17). 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 the *Anellovirus* GC-rich nucleotide sequence of Table 17 (e.g., nucleotides 3054 – 3172 of the nucleic acid sequence of Table 17).

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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table B1 (e.g., nucleotides 185-255 of the nucleic acid sequence of Table B1). 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 the *Anellovirus* GC-rich nucleotide sequence of Table B1 (e.g., nucleotides 3141-3264 of the nucleic acid sequence of Table B1). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table B2 (e.g., nucleotides 185-254 of the nucleic acid sequence of Table B2). 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 the *Anellovirus* GC-rich nucleotide sequence of Table B2 (e.g., nucleotides 3076-3176 of the

nucleic acid sequence of Table B2). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table B3 (e.g., nucleotides 178-248 of the nucleic acid sequence of Table B3). 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 the *Anellovirus* GC-rich nucleotide sequence of Table B3 (e.g., nucleotides 3555-3696 of the nucleic acid sequence of Table B3). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table B4 (e.g., nucleotides 176-246 of the nucleic acid sequence of Table B4). 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 the *Anellovirus* GC-rich nucleotide sequence of Table B4 (e.g., nucleotides 3720-3828 of the nucleic acid sequence of Table B4). 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 the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table B5 (e.g., nucleotides 170-240 of the nucleic acid sequence of Table B5). 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 the *Anellovirus* GC-rich nucleotide sequence of Table B5 (e.g., nucleotides 3716-3815 of the nucleic acid sequence of Table B5).

5' UTR 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 38 and/or Figure 20. 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.

In embodiments, the genetic element 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 (e.g., nucleotides 165 – 235 of the nucleic acid sequence of Table A1). In embodiments, the genetic element 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 A3 (e.g., nucleotides 175 – 245 of the nucleic acid sequence of Table A3). In embodiments, the genetic element 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 A5 (e.g., nucleotides 138 – 208 of the nucleic acid sequence of Table A5). In embodiments, the genetic element 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 A7 (e.g., nucleotides 174 – 244 of the nucleic acid sequence of Table A7). In embodiments, the genetic element 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 A9 (e.g., nucleotides 100 – 171 of the nucleic acid sequence of Table A9). In embodiments, the genetic element 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 A11 (e.g., nucleotides 294 – 364 of the nucleic acid sequence of Table A11).

In embodiments, the genetic element 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 1 (e.g., nucleotides 177 – 247 of the nucleic acid sequence of Table 1). In embodiments, the genetic element 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 3 (e.g., nucleotides 204 – 273 of the nucleic acid sequence of Table 3). In embodiments, the genetic element 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 5 (e.g., nucleotides 170 – 240 of the nucleic acid sequence of Table 5). In embodiments, the genetic element

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 7 (e.g., nucleotides 170 – 238 of the nucleic acid sequence of Table 7). In embodiments, the genetic element 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 9 (e.g., nucleotides 170 – 240 of the nucleic acid sequence of Table 9). In embodiments, the genetic element 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 11 (e.g., nucleotides 174 – 244 of the nucleic acid sequence of Table 11). In embodiments, the genetic element 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 13 (e.g., nucleotides 170 – 240 of the nucleic acid sequence of Table 13). In embodiments, the genetic element 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 15 (e.g., nucleotides 323 – 393 of the nucleic acid sequence of Table 15). In embodiments, the genetic element 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 17 (e.g., nucleotides 117 – 187 of the nucleic acid sequence of Table 17).

In embodiments, the genetic element 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 (e.g., nucleotides 185-255 of the nucleic acid sequence of Table B1). In embodiments, the genetic element 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 B2 (e.g., nucleotides 185-254 of the nucleic acid sequence of Table B2). In embodiments, the genetic element 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 (e.g., nucleotides 178-248 of the nucleic acid sequence of Table B3). In embodiments, the genetic element 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 B4 (e.g., nucleotides 176-246 of the nucleic acid sequence of Table B4). In embodiments, the genetic element 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 B5 (e.g., nucleotides 170-240 of the nucleic acid sequence of Table B5).

Table 38. Exemplary 5' UTR sequences from *Anelloviruses*

Source	Sequence	SEQ ID NO:
Consensus	CGGGTGCCGX$_1$AGGTGAGTTTACACACCGX$_2$AGT CAAGGGGCAATTCGGGCTCX$_3$GGACTGGCCGGG CX$_4$X$_5$TGGG 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	112

	X_1X_2TGGG; wherein X_1 comprises T or C, and wherein X_2 comprises A, C, or T.	
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

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 any of Table 39 and/or Figures 20 and 32. 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) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (SEQ ID NO: 168);

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

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

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

(x) GCGCTACGCGCGCGCGCCGGGGGGCTCTGCCCCCCC (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$_1$GGX$_2$GX$_3$X$_4$X$_5$CGCGCTX$_6$CGCGC GCX$_7$X$_8$X$_9$X$_{10}$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$_{25}$CCCCCCCX$_{26}$CGCGC ATX$_{27}$X$_{28}$GCX$_{29}$CGGGX$_{30}$CCCCCCCCCX$_{31}$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 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	120

		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-CGCCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCC-- GGGGGGGCTCCGCCCCCCCGGCCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCCGCCGCC	125

	Fragment 1	GCGGCGG	126
	Fragment 2	GGGGGCG	127
	Fragment 3	GCCGCG	128
	Fragment 4	TTCGCGCGCCGCCCACCAGGGGGTG	129
	Fragment 5	CTGCG	130
	Fragment 6	CGCCCCCCCCCGCGCAT	131
	Fragment 7	GCGCGGGGCCCCCCCCC	132
	Fragment 8	GGGGGGGCTCCGCCCCCCCGGCCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCCGCCGCC	133
TTV-HD23a	Full sequence	CGGCGGCGGCGGCG-CGCGCGCTGCGCGCGCG---CGCCGGGGGGGCGCCAGCG-CCCCCCCCCCCGCGCAT GCACGGGTCCCCCCCCCCACGGGGGGCTCC G CCCCCCGGCCCCCCCCC	134
	Fragment 1	CGGCGGCGGCGGCG	135
	Fragment 2	CGCGCGCTGCGCGCGCG	136
	Fragment 3	CGCCGGGGGGGCGCCAGCG	137
	Fragment 4	CCCCCCCCCCCGCGCAT	138
	Fragment 5	GCACGGGTCCCCCCCCCCACGGGGGGCTCC G	139
	Fragment 6	CCCCCCGGCCCCCCCCC	140
TTV-JA20	Full sequence	CCGTCGGCGGGGGGGCCGCGCGCTGCGCG CGCGGCCC-CCGGGGGAGGCACAGCCTCCCCCCCCCGCG CGCATGCGCGCGGGTCCCCCCCCCTCCGGG GGGCTCCGCCCCCCGGCCCCCCCC	141
	Fragment 1	CCGTCGGCGGGGGGGCCGCGCGCTGCGCG CGCGGCCC	142
	Fragment 2	CCGGGGGAGGCACAGCCTCCCCCCCCCGCG CGCATGCGCGCGGGTCCCCCCCCCTCCGGG GGGCTCCGCCCCCCGGCCCCCCCC	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-GCGGCGCGCTGCGCGCGCCGCCCAGTAGG GGGAGCCATGCG---CCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCC-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	CCCCCCGGCCCCCCCCG	159
	Fragment 7	CGCGCTGCGCGCGCCGCCCAGTAGGGGGA GCCATGC	160
	Fragment 8	CCGCCATCTTAAGTAGTTGAGGCGGACGGT GGCGTGAGTTCAAAGGTCACCATCAGCCAC ACCTACTCAAAATGGTGG	161
	Fragment 9	CTTAAGTAGTTGAGGCGGACGGTGGCGTGA GTTCAAAGGTCACCATCAGCCACACCTACT	162

		CAAAATGGTGGACAATTTCTTCCGGGTCAA AGGTTACAGCCGCCATGTTAAAACACGTGA CGTATGACGTCACGGCCGCCATTTTGTGAC ACAAGATGGCCGACTTCCTTCC	
Additional GC-rich Sequences (as shown in Figure 32)	36-nucleotide consensus GC-rich region sequence 1	CGCGCTGCGCGCGCCGCCCAGTAGGGGGA GCCATGC	163
	36-nucleotide region consensus sequence 2	GCGCTX$_1$CGCGCGCGCCGGGGGGCTGCG 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 CATGCGCGGGGCCCCCCCCCCGCGGGGGGCT CCGCCCCCCCGGTCCCCCCCCG	804
	TTV-16	CGGCCGTGCGGCGGCGCGCGCGCTTCGCGC GCGCGCCGGGGGCTGCCGCCCCCCCCCCGCG CATGCGCGCGGGGCCCCCCCCCCGCGGGGGG GCTCCGCCCCCCCGGCCCCCCCCCCCCG	805
	TTV-TJN02	CGGCGGCGGCGCGCGCGCTACGCGCGCGC GCCGGGGGGCTGCCGCCCCCCCCCCCGCGCA TGCGCGGGGCCCCCCCCCCGCGGGGGGCTCC	806

		GCCCCCCGGCCCCCC	
	TTV-HD16d	GGCGGCGGCGCGCGCTACGCGCGCGCG CCGGGGAGCTCTGCCCCCCCCCGCGCATGC GCGCGGGTCCCCCCCCCGCGGGGGGCTCCG CCCCCCGGTCCCCCCCCCG	807

In embodiments, the genetic element 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 (e.g., nucleotides 3620 – 3648 of the nucleic acid sequence of Table A1). In embodiments, the genetic element 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 A7 (e.g., nucleotides 3720 – 3742 of the nucleic acid sequence of Table A7). In embodiments, the genetic element 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 (e.g., nucleotides 3844 – 3895 of the nucleic acid sequence of Table A11).

In embodiments, the genetic element 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 1 (e.g., nucleotides 3415 – 3570 of the nucleic acid sequence of Table 1). In embodiments, the genetic element 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 3 (e.g., nucleotides 3302 – 3541 of the nucleic acid sequence of Table 3). In embodiments, the genetic element 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 5 (e.g., nucleotides 3632 – 3753 of the nucleic acid sequence of Table 5). In embodiments, the genetic element 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 7 (e.g., nucleotides 3768 – 3878 of the nucleic acid sequence of Table 7). In embodiments, the genetic element 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 9 (e.g., nucleotides 3302 – 3541 of the nucleic acid sequence of Table 9). In embodiments, the genetic element 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 11 (e.g., nucleotides 3691 – 3794 of the nucleic acid sequence of Table 11). In embodiments, the genetic element 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 13 (e.g., nucleotides 3759 – 3866 of the nucleic acid sequence of Table 13). In embodiments, the genetic element 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 15 (e.g., nucleotides 2868 – 2929 of the nucleic acid sequence of Table 15). In embodiments, the genetic element 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 17 (e.g., nucleotides 3054 – 3172 of the nucleic acid sequence of Table 17).

In embodiments, the genetic element 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 (e.g., nucleotides 3141-3264 of the nucleic acid sequence of Table B1). In embodiments, the genetic element 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 B2 (e.g., nucleotides 3076-3176 of the nucleic acid sequence of Table B2). In embodiments, the genetic element 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 (e.g., nucleotides 3555-3696 of the nucleic acid sequence of Table B3). In embodiments, the genetic element 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 B4 (e.g., nucleotides 3720-3828 of the nucleic acid sequence of Table B4). In embodiments, the genetic element 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 B5 (e.g., nucleotides 3716-3815 of the nucleic acid sequence of Table B5).

Effector

In some embodiments, the genetic element may include one or more sequences that encode 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 in Example 10, 12, or 22. 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 as shown in any of Tables A1-A12, B1-B5, C1-C5, or 1-18).

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 in Example 11.

Regulatory Nucleic Acid

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 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.

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 anellosome 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 peptides or polypeptides

In some embodiments, the genetic element comprises a sequence that encodes a therapeutic peptide or polypeptide, e.g., a protein replacement therapeutic. Such therapeutics include, but are not limited to, 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 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 150 amino acids, or any range therebetween.

Exemplary effectors are described herein, e.g., in the tables below.

Table A. Exemplary enzymatic effectors and corresponding indications

Effector	**Indication**
3-methylcrotonyl-CoA carboxylase	3-methylcrotonyl-CoA carboxylase deficiency
Acetyl-CoA-glucosaminide N-acetyltransferase	Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-C
ADAMTS13	Thrombotic Thrombocytopenic Purpura
adenine phosphoribosyltransferase	Adenine phosphoribosyltransferase deficiency
Adenosine deaminase	Adenosine deaminase deficiency
ADP-ribose protein hydrolase	Glutamyl ribose-5-phosphate storage disease
alpha glucosidase	Glycogen storage disease type 2 (Pompe's disease)
Ammonia monooxygenase	Argininemia
Ammonia monooxygenase	argininosuccinate aciduria
Ammonia monooxygenase	citrullinemia type I
Ammonia monooxygenase	Citrullinemia type II
Ammonia monooxygenase	hyperammonemia/ornithinemia/citrullinemia (ornithine transporter defect)
Arginase	Familial hyperarginemia
Arylsulfatase A	Metachromatic leukodystrophy
Cathepsin K	Pycnodysostosis
Ceramidase	Farber's disease (lipogranulomatosis)
Cystathionine B synthase	Homocystinuria
Dolichol-P-mannose synthase	Congenital disorders of N-glycosylation CDG Ie
Dolicho-P-Glc:Man9GlcNAc2-PP-dolichol glucosyltransferase	Congenital disorders of N-glycosylation CDG Ic

Dolicho-P-Man:Man5GlcNAc2-PP-dolichol mannosyltransferase	Congenital disorders of N-glycosylation CDG Id
Dolichyl-P-glucose:Glc-1-Man-9-GlcNAc-2-PP-dolichyl-α-3-glucosyltransferase	Congenital disorders of N-glycosylation CDG Ih
Dolichyl-P-mannose:Man-7-GlcNAc-2-PP-dolichyl-α-6-mannosyltransferase	Congenital disorders of N-glycosylation CDG Ig
Factor II	Factor II Deficiency
Factor IX	Hemophilia B
Factor V	Owren's disease
Factor VIII	Hemophilia A
Factor X	Stuart-Prower Factor Deficiency
Factor XI	Hemophilia C
Factor XIII	Fibrin Stabilizing factor deficiency
Galactosamine-6-sulfate sulfatase	Mucopolysaccharidosis MPS IV (Morquio's syndrome) Type IV-A
galactose dehydrogenase	Transferase deficient galactosemia (Galactosemia type 1)
Galactosylceramide β-galactosidase	Krabbe's disease
Ganglioside β-galactosidase	GM1 gangliosidosis, generalized
Ganglioside β-galactosidase	GM2 gangliosidosis
Ganglioside β-galactosidase	Sphingolipidosis Type I
Ganglioside β-galactosidase	Sphingolipidosis Type II (juvenile type)
Ganglioside β-galactosidase	Sphingolipidosis Type III (adult type)
Glucosidase I	Congenital disorders of N-glycosylation CDG IIb
Glucosylceramide β-glucosidase	Gaucher's disease
Heparan-S-sulfate sulfamidase	Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-A
homogentisate oxidase	Alkaptonuria
Hyaluronidase	Mucopolysaccharidosis MPS IX (hyaluronidase deficiency)
Iduronate sulfate sulfatase	Mucopolysaccharidosis MPS II (Hunter's syndrome)
Lecithin-cholesterol acyltransferase (LCAT)	Complete LCAT deficiency, Fish-eye disease, atherosclerosis, hypercholesterolemia
Leucine metabolizing enzyme	Isovaleric acidemia
Leucine metabolizing enzyme	Maple syrup urine disease
Lysine oxidase	Glutaric acidemia type I
Lysosomal acid lipase	Cholesteryl ester storage disease (CESD)
Lysosomal acid lipase	Lysosomal acid lipase deficiency
lysosomal acid lipase	Wolman's disease
Lysosomal pepstatin-insensitive peptidase	Ceroid lipofuscinosis Late infantile form (CLN2, Jansky-Bielschowsky disease)
Mannose (Man) phosphate (P) isomerase	Congenital disorders of N-glycosylation CDG Ib
Mannosyl-α-1,6-glycoprotein-β-1,2-N-acetylglucosminyltransferase	Congenital disorders of N-glycosylation CDG IIa
Metalloproteinase-2	Winchester syndrome
methylmalonyl-CoA mutase	Methylmalonic acidemia (vitamin b12 non-responsive)

N-Acetyl galactosamine α-4-sulfate sulfatase (arylsulfatase B)	Mucopolysaccharidosis MPS VI (Maroteaux-Lamy syndrome)
N-acetyl-D-glucosaminidase	Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-B
N-Acetyl-galactosaminidase	Schindler's disease Type I (infantile severe form)
N-Acetyl-galactosaminidase	Schindler's disease Type II (Kanzaki disease, adult-onset form)
N-Acetyl-galactosaminidase	Schindler's disease Type III (intermediate form)
N-acetyl-glucosaminine-6-sulfate sulfatase	Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-D
N-acetylglucosaminyl-1-phosphotransferase	Mucolipidosis ML III (pseudo-Hurler's polydystrophy)
N-Acetylglucosaminyl-1-phosphotransferase catalytic subunit	Mucolipidosis ML II (I-cell disease)
N-acetylglucosaminyl-1-phosphotransferase, substrate-recognition subunit	Mucolipidosis ML III (pseudo-Hurler's polydystrophy) Type III-C
N-Aspartylglucosaminidase	Aspartylglucosaminuria
Neuraminidase 1 (sialidase)	Sialidosis
Oxalate oxidase	Primary hyperoxaluria
Palmitoyl-protein thioesterase-1	Ceroid lipofuscinosis Adult form (CLN4, Kufs' disease)
Palmitoyl-protein thioesterase-1	Ceroid lipofuscinosis Infantile form (CLN1, Santavuori-Haltia disease)
Phenylalanine hydroxylase	Phenylketonuria
Phosphomannomutase-2	Congenital disorders of N-glycosylation CDG Ia (solely neurologic and neurologic-multivisceral forms)
Porphobilinogen deaminase	Acute Intermittent Porphyria
Purine nucleoside phosphorylase	Purine nucleoside phosphorylase deficiency
pyrimidine 5' nucleotidase	Hemolytic anemia and/or pyrimidine 5' nucleotidase deficiency
Sphingomyelinase	Niemann-Pick disease type A
Sphingomyelinase	Niemann-Pick disease type B
Sterol 27-hydroxylase	Cerebrotendinous xanthomatosis (cholestanol lipidosis)
Thymidine phosphorylase	Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE)
Trihexosylceramide α-galactosidase	Fabry's disease
tyrosine phenol-lyase	Tyrosinemia type 1
tyrosinase, e.g., OCA1	albinism, e.g., ocular albinism
UDP-GlcNAc:dolichyl-P NAcGlc phosphotransferase	Congenital disorders of N-glycosylation CDG Ij
UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase, sialin	Sialuria French type
Uricase	Lesch-Nyhan syndrome, gout
uridine diphosphate glucuronyl-transferase (e.g., UGT1A1)	Crigler–Najjar syndrome
α-1,2-Mannosyltransferase	Congenital disorders of N-glycosylation CDG Il (608776)

α-1,2-Mannosyltransferase	Congenital disorders of N-glycosylation, type I (pre-Golgi glycosylation defects)
α-1,3-Mannosyltransferase	Congenital disorders of N-glycosylation CDG Ii
α-D-Mannosidase	α-Mannosidosis, type I (severe) or II (mild)
α-L-Fucosidase	Fucosidosis
α-l-Iduronidase	Mucopolysaccharidosis MPS I H/S (Hurler-Scheie syndrome)
α-l-Iduronidase	Mucopolysaccharidosis MPS I-H (Hurler's syndrome)
α-l-Iduronidase	Mucopolysaccharidosis MPS I-S (Scheie's syndrome)
β-1,4-Galactosyltransferase	Congenital disorders of N-glycosylation CDG IId
β-1,4-Mannosyltransferase	Congenital disorders of N-glycosylation CDG Ik
β-D-Mannosidase	β-Mannosidosis
β-Galactosidase	Mucopolysaccharidosis MPS IV (Morquio's syndrome) Type IV-B
β-Glucuronidase	Mucopolysaccharidosis MPS VII (Sly's syndrome)
β-Hexosaminidase A	Tay-Sachs disease
β-Hexosaminidase B	Sandhoff's disease

In some embodiments, an effector described herein comprises an enzyme of Table A, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment. 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 anellosome encoding an enzyme of Table A, or a functional variant thereof is used for the treatment of a disease or disorder of Table A. In some embodiments, an anellosome 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 anellosome is used to deliver OCA1 or a functional variant thereof to a target cell, e.g., a retinal cell.

Table B. Exemplary non-enzymatic effectors and corresponding indications

Effector	**Indication**
Survival motor neuron protein (SMN)	spinal muscular atrophy
Dystrophin or micro-dystrophin	muscular dystrophy (e.g., Duchenne muscular dystrophy or Becker muscular dystrophy)
Complement protein, e.g., Complement factor C1	Complement Factor I deficiency
Complement factor H	Atypical hemolytic uremic syndrome
Cystinosin (lysosomal cystine transporter)	Cystinosis
Epididymal secretory protein 1 (HE1; NPC2 protein)	Niemann-Pick disease Type C2
GDP-fucose transporter-1	Congenital disorders of N-glycosylation CDG IIc (Rambam-Hasharon syndrome)

GM2 activator protein	GM2 activator protein deficiency (Tay-Sachs disease AB variant, GM2A)
Lysosomal transmembrane CLN3 protein	Ceroid lipofuscinosis Juvenile form (CLN3, Batten disease, Vogt-Spielmeyer disease)
Lysosomal transmembrane CLN5 protein	Ceroid lipofuscinosis Variant late infantile form, Finnish type (CLN5)
Na phosphate cotransporter, sialin	Infantile sialic acid storage disorder
Na phosphate cotransporter, sialin	Sialuria Finnish type (Salla disease)
NPC1 protein	Niemann-Pick disease Type C1/Type D
Oligomeric Golgi complex-7	Congenital disorders of N-glycosylation CDG IIe
Prosaposin	Prosaposin deficiency
Protective protein/cathepsin A (PPCA)	Galactosialidosis (Goldberg's syndrome, combined neuraminidase and β-galactosidase deficiency)
Protein involved in mannose-P-dolichol utilization	Congenital disorders of N-glycosylation CDG If
Saposin B	Saposin B deficiency (sulfatide activator deficiency)
Saposin C	Saposin C deficiency (Gaucher's activator deficiency)
Sulfatase-modifying factor-1	Mucosulfatidosis (multiple sulfatase deficiency)
Transmembrane CLN6 protein	Ceroid lipofuscinosis Variant late infantile form (CLN6)
Transmembrane CLN8 protein	Ceroid lipofuscinosis Progressive epilepsy with intellectual disability
vWF	von Willebrand disease
Factor I (fibrinogen)	Afibrinogenomia

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 anellosome encoding an erythropoietin, or a functional variant thereof is used for stimulating erythropoiesis. In some embodiments, an anellosome 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 anellosome 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 B, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment. In some embodiments, an anellosome encoding a polypeptide of Table B, or a functional variant thereof is used for the treatment of a disease or disorder of Table B. In some embodiments, an anellosome 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 anellosome 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 A or Table B 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 A or Table B herein. In some embodiments, an effector described herein comprises a transporter protein, e.g., a transporter protein listed in Table B 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.

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 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, 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 composition or anellosome described herein includes a polypeptide linked to a ligand that is capable of targeting a specific location, tissue, or cell.

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. In addition, one promoter element can increase an amount of products expressed for multiple DNA sequences attached in tandem. 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, 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 anellosome, e.g., synthetic anellosome, may include sequences that encode one or more replication proteins. In some embodiments, the anellosome 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 anellosome 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 anellosome 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 single stranded DNA virus, e.g., Anellovirus, Bidnavirus, Circovirus, Geminivirus, Genomovirus, Inovirus, Microvirus, Nanovirus, Parvovirus, and Spiravirus. In some embodiments, the sequence has homology or identity to one or more sequence from a double stranded DNA virus, e.g., Adenovirus, Ampullavirus, Ascovirus, Asfarvirus, Baculovirus, Fusellovirus, Globulovirus, Guttavirus, Hytrosavirus, Herpesvirus, Iridovirus, Lipothrixvirus, Nimavirus, and Poxvirus. In some embodiments, the sequence has homology or identity to one or more sequence from an RNA virus, e.g., Alphavirus, Furovirus, Hepatitis virus, Hordeivirus, Tobamovirus, Tobravirus, Tricornavirus, Rubivirus, Birnavirus, Cystovirus, Partitivirus, and Reovirus.

In some embodiments, the genetic element may comprise one or more sequences from a non-pathogenic virus, e.g., a symbiotic virus, e.g., a commensal virus, e.g., a native virus, e.g., an *Anellovirus*. Recent changes in nomenclature have classified the three *Anelloviruses* able to infect human cells into Alphatorquevirus (TT), Betatorquevirus (TTM), and Gammatorquevirus (TTMD) Genera of the *Anelloviridae* family of viruses. To date *Anelloviruses* have not been linked to any human disease. In some embodiments, the genetic element may comprise a sequence with homology or identity to a Torque Teno Virus (TT), a non-enveloped, single-stranded DNA virus with a circular, negative-sense genome. In some embodiments, the genetic element may comprise a sequence with homology or identity to a SEN virus, a Sentinel virus, a TTV-like mini virus, and a TT virus. Different types of TT viruses have been described including TT virus genotype 6, TT virus group, TTV-like virus DXL1, and TTV-like virus DXL2. In some embodiments, the genetic element may comprise a sequence with homology or identity to a smaller virus, Torque Teno-like Mini Virus (TTM), or a third virus with a genomic size in between that of TTV and TTMV, named Torque Teno-like Midi Virus (TTMD). In some embodiments, the genetic element may comprise one or more sequences or a fragment of a sequence from a non-pathogenic virus having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide sequences described herein.

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

Table 41: Examples of *Anelloviruses* and their sequences. Accessions numbers and related sequence information may be obtained at www.ncbi.nlm.nih.gov/genbank/, 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., adenovirus, herpes virus, pox virus, vaccinia virus, SV40, papilloma virus, an RNA virus such as a retrovirus, e.g., lentivirus, a single-stranded RNA virus, e.g., hepatitis virus, or a double-stranded RNA virus e.g., rotavirus. Since, in some embodiments, recombinant retroviruses are defective, assistance may be provided order to produce infectious particles. Such assistance can be provided, e.g., by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences within the LTR. Suitable cell lines for replicating the anellosomes described herein include cell lines known in the art, e.g., A549 cells, which can be modified as described herein. Said genetic element can additionally contain a gene encoding a selectable marker so that the desired genetic elements can be identified.

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.

In some embodiments, the genetic element comprises a nucleotide sequence with at least about 75% nucleotide sequence identity, at least about 80%, 85%, 90% 95%, 96%, 97%, 98%, 99% or 100% nucleotide sequence identity to any one of the nucleotide sequences described herein, e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, 17, or 41. 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.

Gene Editing Component

The genetic element of the anellosome 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 anellosome 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 anellosome. 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 anellosome 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 anellosome 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 anellosome 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 anellosome 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 Cas9 generates only a single-strand break; a catalytically inactive 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 anellosome 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.

As used herein, a "biologically active portion of an effector domain" is a portion that maintains the function (e.g. completely, partially, or minimally) of an effector domain (e.g., a "minimal" or "core" domain). In some embodiments, the anellosome includes a gene encoding a fusion of a dCas9 with all or a portion of one or more effector domains to create a chimeric protein useful in the methods described herein. Accordingly, in some embodiments, the anellosome includes a gene encoding a dCas9-methylase fusion. In other some embodiments, the anellosome includes a gene encoding a dCas9-enzyme fusion with a site-specific gRNA to target an endogenous gene.

In other aspects, the anellosome 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.

Proteinaceous Exterior

In some embodiments, the anellosome, e.g., synthetic anellosome, 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 anellosomes 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 anellosome (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 anellosome (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 sequence or a capsid protein sequence as listed in any of Tables 1-18, A1-A12, B1-B5, C1-C5, D1-D10, or 20-37. 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 any one of the nucleotide sequences described herein, e.g., an Anellovirus capsid sequence or a capsid protein sequence as listed in any of Tables A1-A12, B1-B5, C1-C5, D1-D10, or 20-37. In some embodiments, the protein comprises a capsid protein or a functional fragment of a capsid protein that is encoded by a capsid nucleotide sequence or a sequence having at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% nucleotide sequence identity to any one of the nucleotide sequences described herein, e.g., an Anellovirus capsid sequence or a capsid protein 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.

In some embodiments, the anellosome 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 any one of the amino acid sequences

described herein, e.g., an Anellovirus capsid sequence or a capsid protein sequence 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, the anellosome comprises a nucleotide sequence encoding a capsid protein or a functional fragment of a capsid protein or a sequence having at least about 60%, 65%, 70%, 75%, 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 capsid sequence or a capsid protein sequence 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, the anellosome comprises a nucleotide sequence encoding an amino acid sequence having about position 1 to about position 150 (e.g., or any subset of amino acids within each range, e.g., about position 20 to about position 35, about position 25 to about position 30, about position 26 to about 30), about position 150 to about position 390 (e.g., or any subset of amino acids within each range, e.g., about position 200 to about position 380, about position 205 to about position 375, about position 205 to about 371), about 390 to about position 525, about position 525 to about position 850 (e.g., or any subset of amino acids within each range, e.g., about position 530 to about position 840, about position 545 to about position 830, about position 550 to about 820), about 850 to about position 950 (e.g., or any subset of amino acids within each range, e.g., about position 860 to about position 940, about position 870 to about position 930, about position 880 to about 923) of the amino acid sequences described herein, an Anellovirus amino acid sequence, 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 shown in Figure 1, or a functional fragment thereof. In some embodiments, the protein comprises an amino acid sequence or a functional fragment thereof or a sequence having at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to about position 1 to about position 150 (e.g., or any subset of amino acids within each range as described herein), about position 150 to about position 390, about position 390 to about position 525, about position 525 to about position 850, about position 850 to about position 950 of the amino acid sequences described herein, an Anellovirus amino acid sequence, 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 as shown in Figure 1.

In some embodiments, the protein comprises an amino acid sequence or a functional fragment thereof or a sequence having at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of the amino acid sequences or ranges of amino acids described herein, an Anellovirus amino acid sequence, 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 shown in Figure 1. 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 anellosome lacks lipids in the proteinaceous exterior. In some embodiments, the anellosome lacks a lipid bilayer, e.g., a viral envelope. In some embodiments, the interior of the anellosome is entirely covered (e.g., 100% coverage) by a proteinaceous exterior. In some embodiments, the interior of the anellosome 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 anellosome.

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: 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 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.

II. Vectors

The genetic element described herein may be included in a vector. Suitable vectors as well as methods for their manufacture and their use are well known in the prior art.

In one aspect, the invention includes a vector 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.

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 vector 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 vector includes regulatory elements, nucleic acid sequences homologous to target genes, and 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 vector 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 vector 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.

Model systems for quantifying anellosome activity

Anellosomes described herein can be assayed in a number of animal models and in vitro models. For instance, a mouse model for spinal muscular atrophy can be used to quantify the activity of an anellosome encoding SMN or a functional variant thereof. In some embodiments, a mouse model for muscular dystrophy can be used to quantify the activity of an anellosome encoding micro-dystrophin or a functional variant thereof. In some embodiments, a UGT1A1-deficient mouse model can be used to quantify the activity of an anellosome encoding UGT1A1 or a functional variant thereof.

III. Compositions

The anellosome or vector 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} anellosomes. In some embodiments, the pharmaceutical composition comprises about 10^5-10^{15}, 10^5-10^{10}, or 10^{10}-10^{15}

anellosomes. 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 anellosome. 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 anellosome, e.g., as determined according to the method of Example 18. In some embodiments, the pharmaceutical composition comprises sufficient anellosomes 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 anellosomes per cell to a population of the eukaryotic cells. In some embodiments, the pharmaceutical composition comprises sufficient anellosomes to deliver at least about 1 x 10^4, 1 x 10^5, 1 x 10^6, 1 x or 10^7, or about 1 x 10^4-1 x 10^5, 1 x 10^4-1 x 10^6, 1 x 10^4-1 x 10^7, 1 x 10^5-1 x 10^6, 1 x 10^5-1 x 10^7, or 1 x 10^6-1 x 10^7 copies of a genetic element comprised in the anellosomes 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 anellosome 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 anellosome, *Anellovirus*, anellovector, 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 anellosome or composition comprising a anellosome described herein may also include one or more heterologous moiety. In one aspect, the anellosome or composition comprising a anellosome 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 anellosome. In some embodiments, a heterologous moiety may be administered with the anellosome.

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

In another aspect, the invention includes a pharmaceutical composition comprising a anellosome 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.

Viruses

In some embodiments, the composition may further comprise a virus as a heterologous moiety, e.g., a single stranded DNA virus, e.g., Anellovirus, Bidnavirus, Circovirus, Geminivirus, Genomovirus,

Inovirus, Microvirus, Nanovirus, Parvovirus, and Spiravirus. In some embodiments, the composition may further comprise a double stranded DNA virus, e.g., Adenovirus, Ampullavirus, Ascovirus, Asfarvirus, Baculovirus, Fusellovirus, Globulovirus, Guttavirus, Hytrosavirus, Herpesvirus, Iridovirus, Lipothrixvirus, Nimavirus, and Poxvirus. In some embodiments, the composition may further comprise an RNA virus, e.g., Alphavirus, Furovirus, Hepatitis virus, Hordeivirus, Tobamovirus, Tobravirus, Tricornavirus, Rubivirus, Birnavirus, Cystovirus, Partitivirus, and Reovirus. In some embodiments, the anellosome is administered with a virus as a heterologous moiety.

In some embodiments, the heterologous moiety may comprise a non-pathogenic, e.g., symbiotic, commensal, native, virus. In some embodiments, the non-pathogenic virus is one or more anelloviruses, e.g., Alphatorquevirus (TT), Betatorquevirus (TTM), and Gammatorquevirus (TTMD). In some embodiments, the anellovirus may include a Torque Teno Virus (TT), a SEN virus, a Sentinel virus, a TTV-like mini virus, a TT virus, a TT virus genotype 6, a TT virus group, a TTV-like virus DXL1, a TTV-like virus DXL2, a Torque Teno-like Mini Virus (TTM), or a Torque Teno-like Midi Virus (TTMD). In some embodiments, the non-pathogenic virus comprises one or more sequences having at least at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide sequences described herein, e.g., as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, 17, or 41.

In some embodiments, the heterologous moiety may comprise one or more viruses that are identified as lacking in the subject. For example, a subject identified as having dyvirosis may be administered a composition comprising an anellosome and one or more viral components or viruses that are imbalanced in the subject or having a ratio that differs from a reference value, e.g., a healthy subject.

In some embodiments, the heterologous moiety may comprise one or more non-anelloviruses, e.g., adenovirus, herpes virus, pox virus, vaccinia virus, SV40, papilloma virus, an RNA virus such as a retrovirus, e.g., lenti virus, a single-stranded RNA virus, e.g., hepatitis virus, or a double-stranded RNA virus e.g., rotavirus. In some embodiments, the anellosome or the virus is defective, or requires assistance in order to produce infectious particles. Such assistance can be provided, e.g., by using helper cell lines that contain a nucleic acid, e.g., plasmids or DNA integrated into the genome, encoding one or more of (e.g., all of) the structural genes of the replication defective anellosome or virus under the control of regulatory sequences within the LTR. Suitable cell lines for replicating the anellosomes described herein include cell lines known in the art, e.g., A549 cells, which can be modified as described herein.

Targeting Moiety

In some embodiments, the composition or anellosome 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 anellosome 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 anellosome described herein may further comprise a tag to label or monitor the anellosome 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 anellosome 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 anellosome 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 anellosome 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 anellosome 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 anellosome 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 anellosome 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.

IV. Hosts

The invention is further directed to a host or host cell comprising a anellosome described herein. In some embodiments, the host or host cell is a plant, insect, bacteria, fungus, vertebrate, mammal (e.g., human), or other organism or cell. In certain embodiments, as confirmed herein, provided anellosomes infect a range of different host cells. Target host cells include cells of mesodermal, endodermal, or ectodermal origin. Target host cells include, e.g., epithelial cells, muscle cells, white blood cells (e.g., lymphocytes), kidney tissue cells, lung tissue cells.

In some embodiments, the anellosome is substantially non-immunogenic in the host. The anellosome or genetic element fails to produce an undesired substantial response by the host's immune system. Some immune responses include, but are not limited to, humoral immune responses (e.g., production of antigen-specific antibodies) and cell-mediated immune responses (e.g., lymphocyte proliferation).

In some embodiments, a host or a host cell is contacted with (e.g., infected with) an anellosome. In some embodiments, the host is a mammal, such as a human. The amount of the anellosome in the host can be measured at any time after administration. In certain embodiments, a time course of anellosome growth in a culture is determined.

In some embodiments, the anellosome, e.g., an anellosome as described herein, is heritable. In some embodiments, the anellosome is transmitted linearly in fluids and/or cells from mother to child. In some embodiments, daughter cells from an original host cell comprise the anellosome. In some embodiments, a mother transmits the anellosome to child with an efficiency of at least 25%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%, or a transmission efficiency from host cell to daughter cell at least 25%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%. In some embodiments, the anellosome in a host cell has a transmission efficiency during meiosis of at 25%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%. In some embodiments, the anellosome in a host cell has a transmission efficiency during mitosis of at least 25%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or 99%. In some embodiments, the anellosome in a cell has a transmission efficiency between about 10%-20%, 20%-30%, 30%-40%, 40%-50%, 50%-60%,

60%-70%, 70%-75%, 75%-80%, 80%-85%, 85%-90%, 90%-95%, 95%-99%, or any percentage therebetween.

In some embodiments, the anellosome, e.g., anellosome replicates within the host cell. In one embodiment, the anellosome is capable of replicating in a mammalian cell, e.g., human cell. In other embodiments, the anellosome is replication deficient or replication incompetent.

While in some embodiments the anellosome replicates in the host cell, the anellosome does not integrate into the genome of the host, e.g., with the host's chromosomes. In some embodiments, the anellosome has a negligible recombination frequency, e.g., with the host's chromosomes. In some embodiments, the anellosome has a recombination frequency, e.g., less than about 1.0 cM/Mb, 0.9 cM/Mb, 0.8 cM/Mb, 0.7 cM/Mb, 0.6 cM/Mb, 0.5 cM/Mb, 0.4 cM/Mb, 0.3 cM/Mb, 0.2 cM/Mb, 0.1 cM/Mb, or less, e.g., with the host's chromosomes.

V. Methods of Use

The anellosomes and compositions comprising anellosomes 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 anellosomes may be administered alone or formulated as a pharmaceutical composition.

The anellosomes 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 a anellosome or composition comprising same, e.g., as described herein, may result in delivery of a genetic element comprised by the anellosome to a target cell, e.g., in a subject.

An anellosome 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 anellosome 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 anellosome 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 anellosome, 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 anellosome 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 anellosome 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 anellosome described herein, or a composition comprising the anellosome, 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 anellosome modulates (e.g., increases or decreases) an activity or function in a cell with which the anellosome is contacted. In some embodiments, the anellosome 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 anellosome is contacted. In some embodiments, the anellosome decreases viability of a cell, e.g., a cancer cell, with which the anellosome 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 anellosome 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 anellosome 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 anellosome increases apoptosis of a cell, e.g., a cancer cell, e.g., by increasing caspase-3 activity, with which the anellosome 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 anellosome 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 anellosome is contacted, e.g., by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more.

VI. Methods of Production

Producing the Genetic Element

Methods of making the genetic element of the anellosome 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).

In some embodiments, the genetic element may be designed using computer-aided design tools. The anellosome 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 anellosome. The segments or ORFs may be assembled into the anellosome, e.g., in vitro recombination or unique restriction sites at 5' and 3' ends to enable ligation.

The genetic element can alternatively be synthesized with a design algorithm that parses the anellosome into oligo-length fragments, creating optimal 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 or segment of the genetic element 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). In particular, the AnyDot-chips allow for 10x-50x enhancement of nucleotide fluorescence signal detection. AnyDot.chips and methods for using them are described in part in International Publication Application Nos. WO 02088382, WO 03020968, WO 0303 1947, WO 2005044836, PCTEP 05105657, PCMEP 05105655; and German Patent Application Nos. DE 101 49 786, DE 102 14 395, DE 103 56 837, DE 10 2004 009 704, DE 10 2004 025 696, DE 10 2004 025 746, DE 10 2004 025 694, DE 10 2004 025 695, DE 10 2004 025 744, DE 10 2004 025 745, and DE 10 2005 012 301.

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. The sequence can then be deduced by identifying which base is being incorporated into the growing complementary strand of the target nucleic acid by the catalytic activity of the nucleic acid polymerizing enzyme at each step in the sequence of base additions. A polymerase on the target nucleic acid molecule complex is provided in a position suitable to move along the target nucleic acid molecule

and extend the oligonucleotide primer at an active site. A plurality of labeled types of nucleotide analogs are provided proximate to the active site, with each distinguishably type of nucleotide analog being complementary to a different nucleotide in the target nucleic acid sequence. The growing nucleic acid strand is extended by using the polymerase to add a nucleotide analog to the nucleic acid strand at the active site, where the nucleotide analog being added is complementary to the nucleotide of the target nucleic acid at the active site. The nucleotide analog added to the oligonucleotide primer as a result of the polymerizing step is identified. The steps of providing labeled nucleotide analogs, polymerizing the growing nucleic acid strand, and identifying the added nucleotide analog are repeated so that the nucleic acid strand is further extended and the sequence of the target nucleic acid is determined.

In some embodiments, shotgun sequencing is performed. In shotgun sequencing, DNA is broken up randomly into numerous small segments, which are sequenced using the chain termination method to obtain reads. Multiple overlapping reads for the target DNA are obtained by performing several rounds of this fragmentation and sequencing. Computer programs then use the overlapping ends of different reads to assemble them into a continuous sequence.

In some embodiments, 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, this is done both in context of larger regions of the anellosome genome (e.g., inserting effectors into a specific site in the genome, or replacing viral ORFs with effectors).

In another example, when 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 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 in a form other than a single stranded circular DNA. For example, the genetic element may be introduced into the host cell as 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, e.g., as described in Example 35. Generally, *in vitro* circularized DNA constructs can be produced by digesting 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 anellosome, 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.

Producing the Anellosome

The genetic elements and vectors comprising the genetic elements prepared as described herein can be used in a variety of ways to express the anellosome in appropriate host cells. In some embodiments, the genetic element and vectors comprising the genetic element are transfected in appropriate host cells and the resulting RNA may direct the expression of the anellosome gene products, e.g., non-pathogenic protein and protein binding sequence, at high levels. Host cell systems which provide for high levels of expression include continuous cell lines that supply viral functions, such as cell lines superinfected with APV or MPV, respectively, cell lines engineered to complement APV or MPV functions, etc.

In some embodiments, the anellosome is produced as described in any of Examples 1, 2, 5, 6, or 15-17.

In some embodiments, the anellosome is cultivated in continuous animal cell lines in vitro. 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. 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 susceptible to infection with anellosomes include, but are not limited to cell lines of human or primate origin, such as human or primate kidney carcinoma cell lines.

In some embodiments, the genetic elements and vectors comprising the genetic elements are transfected into cell lines that express a viral polymerase protein in order to achieve expression of the anellosome. To this end, transformed cell lines that express an anellosome 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 anellosome disclosed herein, a genetic element or vector comprising the genetic element disclosed herein may be used to transfect cells which provide anellosome proteins and functions required for replication and production. Alternatively, cells may be transfected with helper virus before, during, or after transfection by the genetic element or vector comprising the genetic element disclosed herein. In some embodiments, a helper virus may be useful to complement production of an incomplete viral particle. The helper virus may have a conditional growth defect, such as host range restriction or temperature sensitivity, which allows the subsequent selection of transfectant viruses. In some embodiments, a helper virus may provide one or more replication proteins utilized by the host cells to achieve expression of the anellosome. 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, a helper virus comprises an antiviral sensitivity.

The genetic element or vector comprising the genetic element disclosed herein can be replicated and produced into anellosome particles by any number of techniques known in the art, as 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.

The production of anellosome-containing cell cultures according to the present invention can be carried out in different scales, such as in flasks, roller bottles or bioreactors. The media used for the cultivation of the cells to be infected are known to the skilled person and can 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 anellosomes.

The purification and isolation of anellosomes can be performed according to methods known by the skilled person in virus production and is described for example by Rinaldi, et al., DNA Vaccines: Methods and Protocols (Methods in Molecular Biology), 3rd ed. 2014, Humana Press.

In one aspect, the present invention includes a method for the in vitro replication and propagation of the anellosome as described herein, which may comprise the following steps: (a) transfecting a linearized genetic element into a cell line sensitive to anellosome infection; (b) harvesting the cells and isolating cells showing the presence of the genetic element; (c) culturing the cells obtained in step (b) 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).

In some embodiments, an anellosome may be introduced to a host cell line grown to a high cell density. In some embodiments, the anellosome 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.

VII. Administration/Delivery

The composition (e.g., a pharmaceutical composition comprising an anellosome 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.

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 anellosome to a subject. The method includes administering a pharmaceutical composition comprising an anellosome as described herein to the subject. In some embodiments, the administered anellosome 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 anellosome may include one or more of the sequences (e.g., nucleic acid sequences or nucleic acid sequences encoding amino acid sequences thereof) in any of Tables A1-A12, B1-B5, C1-C5, or 1-18 or a sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide sequences or a sequence that is complementary to the sequence in any of Tables A1-A12, B1-B5, C1-C5, or 1-18. The anellosome 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 of the sequences in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, 17, or 41. The anellosome 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 any one of the amino acid sequences in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18. The anellosome 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 any one of the amino acid sequences in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18. The anellosome may include one or more of the sequences in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, 17, or 41, or a sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide

sequences or a sequence that is complementary to the sequence in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, 17, or 41.

In some embodiments, the anellosome 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 anellosome 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 anellosome 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 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 anellosome 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 anellosome 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 anellosome is sufficient to compete with chronic or acute viral infection. In certain embodiments, the anellosome may be administered prophylactically to protect from viral infections (e.g. a provirotic). In some embodiments, the anellosome 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).

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 1: Preparation of Anellosomes: Design and synthesis of a synthetic anellosome that inhibits interferon (IFN) expression

Example 2: Large-Scale Production of Anellosomes (Anellosome A and/or B): Production and propagation of anellosomes

Example 3: Effects of Anellosomes in vitro (Anellosome A): In vitro assessment of expression and effector function, e.g., expression of the miRNA, of the anellosome after cell infection

Example 4: Immunologic effects of Anellosomes (Anellosome A): in vivo effector function, e.g., expression of the miRNA, of the anellosome after administration

Example 5: Preparation of synthetic anellosomes: In vitro production of a synthetic anellosome

Example 6: Assembly and infection of anellosomes: In vitro production of infectious anellosomes using synthetic DNA sequences as described in Example 5

Example 7: Selectivity of anellosomes: Synthetic anellosomes produced *in vitro* infect cell lines of a variety of tissue origins

Example 8: Identification and use of protein binding sequences: putative protein-binding sites in the *Anellovirus* genome

Example 9: An *Anellovirus* genome

Example 10: Nucleotide insertions of various lengths into an *Anellovirus* genome: addition of DNA sequences of various lengths into an *Anellovirus* genome

Example 11: Exemplary cargo to be delivered: exemplary classes of nucleic acid and protein payloads in an anellosome

Example 12: Exemplary payload integration loci

Example 13: Defined categories of *Anellovirus* and conserved regions thereof

Example 14: Replication-deficient anellosomes and helper viruses

Example 15: Manufacturing process for replication-competent anellosomes

Example 16: Manufacturing process of replication-deficient anellosomes: recovery and scaling up of production of replication-deficient anellosomes

Example 17: Production of anellosomes using suspension cells: production of anellosomes in cells in suspension.

Example 18: Quantification of anellosome genome equivalents by qPCR: development of a hydrolysis probe-based quantitative PCR assay to quantify anellosomes

Example 19: Utilizing anellosomes to express an exogenous protein in mice: use of an anellosome to express a functional model protein in vivo

Example 20: Genome alignments to determine whether anellosome DNA integrated into host genomes

Example 21: Assessment of anellosome integration into a host genome

Example 22: Functional effects of an anellosome expressing an exogenous microRNA sequence: use of an anellosome to express a functional nucleic acid effector

Example 23: Preparation and production of anellosomes to express exogenous non-coding RNAs: use of anellosomes to express exogenous small non-coding RNAs

Example 24: Conservation in *Anellovirus* clades: identification of seven clades within the *Alphatorquevirus* genus

Example 25: Expression of an endogenous miRNA from an anellosome and deletion of the endogenous miRNA

Example 26: Localization of *Anellovirus* ORFs

Example 27: Characterization of regions required for anellosome development

Example 28: Anellosome delivery of exogenous proteins *in vivo*: This example demonstrates in vivo effector function (e.g. expression of proteins) of anellosomes after administration

Example 29: Identification of precursor miRNAs (pre-mIRs) in Anelloviruses: computational and experimental approaches to identify novel precursor miRNAs encoded by various Anelloviruses

Example 30: Determination of the endogenous target of Anellovirus pre-miRs: analysis to determine endogenous targets and potentially therapeutically relevant target pathways of pre-miRs encoded by various strains of Anelloviruses

Example 31: Making an anellosome encoding a native Anellovirus pre-miR: a process to package either the replicating or non-replicating form of anellosomes expressing native Anellovirus pre-miRs

Example 32: Utilizing Anellovirus pre-miRs a tumor suppressor in an *in vitro* cell culture model: phenotypic effect of candidate pre-miRs identified as tumor suppressive from analysis, e.g., as described in Example 29

Example 33: Utilizing Anellovirus pre-miRs as tumor supppresors *in vivo*: *in vivo* experiments to confirm the tumor suppressive effect of a tumor suppressive Anellovirus pre-miRs and cancer cell lines from *in vitro* analysis, as described in Example 32

Example 34: Tandem copies of the Anellovirus genome

Example 35: *In vitro* circularized Anellovirus genomes: constructs comprising circular, double stranded Anelloviral genome DNA with minimal non-viral DNA

Example 36: Modelling ORF1 and identification of conserved residues and domains: modelling of ORF1 proteins of Betatorqueviruses and defining putative domains

Example 37: Production of anellosomes containing chimeric ORF1 with hypervariable domains from different Torque Teno Virus strains

Example 38: Production of chimeric ORF1 containing non-TTV protein/peptides in place of hypervariable domains

Example 39: Anellosome delivery of secreted enzyme *in vivo*

Example 40: Design of an anellosome harboring a DNA payload

Example 41: Anellosomes based on tth8 and LY2 each successfully transduced the EPO gene into lung cancer cells

Example 42: Anellosomes with therapeutic transgenes can be detected in vivo after intravenous (i.v.) administration

Example 43: Coding sequence size distribution in Anelloviruses

Example 44: A highly conserved motif to characterize ORF2

Example 45: Evidence for full-length Anellovirus ORF1 mRNA in humans
Example 46: In vitro circularized genome as input material for producing anellosomes in vitro
Example 47: Identification of conserved secondary structural motifs in Anellovirus ORF1

Example 1: Preparation of Anellosomes

This example describes the design and synthesis of a synthetic anellosome that inhibits interferon (IFN) expression.

An anellosome (Anellosome A) is designed starting with 1) a DNA sequence for a capsid gene encoding a non-pathogenic packaging enclosure (Arch Virol (2007) 152: 1961-1975), Accession Number: A7XCE8.1 (ORF11_TTW3); 2) a DNA sequence coding for a microRNA that targets a host gene (e.g. IFN) (PLOS Pathogen (2013), 9(12), e1003818), Accession number: AJ620231.1; and 3) a DNA sequence (Journal of Virology (2003), 77(24), 13036-13041) that binds to a specific region in the capsid protein, (e.g., specific region of capsid having an Accession Number: Q99153.1).

To this sequence is added 1kb non-coding DNA sequences (Anellosome B). The designed anellosome (Figure 2) is chemically synthesized into 3 kb (total size), which is sequence verified.

The anellosome sequence is transfected into human embryonic kidney 293T cells (1 mg per 10^5 cells on 12-well plates) with JetPEI reagent (PolyPlus-transfection, Illkirch, France) as recommended by the manufacturer. Controls transfections are included with vector alone or cells transfected with JetPEI alone and transfection efficiencies are optimized with a reporter plasmid encoding GFP. Fluorescence of control transfections is measured to ensure properly transfected cells. Transfected cultures are incubated overnight at 37°C and 5% carbon dioxide.

After 18 hrs, the cells are washed three times with PBS before adding fresh medium. The supernatant is collected for ultracentrifugation and harvest of anellosomes as follows. The medium is cleared by centrifugation at 4,000 x *g* for 30 min and then at 8,000 x *g* for 15 min to remove cells and cell debris. The supernatant is then filtered through 0.45-µm-pore-size filters. Anellosomes are pelleted at 27,000 rpm for 1 hr through a 5% sucrose cushion (5 ml) and resuspended in 1x phosphate-buffered saline (PBS) plus 0.1% bacitracin in 1/100 of the original volume. The concentrated anellosomes are centrifuged through a 20 to 35% sucrose step gradient at 24,000 rpm for 2 hr. The anellosome band at the gradient junction is collected. The anellosomes are then diluted with 1x PBS and pelleted at 27,000 rpm for 1 hr. The anellosome pellets are resuspended in 1x PBS and further purified through a 20 to 35% continuous sucrose gradient.

Example 2: Large-Scale Production of Anellosomes (Anellosome A and/or B)

This example describes production and propagation of anellosomes.

Purified anellosomes as described in Example 1 are prepared for large-scale amplification in spinner flasks with producer A549 cells grown in suspension. A549 cells are maintained in F12K medium, 10% fetal bovine serum, 2 mM glutamine and antibiotics. A549 cells are infected with anellosomes at an anellosome load of 10^6 anellosomes to produce ~$1x10^7$ anellosome particles after an incubation at 37°C and 5% carbon dioxide for 24 hrs. Cells are then washed three times with PBS and incubated with fresh medium for 6 hrs.

For anellosome purification, two ultracentrifugation steps based on cesium chloride gradients are performed followed by dialysis as follows (Bio-Protocol (2012) Bio101: e201). Cells are removed by centrifugation (6000 x g for 10 min) and the supernatant is filtered through 0.8 and then 0.2 µm filters. The filtrate is concentrated by passage through filter membranes (100,000 mw) to a volume of 8 ml. The retentate is loaded into a cesium sulfate solution and centrifuged at 247,000 x g for 20 h. Anellosome bands are removed, placed into 14,000 mw cutoff dialysis tubing, and dialyzed. A further concentration may be performed, if desired.

Example 3: Effects of Anellosomes in vitro (Anellosome A)

This example describes in vitro assessment of expression and effector function, e.g., expression of the miRNA, of the anellosome after cell infection.

The effect of purified anellosomes as described in Example 1 is assessed in vitro through endogenous gene regulation (e.g. IFN signaling). HEK293T cells are co-transfected with dual luciferase plasmids (firefly luciferase with an interferon-stimulated response element (ISRE) based promoter and transfection control Renilla luciferase with constitutive promoter): Luciferase reporter mix (pcDNA3.1dsRluc to pISRE-Luc at 1:4 ratio (Clonetech)) (J Virol (2008), 82: 9823–9828).

Anellosomes are administered at multiplicity of infection of 10^7 to HEK293T cells seeded in a 6-well plate (2 sets of triplicates-3 control wells and 3 experimental wells with Anellosome A).

After 48 hours, the media is replaced with new media with or without 100 u/ml of universal type I interferon (PBL, Piscataway, NJ). Sixteen hours after IFN treatment, a dual-luciferase assay (J Virol (2008), 82: 9823–9828) is performed to determine IFN signaling. Firefly luciferase is normalized to Renilla luciferase expression to control for transfection differences. The fold induction of the ISRE ffLuc reporter is calculated by dividing the comparable experimental wells by the control wells and induction of each condition is compared relative to the negative control.

In an embodiment, a decreased luciferase signal in the anellosome treatment group compared to a control will indicate that the anellosomes decrease IFN production in the cells.

Example 4: Immunologic effects of Anellosomes (Anellosome A)

This example describes in vivo effector function, e.g., expression of the miRNA, of the anellosome after administration.

Purified anellosomes prepared as described in Examples 1 and 2 are intravenously administered to healthy pigs at various doses using hundred-fold dilutions starting from 10^{14} genome equivalents per kilogram down to 0 genome equivalents per kilogram. In order to evaluate the effects on immune tolerance, pigs are injected daily for 3 days with the dosages of anellosomes specified above or vehicle control PBS and sacrificed after 3 days.

Spleen, bone marrow and lymph nodes are harvested. Single cell suspensions are prepared from each of the tissues and stained with extracellular markers for MHC-II, CD11c, and intracellular IFN. MHC+, CD11c+, IFN+ antigen presenting cells are analyzed via flow cytometry from each tissue, e.g., wherein a cell that is positive for a given one of the above-mentioned markers is a cell that exhibits higher fluorescence than 99% of cells in a negative control population that lack expression of the marker but is otherwise similar to the the assay population of cells, under the same conditions.

In an embodiment, a decreased number of IFN+ cells in the anellosome treatment group compared to the control will indicate that the anellosomes decrease IFN production in cells after administration.

Example 5: Preparation of synthetic anellosomes.

This example demonstrates in vitro production of a synthetic anellosome.

DNA sequences from LY1 and LY2 strains of TTMiniV (Eur Respir J. 2013 Aug;42(2):470-9), between the EcoRV restriction enzyme sites, were cloned into a kanamycin vector (Integrated DNA Technologies). Anellosomes including DNA sequences from the LY1 and LY2 strains of TTMiniV are referred to as Anellosome 1 (Anello 1) and Anellosome 2 (Anello 2) respectively, in Examples 6 and 7 and in Figures 6A-10B. Cloned constructs were transformed into 10-Beta competent E.coli. (New England Biolabs Inc.), followed by plasmid purification (Qiagen) according to the manufacturer's protocol.

DNA constructs (Figure 3 and Figure 4) were linearized with EcoRV restriction digest (New England Biolabs, Inc.) at 37 degree Celsius for 6 hours, yielding double-stranded linear DNA fragments containing the TTMiniV genome, and excluding bacterial backbone elements (such as the origin of replication and selectable markers). This was followed by agarose gel electrophoresis, excision of a correctly size DNA band for the TTMiniV genome fragment (2.9 kilobase pairs), and gel purification of DNA from excised agarose bands using a gel extraction kit (Qiagen) according to the manufacturer's protocol.

Example 6: Assembly and infection of anellosomes

This example demonstrates successful in vitro production of infectious anellosomes using synthetic DNA sequences as described in Example 5.

The double-stranded linearized gel-purified Anellovirus genome DNA (obtained in Example 5) was transfected into either HEK293T cells (human embryonic kidney cell line) or A549 cells (human lung carcinoma cell line), either in an intact plasmid or in linearized form, with lipid transfection reagent (Thermo Fisher Scientific). 6 ug of plasmid or 1.5 ug of linearized Anellovirus genome DNA was used for transfection of 70% confluent cells in T25 flasks. Empty vector backbone lacking the viral sequences included in the anellosome was used as a negative control. Six hours post-transfection, cells were washed with PBS twice and were allowed to grow in fresh growth medium at 37 degrees Celsius and 5% carbon dioxide. DNA sequences encoding the human Ef1alpha promoter followed by YFP gene were synthesized from IDT. This DNA sequence was blunt end ligated into a cloning vector (Thermo Fisher Scientific). The resulting vector was used as a control to assess transfection efficiency. YFP was detected using a cell imaging system (Thermo Fisher Scientific) 72 hours post transfection. The transfection efficiencies of HEK293T and A549 cells were calculated as 85% and 40% respectively (Figure 5).

Supernatants of 293T and A549 cells transfected with anellosomes were harvested 96 hours post transfection. The harvested supernatants were spun down at 2000 rpm for 10 minutes at 4 degrees Celsius to remove any cell debris. Each of the harvested supernatants was used to infect new 293T and A549 cells, respectively, that were 70% confluent in wells of 24 well plates. Supernatants were washed away after 24 hours of incubation at 37 degrees Celsius and 5% carbon dioxide, followed by two washes of PBS, and replacement with fresh growth medium. Following incubation of these cells at 37 degrees and 5% carbon dioxide for another 48 hours, cells were individually harvested for genomic DNA extraction. Genomic DNA from each of the samples was harvested using a genomic DNA extraction kit (Thermo Fisher Scientific), according to manufacturer's protocol.

To confirm the successful infection of 293T and A549 cells by anellosomes produced in vitro, 100 ng of genomic DNA harvested as described herein was used to perform quantitative polymerase chain reaction (qPCR) using primers specific for beta-torqueviruses or LY2 specific sequences. SYBR green reagent (Thermo Fisher Scientific) was used to perform qPCR, as per manufacturer's protocol. qPCR for primers specific to genomic DNA sequence of GAPDH was used for normalization. The sequences for all the primers used are listed in Table 42.

Table 42:

	Primer sequence (5' > 3')	
Target	**Forward**	**Reverse**
Betatorqueviruses	ATTCGAATGGCTGAGTTTATGC (SEQ ID NO: 690)	CCTTGACTACGGTGGTTTCAC (SEQ ID NO: 693)
LY2 TTMiniV strain	CACGAATTAGCCAAGACTGGGCAC (SEQ ID NO: 691)	TGCAGGCATTCGAGGGCTTGTT (SEQ ID NO: 694)
GAPDH	GCTCCCACTCCTGATTTCTG (SEQ ID NO: 692)	TTTAACCCCCTAGTCCCAGG (SEQ ID NO: 695)

As shown in the qPCR results depicted in Figures 6A, 6B, 7A, and 7B, the anellosomes produced *in* vitro and as described in this example were infectious.

Example 7: Selectivity of anellosomes

This example demonstrates the ability of synthetic anellosomes produced *in vitro* to infect cell lines of a variety of tissue origins.

Supernatants with the infectious TTMiniV anellosomes (described in Example 5) were incubated with 70% confluent 293T, A549, Jurkat (an acute T cell leukemia cell line), Raji (a Burkitt's lymphoma B cell line), and Chang cell lines at 37 degrees and 5% carbon dioxide in wells of 24 well plates. Cells were washed with PBS twice, 24 hours post infection, followed by replacement with fresh growth medium. Cells were then incubated again at 37 degrees and 5% carbon dioxide for another 48 hours, followed by harvest for genomic DNA extraction. Genomic DNA from each of the samples was harvested using a genomic DNA extraction kit (Thermo Fisher Scientific), according to manufacturer's protocol.

To confirm successful infection of these cell lines by anellosomes produced in the previous Example, 100 ng of genomic DNA harvested as described herein was used to perform quantitative polymerase chain reaction (qPCR) using primers specific for beta-torqueviruses or LY2 specific sequences. SYBR green reagent (Thermo Fisher Scientific) was used to perform qPCR, as per manufacturer's protocol. qPCR for primers specific to genomic DNA sequence of GAPDH was used for normalization. The sequences for all the primers used are listed in Table 42.

As shown in the qPCR results depicted in Figures 6A-10B, not only were anellosomes produced *in vitro* infectious, they were able to infect a variety of cell lines, including examples of epithelial cells, lung tissue cells, liver cells, carcinoma cells, lymphocytes, lymphoblasts, T cells, B cells, and kidney cells. It was also observed that a synthetic anellosome was able to infect HepG2 cells (a liver cell line), resulting in a greater than 100-fold increase relative to a control.

Example 8: Identification and use of protein binding sequences

This example describes putative protein-binding sites in the *Anellovirus* genome, which can be used for amplifying and packaging effectors, e.g., in an anellosome as described herein. In some instances, the protein-binding sites may be capable of binding to an exterior protein, such as a capsid protein.

Two conserved domains within the *Anellovirus* genome are putative origins of replication: the 5' UTR conserved domain (5CD) and the GC-rich domain (GCR) (de Villiers et al., Journal of Virology 2011; Okamoto et al., Virology 1999). In one example, in order to confirm whether these sequences act as DNA replication sites or as capsid packaging signals, deletions of each region are made in plasmids harboring TTMV-LY2. A539 cells are transfected with pTTMV-LY2Δ5CD or pTTMV-LY2ΔGCR. Transfected cells are incubated for four days, and then virus is isolated from supernatant and cell pellets. A549 cells are infected with virus, and after four days, virus is isolated from the supernatant and infected cell pellets. qPCR is performed to quantify viral genomes from the samples. Disruption of an origin of replication prevents viral replicase from amplifying viral DNA and results in reduced viral genomes isolated from transfected cell pellets compared to wild-type virus. A small amount of virus is still packaged and can be found in the transfected supernatant and infected cell pellets. In some embodiments, disruption of a packaging signal will prevent the viral DNA from being encapsulated by capsid proteins. Therefore, in embodiments, there will still be an amplification of viral genomes in the transfected cells, but no viral genomes are found in the supernatant or infected cell pellets.

In a further example, in order to characterize additional replication or packaging signals in the DNA, a series of deletions across the entire TTMV-LY2 genome is used. Deletions of 100bp are made stepwise across the length of the sequence. Plasmids harboring TTMV-LY2 deletions are transfected into A549 and tested as described above. In some embodiments, deletions that disrupt viral amplification or packaging will contain potential *cis*-regulatory domains.

Replication and packaging signals can be incorporated into effector-encoding DNA sequences (e.g., in a genetic element in an anellosome) to induce amplification and encapsulation. This is done both in context of larger regions of the anellosome genome (i.e., inserting effectors into a specific site in the genome, or replacing viral ORFs with effectors, etc.), or by incorporating minimal *cis* signals into the effector DNA. In cases where the anellosome lacks *trans* replication or packaging factors (e.g., replicase and capsid proteins, etc.), the *trans* factors are supplied by helper genes. The helper genes express all of the proteins and RNAs sufficient to induce amplification and packaging, but lack their own packaging signals. The anellosome DNA is co-transfected with helper genes, resulting in amplification and packaging of the effector but not of the helper genes.

Example 9: An *Anellovirus* genome

This Example describes deletions in the *Anellovirus* genome.

A 172-nucleotide (nt) deletion was made in the non-coding region (NCR) of TTV-tth8 downstream of the ORFs but upstream of the GC-rich region (nts 3436 to 3607). A random 56-nt sequence (TTTGTGACACAAGATGGCCGACTTCCTTCCTCTTTAGTCTTCCCCAAAGAAGACAA (SEQ ID NO: 696)) was inserted into the deletion. pTTV-tth8(3436-3707::56nt), a DNA plasmid harboring the altered TTV-tth8, was generated. 2 µg of double-stranded circular plasmid or double-stranded SmaI linearized DNA (yielding a TTV-tth8 genome fragment separated from bacterial backbone elements) was transfected into HEK293 or A549 cells at 60% confluency in a 6cm plate using lipofectamine 2000, in duplicate. Virus was isolated from cell pellets and supernatant 96 hours post transfection by freeze thaw, alternating three times between liquid nitrogen and 37˚C water bath. Virus from supernatant was used to re-infect cells (HEK293 cells infected by virus isolated from HEK293, and A549 cells infected by virus isolated from A549). 72 hours after infection, virus was isolated from cell pellets and supernatant by freeze thaw. qPCR was performed on all samples. As shown in Table 43 below, TTV-tth8 was observed in both the cell pellet and supernatant of infected cells, indicating successful virus production by pTTV-tth8(3436-3707::56nt). Therefore, TTV-tth8 is able to tolerate deletion of nts 3436 to 3707.

Table 43: TTV-tth8(3436-3707::56nt) infections in HEK293 and A549 result in viral amplification. Average genome equivalents from duplicate experiments compared to negative control cells with no plasmid or virus added.

Genome Equivalents/Rx		HEK293 P0	HEK293 P1	A549 P0	A549 P1
TTH8 Linear	Sup	2.45E+06	1.02E+03	1.87E+07	1.00E+04
	Cell	2.52E+08	3.92E+05	2.89E+08	7.57E+05
TTH8 circular	Sup	1.69E+06	6.83E+02	5.07E+02	1.05E+04
	Cell	2.00E+08	3.75E+05	2.61E+08	8.36E+05

Negatives	
293 Empty	1.42E+02
293 Neg	5.08E+02
549 Empty	1.73E+01
549 Neg	2.08E+01

An engineered version of TTMV-LY2 was assembled, deleting nucleotides 574 to 1371 and 1432 to 2210 (1577 bp deletion) and inserting a 513 bp NanoLuc (nLuc) reporter ORF at the C-terminus of ORF1 (after nt 2609 in wild-type TTMV-LY2). Plasmids harboring the DNA sequence for the engineered TTMV-LY2 (pVL46-015B) were transfected into A549 cells, and then virus was isolated and used to infect new A549 cells, as described in Example 17. nLuc luminescence was detected in the cell pellets

and supernatant of the infected cells, indicating viral replication (Figures 11A-11B). This demonstrates that TTMV-LY2 can tolerate at least a 1577 bp deletion in the ORF region.

To further characterize the viral genome, a series of deletions are made in the TTMV-LY2 DNA. A TTMV-LY2 with deletions of nts 574-1371 and 1432-2210 but no nLuc insertion is made and tested for viral replication as described previously. Further deletions are made to TTMV-LY2Δ574-1371,Δ1432-2210. Nts 1372-1431 are deleted to create TTMV-LY2Δ574-2210. Additionally, ORF3 sequence downstream of ORF1 is deleted (Δ2610-2809). Finally, to test deletions in non-coding regions, a series of 100 bp deletions are made sequentially across the NCR. All deletion mutants are tested for viral replication as previously described. Deletions that result in successful viral production (indicating that the deleted region is not essential for viral replication) are combined to make variants of TTMV-LY2 with more deleted nucleotides. To identify the viral genome that can be amplified with helpers, each of the deletion mutants that disrupted viral replication is tested alongside helper genes carrying *trans* replication and packaging elements. Deletions rescued by *trans* expression of replication elements indicate areas of the viral genome that can be deleted without blocking virus formation when helper genes are provided from a separate source.

Example 10: Nucleotide insertions of various lengths into an *Anellovirus* genome

This example describes the addition of DNA sequences of various lengths into an *Anellovirus* genome, which can, in some instances, be used to generate an anellosome as described herein.

DNA sequences are cloned into plasmids harboring TTV-tth8 (GenBank accession number AJ620231.1) and TTMV-LY2 (GenBank accession number JX134045.1). Insertions are made in the noncoding regions (NCR) 3' of the open reading frames and 5' of the GC-rich region: after nucleotide 3588 in TTV-tth8, or nucleotide 2843 in TTMV-LY2.

Randomized DNA sequences of the following lengths are inserted into the NCRs of TTV-tth8 and TTMV-LY2: 100 base pairs (bp), 200 bp, 500 bp, 1000 bp, and 2000 bp. These sequences are designed to match the relative GC-content of each viral genome: approximately 50% GC for insertions into TTV-tth8, and approximately 38% GC for TTMV-LY2. In addition, several *trans* genes are inserted into the NCR. These include a miRNA (e.g., FF4 miRNA) driven by a U6 promoter (351 bp) and EGFP driven by a constitutive hEF1a promoter (2509 bp).

TTV-tth8 and TTMV-LY2 variants harboring various sized DNA inserts are transfected into mammalian cell lines, including HEK293 and A549, as previously described. Virus is isolated from the supernatant or cell pellets. Isolated virus is used to infect additional cells. Production of virus from the infected cells is monitored by quantitative PCR. In some embodiments, successful production of virus will indicate tolerance of insertions.

Example 11: Exemplary cargo to be delivered

This example describes exemplary classes of nucleic acid and protein payloads that may be delivered with an anellosome, e.g., an anellosome based on an *Anellovirus*, e.g., as described herein.

One example of a payload is mRNA for protein expression. A coding sequence of interest is transcribed from either a viral promoter native to the source virus (e.g., an *Anellovirus*) or from a promoter introduced with the payload as part of a *trans* gene. Alternatively, the mRNA is encoded within the open reading frames of the viral mRNAs, resulting in fusions between viral proteins and the protein of interest. Cleavage domains, for example, the 2A peptide or a proteinase target site, may be used to separate the protein of interest from the viral proteins when desired.

Non-coding RNAs (ncRNAs) are another example of a payload. These RNAs are generally transcribed using RNA polymerase III promoters, such as U6 or VA. Alternatively, an ncRNA is transcribed using RNA polymerase II, such as the native viral promoter or regulatable synthetic promoters. When expressed from RNA polymerase II promoters, the ncRNAs are encoded as part of the mRNA exon, introns, or as extra RNA transcribed downstream of the poly-A signal. ncRNAs are often encoded as part of a larger RNA molecule or are cleaved apart using ribozymes or endoribonucleases. ncRNAs that can be encoded as cargo in the genome of an anellosome include micro-RNA (miRNA), small-interfering RNAs (siRNA), short hairpin RNA (shRNA), antisense RNA, miRNA sponges, long-noncoding RNA (lncRNA), and guide RNA (gRNA).

DNA may be used as a functional element without requiring RNA transcription. For example, DNA may be used as a template for homologous recombination. In another example, a protein-binding DNA sequence may be used to drive packaging of proteins of interest into a capsid (e.g., in a proteinaceous exterior of an anellosome). For homologous recombination, regions of homology to human genomic DNA are encoded into the vector DNA to act as homology arms. Recombination can be driven by a targeted endonuclease (such as Cas9 with a gRNA, or a zinc-finger nuclease), which can be expressed either from the vector or from a separate source. Inside the cell, a single-stranded DNA genome is converted to double-stranded DNA, which then acts as a template for homologous recombination at the genomic DNA break site. For recruiting proteins of interest, a protein-binding sequence can be encoded in the anellosome DNA. A DNA-binding protein of interest, or a protein of interest fused to a DNA-binding protein (such as Gal4), binds to the anellosome DNA. When the anellosome DNA is encapsulated by the capsid proteins, the DNA-binding protein is encapsulated too, and can be delivered to cells with the anellosome.

Example 12: Exemplary payload integration loci

This example describes exemplary loci in the genomes of TTV-tth8 (GenBank accession number AJ620231.1) and TTMV-LY2 (GenBank accession number JX134045) into which nucleic acid payloads can be inserted.

Several strategies can be employed for insertions into the open reading frame (ORF) regions of TTV-tth8 (nucleotides 336 to 3015) and TTMV-LY2 (nucleotides 424 to 2812). In one example, in order to tag viral proteins or create fusion proteins, a payload is inserted in frame within the specific ORF of interest. Alternatively, part or all of the ORF region is deleted, which may or may not disrupt viral protein function. The payload is then inserted into the deleted region. Additionally, a hyper-variable domain (HVD) in ORF1 of TTV-tth8 (between nucleotides 716 and 2362) or TTMV-LY2 (between nucleotides 724 and 2273) can be used as an insertion site. In some instances, insertions or nucleotide replacements in the HVD may be better tolerated and/or disrupt viral function to a lesser degree.

Alternatively, payload insertions are made into regions of the vector comparable to the non-coding regions (NCRs) of TTV-tth8 or TTMV-LY2. In particular, insertions are made in the 5' NCR upstream of the TATA box, in the 5' untranslated region (UTR), in the 3' NCR downstream of the poly-A signal and upstream of the GC-rich region. Additionally, insertions are made into the miRNA region of TTV-tth8 (nucleotides 3429 to 3506). For the 5' NCR region, insertions are made upstream of the TATA box (between nucleotides 1 and 82 in TTV-tth8, and nucleotides 1 and 236 in TTMV-LY2). In some embodiments, *trans* genes are inserted in the reverse orientation to reduce promoter interference. For the 5' UTR, insertions are made downstream of the transcriptional start site (nucleotide 111 in TTV-tth8, and nucleotide 267 in TTMV-LY2) and upstream of the ORF2 start codon (nucleotide 336 in TTV-tth8, and nucleotide 421 in TTMV-LY2). 5' UTR insertions add or replace nucleotides in the 5' UTR. 3' NCR insertions are made upstream of the GC-rich region, in particular after nucleotide 3588 in TTV-tth8 or nucleotide 2843 in TTMV-LY2, as described in Example 10. The miRNA of TTV-tth8 is replaced by alternative natural or synthetic miRNA hairpins.

Example 13: Defined categories of *Anellovirus* and conserved regions thereof

There are three genera of *Anellovirus* present in humans: *Alphatorquevirus* (Torque Teno Virus, TTV), *Betatorquevirus* (Torque Teno Midi Virus, TTMDV), and *Gammatorquevirus* (Torque Teno Mini Virus, TTMV). *Alphatorquevirus* includes at least five (e.g., seven) well-supported phylogenetic clades (Figure 11C). It is contemplated that any of these *Anelloviruses* can be used as a source virus (e.g., a source of viral DNA sequences) for producing an anellosome as described herein.

Among these sequences, the highest conservation is found in the 5' UTR domain (about 75% conserved) and the GC-rich domain (greater than 100 base pairs, greater than 70% GC-content, about

70% conserved). Additional, a hypervariable domain (HVD) in the sequences has very low conservation (about 30% conserved). All *Anelloviruses* also contain a region in which all three reading frames are open. In some instances, the 5'UTR or the GC-rich region may function as an origin of replication.

Also provided herein are exemplary sequences of representative viruses from each of the TTV clades, and of TTMDV and TTMV, annotated with the conserved regions (see, e.g., Tables A1-A12, B1-B5, C1-C5, or 1-18).

Example 14: Replication-deficient anellosomes and helper viruses

For replication and packaging of an anellosome, some elements can be provided in *trans*. These include proteins or non-coding RNAs that direct or support DNA replication or packaging. *Trans* elements can, in some instances, be provided from a source alternative to the anellosome, such as a helper virus, plasmid, or from the cellular genome.

Other elements are typically provided in *cis*. These elements can be, for example, sequences or structures in the anellosome DNA that act as origins of replication (e.g., to allow amplification of anellosome DNA) or packaging signals (e.g., to bind to proteins to load the genome into the capsid). Generally, a replication deficient virus or anellosome will be missing one or more of these elements, such that the DNA is unable to be packaged into an infectious virion or anellosome even if other elements are provided in *trans*.

Replication deficient viruses can be useful as helper viruses, e.g., for controlling replication of an anellosome (e.g., a replication-deficient or packaging-deficient anellosome) in the same cell. In some instances, the helper virus will lack *cis* replication or packaging elements, but express *trans* elements such as proteins and non-coding RNAs. Generally, the therapeutic anellosome would lack some or all of these *trans* elements and would therefore be unable to replicate on its own, but would retain the *cis* elements. When co-transfected/infected into cells, the replication-deficient helper virus would drive the amplification and packaging of the anellosome. The packaged particles collected would thus be comprised solely of therapeutic anellosome, without helper virus contamination.

To develop a replication deficient anellosome, conserved elements in the non-coding regions of *Anellovirus* will be removed. In particular, deletions of the conserved 5' UTR domain and the GC-rich domain will be tested, both separately and together. Both elements are contemplated to be important for viral replication or packaging. Additionally, deletion series will be performed across the entire non-coding region to identify previously unknown regions of interest.

Successful deletion of a replication element will result in reduction of anellosome DNA amplification within the cell, e.g., as measured by qPCR, but will support some infectious anellosome production, e.g., as monitored by assays on infected cells that can include any or all of qPCR, western

blots, fluorescence assays, or luminescence assays. Successful deletion of a packaging element will not disrupt anellosome DNA amplification, so an increase in anellosome DNA will be observed in transfected cells by qPCR. However, the anellosome genomes will not be encapsulated, so no infectious anellosome production will be observed.

Example 15: Manufacturing process for replication-competent anellosomes

This example describes a method for recovery and scaling up of production of replication-competent anellosomes. Anellosomes are replication competent when they encode in their genome all the required genetic elements and ORFs necessary to replicate in cells. Since these anellosomes are not defective in their replication they do not need a complementing activity provided in trans. They might, however need helper activity, such as enhancers of transcriptions (e.g. sodium butyrate) or viral transcription factors (e.g. adenoviral E1, E2 E4, VA; HSV Vp16 and immediate early proteins).

In this example, double-stranded DNA encoding the full sequence of a synthetic anellosome either in its linear or circular form is introduced into 5E+05 adherent mammalian cells in a T75 flask by chemical transfection or into 5E+05 cells in suspension by electroporation. After an optimal period of time (e.g., 3-7 days post transfection), cells and supernatant are collected by scraping cells into the supernatant medium. A mild detergent, such as a biliary salt, is added to a final concentration of 0.5% and incubated at 37°C for 30 minutes. Calcium and Magnesium Chloride is added to a final concentration of 0.5mM and 2.5mM, respectively. Endonuclease (e.g. DNAse I, Benzonase), is added and incubated at 25-37°C for 0.5-4 hours. Anellosome suspension is centrifuged at 1000 x g for 10 minutes at 4°C. The clarified supernatant is transferred to a new tube and diluted 1:1 with a cryoprotectant buffer (also known as stabilization buffer) and stored at -80°C if desired. This produces passage 0 of the anellosome (P0). To bring the concentration of detergent below the safe limit to be used on cultured cells, this inoculum is diluted at least 100-fold or more in serum-free media (SFM) depending on the anellosome titer.

A fresh monolayer of mammalian cells in a T225 flask is overlaid with the minimum volume sufficient to cover the culture surface and incubated for 90 minutes at 37°C and 5% carbon dioxide with gentle rocking. The mammalian cells used for this step may or may not be the same type of cells as used for the P0 recovery. After this incubation, the inoculum is replaced with 40ml of serum-free, animal origin-free culture medium. Cells are incubated at 37°C and 5% carbon dioxide for 3-7 days. 4 ml of a 10X solution of the same mild detergent previously utilized is added to achieve a final detergent concentration of 0.5%, and the mixture is then incubated at 37°C for 30 minutes with gentle agitation. Endonuclease is added and incubated at 25-37°C for 0.5-4 hours. The medium is then collected and

centrifuged at 1000 x g at 4°C for 10 minutes. The clarified supernatant is mixed with 40 ml of stabilization buffer and stored at-80°C. This generates a seed stock, or passage 1 of anellosome (P1).

Depending on the titer of the stock, it is diluted no less than 100-fold in SFM and added to cells grown on multilayer flasks of the required size. Multiplicity of infection (MOI) and time of incubation is optimized at smaller scale to ensure maximal anellosome production. After harvest, anellosomes may then be purified and concentrated as needed. A schematic showing a workflow, e.g., as described in this example, is provided in Figure 12.

Example 16: Manufacturing process of replication-deficient anellosomes

This example describes a method for recovery and scaling up of production of replication-deficient anellosomes.

Anellosomes can be rendered replication-deficient by deletion of one or more ORFs (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3) involved in replication. Replication-deficient anellosomes can be grown in a complementing cell line. Such cell line constitutively expresses components that promote anellosome growth but that are missing or nonfunctional in the genome of the anellosome.

In one example, the sequence(s) of any ORF(s) involved in anellosome propagation are cloned into a lentiviral expression system suitable for the generation of stable cell lines that encode a selection marker, and lentiviral vector is generated as described herein. A mammalian cell line capable of supporting anellosome propagation is infected with this lentiviral vector and subjected to selective pressure by the selection marker (e.g., puromycin or any other antibiotic) to select for cell populations that have stably integrated the cloned ORFs. Once this cell line is characterized and certified to complement the defect in the engineered anellosome, and hence to support growth and propagation of such anellosomes, it is expanded and banked in cryogenic storage. During expansion and maintenance of these cells, the selection antibiotic is added to the culture medium to maintain the selective pressure. Once anellosomes are introduced into these cells, the selection antibiotic may be withheld.

Once this cell line is established, growth and production of replication-deficient anellosomes is carried out, e.g., as described in Example 15.

Example 17: Production of anellosomes using suspension cells

This example describes the production of anellosomes in cells in suspension.

In this example, an A549 or 293T producer cell line that is adapted to grow in suspension conditions is grown in animal component-free and antibiotic-free suspension medium (Thermo Fisher Scientific) in WAVE bioreactor bags at 37 degrees and 5% carbon dioxide. These cells, seeded at 1 x 10^6

viable cells/ mL, are transfected using lipofectamine 2000 (Thermo Fisher Scientific) under current good manufacturing practices (cGMP), with a plasmid comprising anellosome sequences, along with any complementing plasmids suitable or required to package the anellosome (e.g., in the case of a replication-deficient anellosome, e.g., as described in Example 16). The complementing plasmids can, in some instances, encode for viral proteins that have been deleted from the anellosome genome (e.g., an anellosome genome based on a viral genoe, e.g., an *Anellovirus* genome, e.g., as described herein) but are useful or required for replication and packaging of the anellosomes. Transfected cells are grown in the WAVE bioreactor bags and the supernatant is harvested at the following time points: 48, 72, and 96 hours post transfection. The supernatant is separated from the cell pellets for each sample using centrifugation. The packaged anellosome particles are then purified from the harvested supernatant and the lysed cell pellets using ion exchange chromatography.

The genome equivalents in the purified prep of the anellosomes can be determined, for example, by using a small aliquot of the purified prep to harvest the anellosome genome using a viral genome extraction kit (Qiagen), followed by qPCR using primers and probes targeted towards the anellosome DNA sequence, e.g., as described in Example 18.

The infectivity of the anellosomes in the purified prep can be quantified by making serial dilutions of the purified prep to infect new A549 cells. These cells are harvested 72 hours post transfection, followed by a qPCR assay on the genomic DNA using primers and probes that are specific to the anellosome DNA sequence.

Example 18: Quantification of anellosome genome equivalents by qPCR

This example demonstrates the development of a hydrolysis probe-based quantitative PCR assay to quantify anellosomes. Sets of primers and probes were designed based on selected genome sequences of TTV (Accession No. AJ620231.1) and TTMV (Accession No. JX134045.1) using the software Geneious with a final user optimization. Primer sequences are shown in Table 44 below.

Table 44: Sequences of forward and reverse primers and hydrolysis probes used to quantify TTMV and TTV genome equivalents by quantitative PCR.

TTMV		SEQ ID NO:
Forward Primer	5'-GAAGCCCACCAAAAGCAATT-3'	697
Reverse Primer	5'-AGTTCCCGTGTCTATAGTCGA-3'	698
Probe	5'-ACTTCGTTACAGAGTCCAGGGG-3'	699

TTV		
Forward Primer	5'-AGCAACAGGTAATGGAGGAC-3'	700
Reverse Primer	5'-TGGAAGCTGGGGTCTTTAAC-3'	701
Probe	5'-TCTACCTTAGGTGCAAAGGGCC-3'	702

As a first step in the development process, qPCR is run using the TTV and TTMV primers with SYBR-green chemistry to check for primer specificity. Figure 13 shows one distinct amplification peak for each primer pair.

Hydrolysis probes were ordered labeled with the fluorophore 6FAM at the 5' end and a minor groove binding, non-fluorescent quencher (MGBNFQ) at the 3' end. The PCR efficiency of the new primers and probes was then evaluated using two different commercial master mixes using purified plasmid DNA as component of a standard curve and increasing concentrations of primers. The standard curve was set up by using purified plasmids containing the target sequences for the different sets of primers-probes. Seven tenfold serial dilutions were performed to achieve a linear range over 7 logs and a lower limit of quantification of 15 copies per 20ul reaction. Master mix #2 was capable of generating a PCR efficiency between 90-110%, values that are acceptable for quantitative PCR (Figure 14). All primers for qPCR were ordered from IDT. Hydrolysis probes conjugated to the fluorophore 6FAM and a minor groove binding, non-fluorescent quencher (MGBNFQ) as well as all the qPCR master mixes were obtained from Thermo Fisher. An exemplary amplification plot is shown in Figure 15.

Using these primer-probe sets and reagents, the genome equivalent (GEq)/ml in anellosome stocks was quantified. The linear range was between 1.5E+07 – 15 GEq per 20ul reaction, which was then used to calculate the GEq/ml, as shown in Figures 16A-16B. Samples with higher concentrations than the linear range can be diluted as needed.

Example 19: Utilizing anellosomes to express an exogenous protein in mice

This example describes the usage of an anellosome in which the Torque Teno Mini Virus (TTMV) genome is engineered to express the firefly luciferase protein in mice.

The plasmid encoding the DNA sequence of the engineered TTMV encoding the firefly-luciferase gene is introduced into A549 cells (human lung carcinoma cell line) by chemical transfection. 18 ug of plasmid DNA is used for transfection of 70% confluent cells in a 10 cm tissue culture plate. Empty vector backbone lacking the TTMV sequences is used as a negative control. Five hours post-transfection, cells are washed with PBS twice and are allowed to grow in fresh growth medium at 37°C and 5% carbon dioxide.

Transfected A549 cells, along with their supernatant, are harvested 96 hours post transfection. Harvested material is treated with 0.5% deoxycholate (weight in volume) at 37°C for 1 hour followed by endonuclease treatment. Anellosome particles are purified from this lysate using ion exchange chromatography. To determine anellosome concentration, a sample of the anellosome stock is run through a viral DNA purification kit and genome equivalents per ml are measured by qPCR using primers and probes targeted towards the anellosome DNA sequence.

A dose-range of genome equivalents of anellosomes in 1x phosphate-buffered saline is performed via a variety of routes of injection (e.g. intravenous, intraperitoneal, subcutaneous, intramuscular) in mice at 8-10 weeks of age. Ventral and dorsal bioluminescence imaging is performed on each animal at 3, 7, 10 and 15 days post injection. Imaging is performed by adding the luciferase substrate (Perkin-Elmer) to each animal intraperitoneally at indicated time points, according to the manufacturer's protocol, followed by intravital imaging.

Example 20: Genome alignments to determine whether anellosome DNA integrated into host genomes

This example describes the computational analysis performed to determine whether anellosome DNA can integrate into the host genome, by examining whether Torque Teno Virus (TTV) has integrated into the human genome.

The complete genomes of one representative TTV sequence from each of five exemplary *Alphatorquevirus* clades were aligned against the human genome sequence using the Basic Local Alignment Search Tool (BLAST) that finds regions of local similarity between sequences. The representative TTV sequences shown in Table 45 were analyzed:

Table 45: Representative TTV sequences

TTV Clade	NCBI Accession No.
Clade A	AB064597.1
Clade B	AB028669.1
Clade C	AJ20231.1
Clade D	AF122914.3
Clade E	AF298585.1

Sequences from none of the aligned TTVs were found to have any significant similarity to the human genome, indicating that the TTVs have not integrated into the human genome.

Example 21: Assessment of anellosome integration into a host genome

In this example, A549 cells (human lung carcinoma cell line) and HEK293T cells (human embryonic kidney cell line) are infected with either anellosome particles or AAV particles at MOIs of 5, 10, 30 or 50. The cells are washed with PBS 5 hours post infection and replaced with fresh growth medium. The cells are then allowed to grow at 37 degrees and 5% carbon dioxide. Cells are harvested five days post infection and they are processed to harvest genomic DNA, using the genomic DNA extraction kit (Qiagen). Genomic DNA is also harvested from uninfected cells (negative control). Whole-genome sequencing libraries are prepared for these harvested DNAs, using the Nextera DNA library preparation kit (Illumina), according to manufacturers protocol. The DNA libraries are sequenced using the NextSeq 550 system (Illumina) according to manufacturer's protocol. Sequencing data is assembled to the reference genome and analyzed to look for junctions between anellosome or AAV genomes and host genome. In cases where junctions are detected they are verified in the original genomic DNA sample prior sequencing library preparation by PCR. Primers are designed to amplify the region containing and around the junctions. The frequency of integration of anellosomes into the host genome is determined by quantifying the number of junctions (representing integration events) and the total number of anellosome copies in the sample by qPCR. This ratio can be compared to that of AAV.

Example 22: Functional effects of an anellosome expressing an exogenous microRNA sequence

This example demonstrates the successful expression of an exogenous miRNA (miR-625) from anellosome genome using a native promoter.

500 ng of following plasmid DNAs were transfected into 60% confluent wells of HEK293T cells in a 24 well plate:

i) Empty plasmid backbone

ii) Plasmid containing TTV-tth8 genome in which endogenous miRNA is knocked out (KO)

iii) TTV-tth8 in which endogenous miRNA is replaced with a non-targeting scramble miRNA

iv) TTV-tth8 in which endogenous miRNA sequence is replaced with miRNA encoding miR-625

72 hours post transfection, total miRNA was harvested from the transfected cells using the Qiagen miRNeasy kit, followed by reverse transcription using miRNA Script RT II kit. Quantitative PCR was performed on the reverse transcribed DNA using primer that should specifically detect miRNA-625 or RNU6 small RNA. RNU6 small RNA was used as a housekeeping gene and data is plotted in Figure 17 as a fold change relative to empty vector. As shown in Figure 17, miR-625 anellosome resulted in approximately 100-fold increase in miR-625 expression, whereas no signal was detected for empty vector, miR-knockout (KO), and scrambled miR.

Example 23: Preparation and production of anellosomes to express exogenous non-coding RNAs

This example describes the synthesis and production of anellosomes to express exogenous small non-coding RNAs.

The DNA sequence from the tth8 strain of TTV (Jelcic et al, *Journal of Virology*, 2004) is synthesized and cloned into a vector containing the bacterial origin of replication and bacterial antibiotic resistance gene. In this vector, the DNA sequence encoding the TTV miRNA hairpin is replaced by a DNA sequence encoding an exogenous small non-coding RNA such as miRNA or shRNA. The engineered construct is then transformed into electro-competent bacteria, followed by plasmid isolation using a plasmid purification kit according to the manufacturer's protocols.

The anellosome DNA encoding the exogenous small non-coding RNAs is transfected into an eukaryotic producer cell line to produce anellosome particles. The supernatant of the transfected cells containing the anellosome particles is harvested at different time points post transfection. Anellosome particles, either from the filtered supernatant or after purification, are used for downstream applications, e.g., as described herein.

Example 24: Conservation in *Anellovirus* clades

This example describes the identification of seven clades within the *Alphatorquevirus* genus. Representative sequences between these clades showed 54.7% pairwise identity across the sequences (Figure 18). The pairwise identity was lowest among the open reading frames (~48.8%), and higher in the non-coding regions (69.1% in the 5' NCR, 74.6% in the 3' NCR) (Figure 18). This suggests that DNA sequences or structures in the non-coding regions play important roles in viral replication.

The amino acid sequences of the putative proteins in Alphatorquevirus were also compared. The DNA sequences showed approximately 47-50% pairwise identity, while the amino acid sequences showed approximately 32-38% pairwise identity (Figure 19). Interestingly, the representative sequences from the Alphatorquevirus clades are able to successfully replicate *in vivo* and are observed in the human population. This suggests that the amino acid sequences for Anellovirus proteins can vary widely while retaining functionalities such as replication and packaging.

Anelloviruses were found to have regions of local high conservation in the non-coding regions. In the region downstream of the promoter is a 71-bp 5' UTR conserved domain that exhibited 95.2% pairwise identity across the seven alphatorquevirus clades (Figure 20). Downstream of the open reading frames in the 3' non-coding region of alphatorqueviruses, there was a region with substantial pairwise identity between the representative sequences. Near the 3' end of this 3' conserved non-coding region is a highly conserved sequence. The *Anelloviruses* also included a GC-rich region having greater than 70% GC content, which exhibited 75.4% pairwise identity in areas where they align (Figure 21).

Example 25: Expression of an endogenous miRNA from an anellosome and deletion of the endogenous miRNA

In one example, anellosomes comprising a modified TTV-tth8 genome, in which the TTV-tth8 genome was modified with a deletion in the GC-rich region as described in Example 27, were used to infect Raji B cells in culture. These anellosomes comprised a sequence encoding the endogenous payload of the TTV-tth8 *Anellovirus*, which is a miRNA targeting the mRNA encoding n-myc interacting protein (NMI), and were produced by introducing a plasmid comprising the *Anellovirus* genome into a host cell. NMI operates downstream of the JAK/STAT pathway to regulate the transcription of various intracellular signals, including interferon-stimulated genes, proliferation and growth genes, and mediators of the inflammatory response. As shown in Figure 22, viral genomes were detected in target Raji B cells. Successful knockdown of NMI was also observed in target Raji B cells compared to control cells (Figure 23). Anellosome comprising the miRNA against NMI induced a greater than 75% reduction in NMI protein levels compared to control cells. This example demonstrates that an anellosome with a native *Anellovirus* miRNA can knock down a target molecule in host cells.

In another example, the endogenous miRNA of an *Anellovirus*-based anellosome was deleted. The resultant anellosome (Δ miR) was then incubated with host cells. Genome equivalents of Δ miR anellosome genetic elements was then compared to that of corresponding anellosomes in which the endogenous miRNA was retained. As shown in Figure 24, anellosome genomes in which the endogenous miRNA were deleted were detected in cells at levels comparable to those observed for anellosome genomes in which the endogenous miRNA was still present. This example demonstrates that the endogenous miRNA of an *Anellovirus*-based anellosome can be mutated, or deleted entirely and the anellosome genome can still be detected in target cells.

Example 26: Localization of *Anellovirus* ORFs

This Example describes novel functionality of various putative ORFs of *Anelloviruses*. In this example, putative open reading frame (ORF) sequences were designed downstream of a tagged protein (i.e. nanoLuciferase) at the N-terminus of each ORF. Each ORF-nLuc plasmid was introduced into 5E+05 adherent cells (Vero or HEK293T) in a 12-well plate by chemical transfection or into 5E+05 cells in suspension by electroporation. After an optimal period of time (e.g., 3-7 days post transfection), cells were fixed with 4% paraformaldehyde (ThermoFisher cat# 28908) in PBS and permeabilized with 0.5% Triton X-100 and stained for nLuc with a rabbit polyclonal anti nLuc antibody (kind gift of Promega Corp.) followed by goat anti-rabbit Alexa488 (ThermoFisher cat# A-11008) conjugated secondary antibody. The nuclei were stained with DAPI (ThermoFisher Cat# D3571). The stained cells were

visualized on a Zeiss AxioVert A1 with a 20X objective and a monochrome Axiocam 506 camera for tagged protein cellular localization.

As shown in Figures 25A-25B, ORF2 was observed localized the cytoplasm and ORF1/1 was observed localized to the nucleus in both Vero cells and HEK293 cells. Figure 25C shows the localization for ORF1/2 and ORF2/2.

Example 27: Characterization of regions required for anellosome development

This Example describes deletions in the Anellovirus genome to help characterize the portions of the genome sufficient for replicating virus and anellosome production. A series of deletions were made in the non-coding region (NCR) of TTV-tth8 downstream of the ORFs (nts 3016 to 3753). A 36-nucleotide (nt) sequence (CGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGC (SEQ ID NO: 160)) was deleted from the GC region (labeled Δ36nt (GC)). Additionally, a 78-nt pre-microRNA sequence (CCGCCATCTTAAGTAGTTGAGGCGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACA CCTACTCAAAATGGTGG (SEQ ID NO: 161)) was deleted from the 3' NCR (labeled Δ36nt (GC) ΔmiR). And lastly, an extra 171 nts in the 3'NCR of Δ36nt (GC) was deleted (CTTAAGTAGTTGAGGCGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACACCTACTC AAAATGGTGGACAATTTCTTCCGGGTCAAAGGTTACAGCCGCCATGTTAAAACACGTGACGT ATGACGTCACGGCCGCCATTTTGTGACACAAGATGGCCGACTTCCTTCC (SEQ ID NO: 162)) and labeled Δ3'NCR (Figure 26). 2 μg of circular pTTV-tth8 (WT), pTTV-tth8(Δ36nt (GC)), pTTV-tth8(Δ36nt (GC) ΔmiR), pTTV-tth8(Δ3'NCR) DNA plasmids harboring the altered 3'NCRs TTV-tth8 respectively described above, were transfected into HEK293 at 60% confluency in a 12-well plate using lipofectamine 2000, in triplicates. 48 hours after transfection, cell pellets were harvested and lysed to isolate mRNA transcripts (RNeasy, Qiagen cat# 74104) and converted to cDNA (High-Capacity cDNA Reverse Transcription kit, ThermoFisher, cat# 4368814). qPCR was performed on all samples measuring viral transcripts expression with each deletion and normalized to the internal control mRNA of GAPDH.

As shown in Figures 27A-27D, all three of the deletion mutants significantly inhibited viral transcript expression *in vitro*. Therefore, the 3' NCR of TTV-tth8 is necessary for anellosome production for expression of transgene.

The TTV strain tth8, GeneBank Accession No. AJ620231.1, was deposited as a full-genome sequence. In the GC-rich region, however, there is a stretch of 36 nucleotides annotated as generic Ns. This region is highly conserved among TTV strains and therefore might be important for the biology of these viruses. The DNA sequences of several hundred TTV strains were computationally aligned and used to generate a strong consensus sequence for those 36 nucleotides (CGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGC (SEQ ID NO: 160)). The TTV-tth8

genome sequence referred to herein as the "wild-type" sequence accordingly had this consensus sequence inserted in place of the stretch of 36 Ns listed in the publicly available TTV-tth8 sequence.

Example 28: Anellosome delivery of exogenous proteins *in vivo*

This example demonstrates in vivo effector function (e.g. expression of proteins) of anellosomes after administration.

Anellosomes comprising a transgene encoding nano-luciferase (nLuc) (Figures 28A-28B) were prepared. Briefly, double-stranded DNA plasmids harboring the TTMV-LY2 non-coding regions and an nLuc expression cassette were transfected into HEK293T cells along with double-stranded DNA plasmids encoding the full TTMV-LY2 genome to act as *trans* replication and packaging factors. After transfection, cells were incubated to permit anellosome production and anellosome material was harvested and enriched via nuclease treatment, ultrafiltration/diafiltration, and sterile filtration. Additional HEK293T cells were transfected with non-replicating DNA plasmids harboring nLuc expression cassettes and TTMV-LY2 ORF transfection cassettes, but lacking non-coding domains essential for replication and packaging, to act as a "non-viral" negative control. The non-viral samples were prepared following the same protocol as the anellosome material.

Anellosome preparation was administered to a cohort of three healthy mice intramuscularly, and monitored by IVIS Lumina imaging (Bruker) over the course of nine days (Figure 29A). As a non-viral control, the non-replicating preparation was administered to three additional mice (Figure 29B). Injections of 25µL of anellosome or non-viral preparations were administered to the left hind leg on Day 0, and re-administered to the right hind leg on Day 4 (See arrows in Figure 29A and B). After 9 days of IVIS imaging, more occurences of nLuc luminescent signal were observed in mice injected with the anellosome preparation (Figure 29A) than the non-viral preparation (Figure 29B), which is consistent with *trans* gene expression after *in vivo* anellosome transduction.

Example 29: Identification of precursor miRNAs (pre-mIRs) in Anelloviruses

This example describes various computational and experimental approaches to identify novel precursor miRNAs encoded by various Anelloviruses.

Computational Methods

Anellovirus strains are very diverse from each other at the level of nucleotide sequence. However, Anellovirus strains, especially the ones within the same clade, can show significant similarity to each other in terms of genomic organization of various components such as promoter, GC rich region, non-coding region, and coding regions (see, e.g., Figure 29D). Herein is described a method in which the pre-

miR sequences of various Anellovirus strains (whose pre-miR sequences are unknown) are predicted by aligning with Anellovirus strains whose pre-miR sequences are already experimentally validated.

Briefly, various publicly available small RNA sequencing data sets for small RNAs from cell lines and various human samples are mined to discover novel pre-miR sequences encoded by various strains of Anelloviruses. Publicly available computational tools and algorithms that are based on structure prediction or machine-learning classification, such as the mFold program, miRANDA algorithm, miRScan, miRanalyzer, miRDeep (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1559940/, https://www.frontiersin.org/articles/10.3389/fbioe.2015.00007/full) are used to predict novel miRNAs encoded by various anellos. Northern blots with probes designed to specific miRNA sequences and/ or RT-qPCR using primers specific to miRNAs are then used to confirm, validate and quantitate the expression of novel miRNAs.

Experimental Methods

In one example, high throughput small RNA sequencing is performed on human tissue or blood samples that are infected with anellos to discover novel Anellovirus-encoded pre-miRNAs. To perform this, RNA is harvested from homogenized human tissue samples or human blood samples. Small RNA libraries are prepared and sequenced using Illumina kits and sequencing platforms. Sequencing reads are stored, aligned, and analyzed on BaseSpace Sequence Hub (Illumina).

In a second example, high throughput small RNA sequencing is performed on various cell lines treated with the following conditions to discover novel pre-miRNAs encoded by Anelloviruses: (a) cell lines infected with naturally occurring Anelloviruses, cell lines transfected with Anellovirus genomes synthesized *in vitro*, and (c) cell lines infected with Anelloviruses packaged *in vitro* using synthetic genomes. Northern blots with probes designed to specific miRNA sequences and/ or RT-qPCR using primers specific to miRNAs are used to confirm, validate and quantitate the expression of novel miRNAs.

Example 30: Determination of the endogenous target of Anellovirus pre-miRs

This example describes the analysis to determine endogenous targets and potentially therapeutically relevant target pathways of pre-miRs encoded by various strains of Anelloviruses. Computationally predicted and/or experimentally validated individual pre-miRNA sequences encoded by various Anelloviruses are cloned into a lentiviral vector, driven by a U6 promoter. A non-targeting scramble miRNA sequence, driven by a U6 promoter is also cloned in a similar way that is used as a control. The lentiviral plasmid is cloned such that when packaged, its genome will contain (i) a pre-miRNA sequence driven by a U6 promoter, (ii) a puromycin resistance gene driven by a SV40 promoter, and (iii) a Green Fluorescent Protein (GFP) gene driven by a CMV promoter. Each of these lentiviral

plasmids are individually co-transfected into HEK-293T cells along with the lentiviral helper plasmids to package the virus. Six hours after transfection, the medium of the transfected cells is aspirated, washed with PBS once and replaced with fresh medium. This medium containing the lentivirus is harvested 72 hours post transfection. The medium is filtered through 0.4 um filter to remove any cells and then used to infect cell type of interest such as HeLa, Raji, and THP1, in triplicates. Cells containing the integrated lentiviral genomes are selected by treatment with puromycin that is initiated 3 days post infection. RNA is harvested from stably selected cell lines using the RNA extraction kits (Qiagen), followed by reverse transcription into cDNA using reverse transcriptase kit (Thermo Fisher Scientific). cDNA samples are processed to generate indexed short-read libraries. Uniquely indexed short read libraries are multiplexed to sequence to generate about 20 million reads per sample, using the Illumina sequencing platform. Sequencing reads are stored, aligned, and analyzed using the BaseSpace Sequence Hub (Illumina). Targets of each individual candidate pre-miR are determined by comparing expression of genes in cell lines expressing the candidate pre-miR compared to in cell lines expressing the scrambled pre-miR. Ingenuity Pathway analysis is performed to test whether the pre-miRNas target specific pathways, especially therapeutically relevant pathways. A schematic of the workflow described in this Example is shown in Figure 30.

Example 31: Making an anellosome encoding a native Anellovirus pre-miR

This example describes a process to package either the replicating or non-replicating form of anellosomes expressing native Anellovirus pre-miRs.

The genome of the non-replicating form of the anellosome is synthesized containing the following components: (i) origin of replication, (ii) sequence encoding Anellovirus pre-miRNA, (iii) RNA polymerase III such as U6 or H1 driving the expression of pre-miRNA, and (iv) packaging signal. This genome is packaged by transfecting into a helper cell line that stably expresses all of the required proteins for viral packaging. The transfected cells are harvested 7 days post transfection and processed to make an anellosome preparation, as described herein. The genome equivalence titer of the anellosome preparation is determined by performing qPCR, as described herein. An appropriate dose of the anellosome preparation is then used for downstream applications.

The genome of the replicating form of the anellosome can be synthesized, for example, by generating the native Anellovirus, except that the expression of pre-miRNA sequence is manipulated using an exogenous promoter such as U6 or tissue specific promoters. The genome is packaged by transfecting into HEK-293T cells. The transfected cells are harvested 7 days post transfection and processed to make an anellosome preparation, as described herein. The genome equivalence titer of the

anellosome preparation is determined by performing qPCR, as described herein. An appropriate dose of the anellosome preparation is used for downstream applications.

Example 32: Utilizing Anellovirus pre-miRs a tumor suppressor in an *in vitro* cell culture model

This example describes studies to confirm the phenotypic effect of candidate pre-miRs identified as tumor suppressive from analysis, e.g., as described in Example 29.

Candidate pre-miRNAs having a tumor suppressive effect are identified based on analysis as described in Example 29. Anellosome preparations of the replicating form of anellosomes encoding these candidate pre-miRNAs, as well as scrambled pre-miRNAs, are prepared as described in Example 31. Cancer cell lines from the NCI-60 cancer cell line panel are plated in 96 well plates. When 30% confluent, these cell lines are treated with anellosomes comprising the candidate pre-miRs or scrambled pre-miRs at a dosage of five genome equivalents per cell. The anellosome-containing medium is aspirated five hours after infection, followed by washing with PBS twice, and replacing with fresh medium. Alamar blue assay is performed on the treated cells three days after treatment to determine which of the pre-miRs inhibit the proliferation of the cancer cell lines.

Example 33: Utilizing Anellovirus pre-miRs as tumor supppresors *in vivo*

This example describes *in vivo* experiments to confirm the tumor suppressive effect for shortlisted candidate tumor suppressive Anellovirus pre-miRs and cancer cell lines from *in vitro* analysis, as described in Example 32.

Xenografts are generated by subcutaneously injecting cancer cell lines shortlisted from the analysis described in Example 32, along with Matrigel, into the flanks of athymic mice. Once the xenograft tumors become palpable, local tumor injection of 3×10^6 genome equivalents of anellosomes encoding tumor suppressive pre-miRNAs or scramble pre-miRNA is performed. Effect of anellosome injection on tumor growth is determined by routine tumor growth measurements for three weeks, tumor weight measurement of the xenograft tumor at the end of the experiment, as well as by BrdU incorporation assays.

Example 34: Tandem copies of the Anellovirus genome

This example describes plasmid-based expression vectors harboring two copies of a single anelloviral genome, arranged in tandem such that the GC-rich region of the upstream genome is near the 5' region of the downstream genome (Figure 31A).

Anelloviruses replicate via rolling circle, in which a replicase (Rep) protein binds to the genome at an origin of replication and initiates DNA synthesis around the circle. For anellovirus genomes

contained in plasmid backbones, this requires either replication of the full plasmid length, which is longer than the native viral genome, or recombination of the plasmid resulting in a smaller circle comprising the genome with minimal backbone. Therefore, viral replication off of a plasmid can be inefficient. To improve viral genome replication efficiency, plasmids were engineered with tandem copies of TTV-tth8 and TTMV-LY2. These plasmids presented every possible circular permutation of the anelloviral genome: regardless of where the Rep protein binds, it will be able to drive replication of the viral genome from the upstream origin of replication to the downstream origin. A similar strategy has been used to produce porcine Anelloviruses (Huang et al., 2012, Journal of Virology 86 (11) 6042-6054).

Tandem TTV-tth8 was assembled by sequentially cloning copies of the genome into a plasmid backbone, leaving 12bp of non-viral DNA between the two sequences. Several TTV-tth8 variants were assembled into tandem plasmids, including wild-type and TTV-tth8(Δ36GC) (i.e., a TTV-tth8 genome engineered to include the 36-nucleotide GC-rich sequence described herein), which is missing 36 base pairs from the GC-rich region. Tandem TTMV-LY2 was assembled via Golden-gate assembly, simultaneously incorporating two copies of the genome into a backbone and leaving no extra nucleotides between the genomes.

Plasmid harboring tandem copies of TTV-tth8(Δ36GC) was transfected into HEK239T cells. Cells were incubated for five days, then lysed using 0.1% Triton X-100 and treated with nucleases to digest DNA not protected by viral capsids. qPCR was then performed using Taqman probes for the TTV-tth8 genome sequence and the plasmid backbone. TTV-tth8 genome copies were normalized to backbone copies. As shown in Figure 31B, tandem TTV-tth8 produced more than four times the number of viral genomes than single-copy harboring plasmids. When accounting for the doubled number of TTV-tth8 genome sequences, the tandem plasmid produced more than twice as many newly synthesized genome copies per transfected copy. These data demonstrated that engineering tandem Anelloviral genomes can increase viral genome replication and can be used as a strategy for increasing Anellovirus production.

Example 35: *In vitro* circularized Anellovirus genomes

This example describes constructs comprising circular, double stranded Anelloviral genome DNA with minimal non-viral DNA. These circular viral genomes more closely match the double-stranded DNA intermediates found during wild-type Anellovirus replication. When introduced into a cell, such circular, double stranded Anelloviral genome DNA with minimal non-viral DNA can undergo rolling circle replication to produce, for example, a genetic element as described herein.

In one example, plasmids harboring TTV-tth8 variants and TTMV-LY2 were digested with restriction endonucleases recognizing sites flanking the genomic DNA. The resulting linearized genomes were then ligated to form circular DNA. These ligation reactions were done with varying DNA

concentrations to optimize the intramolecular ligations. The ligated circles were either directly transfected into mammalian cells, or further processed to remove non-circular genome DNA by digesting with restriction endonucleases to cleave the plasmid backbone and exonucleases to degrade linear DNA. For TTV-tth8, *Xma*I endonuclease was used to linearize the DNA; the ligated circle contained 53bp of non-viral DNA between the GC-rich region and the 5' non-coding region. For TTMV-LY2, the type IIS restriction enzyme *Esp3*I was used, yielding a viral genomic DNA circle with no non-viral DNA. This protocol was adapted from previously published circularizations of TTV-tth8 (Kincaid et al., 2013, PLoS Pathogens 9(12): e1003818). To demonstrate the improvements in Anellovirus production, circularized TTV-tth8 and TTMV-LY2 were transfected into HEK293T cells. After 7 days of incubation, cells were lysed, and qPCR was performed to compare the levels of anellovirus genome between circularized and plasmid-based anelloviral genomes. Increased levels of Anelloviral genomes show that circularization of the viral DNA is a useful strategy for increasing Anellovirus production.

In another example, TTMV-LY2 plasmid (pVL46-240) and TTMV-LY2-nLuc were linearized with Esp3I or EcoRV-HF, respectively. Digested plasmid was purified on 1% agarose gels prior to electroelution or Qiagen column purification and ligation with T4 DNA Ligase. Circularized DNA was concentrated on a 100 kDa UF/DF membrane before transfection. Circularization was confirmed by gel electrophoresis, as shown in Figure 31C. T-225 flasks were seeded with HEK293T at 3 x 10^4 cells/cm^2 one day prior to lipofection with Lipofectamine 2000. Nine micrograms of circularized TTMV-LY2 DNA and 50 μg of circularized TTMV-LY2-nLuc were co-transfected one day post flask seeding. As a comparison, an additional T-225 flask was co-transfected with 50 μg of linearized TTMV-LY2 and 50 μg of linearized TTMV-LY2-nLuc.

Anellosome production proceeded for eight days prior to cell harvest in Triton X-100 harvest buffer. Generally, anellosomes can be enriched, e.g., by lysis of host cells, clarification of the lysate, filtration, and chromatography. In this example, harvested cells were nuclease treated prior to sodium chloride adjustment and 1.2 μm / 0.45 μm normal flow filtration. Clarified harvest was concentrated and buffer exchanged into PBS on a 750 kDa MWCO mPES hollow fiber membrane. The TFF retentate was filtered with a 0.45 μm filter before loading on a Sephacryl S-500 HR SEC column pre-equilibrated in PBS. Anellosomes were processed across the SEC column at 30 cm/hr. Individual fractions were collected and assayed by qPCR for viral genome copy number and transgene copy number, as shown in Figure 31D. Viral genomes and transgene copies were observed beginning at the void volume, Fraction 7, of the SEC chromatogram. A residual plasmid peak was observed at Fraction 15. Copy number for TTMV-LY2 genomes and TTMV-LY2-nLuc transgene were in good agreement for Anellosomes produced using circularized input DNA at Fraction 7 – Fraction 10, indicating packaged Anellosomes

containing nLuc transgene. SEC fractions were pooled and concentrated using a 100 kDa MWCO PVDF membrane and then 0.2 µm filtered prior to *in vivo* administration.

Circularization of input Anellosome DNA resulted a threefold increase in a percent recovery of nuclease protected genomes throughout the purification process when compared to linearized Anellosome DNA, indicating improved manufacturing efficiency using the circularized input Anellosome DNA as shown in Table 46.

Table 46. Purification Process Yields

Step	Linearized TTMV-LY2		Circularized TTMV-LY2	
	Total viral genome copies	**Total nLuc transgene genome copies**	**Total viral genome copies**	**Total nLuc transgene genome copies**
Harvest pre-nuclease	2.78E+12	2.17E+12	1.04E+11	4.39E+11
Clarified Harvest	9.96E+09	5.48E+09	6.55E+08	9.81E+08
TFF	1.01E+10	7.66E+09	2.58E+08	3.56E+08
SEC	3.18E+07	8.73E+06	9.16E+06	7.75E+06
UF/DF	8.82E+06	3.25E+06	1.78E+06	2.73E+06
Sterile Filtration	5.60E+06	2.64E+06	8.66E+05	1.63E+06
Purification Process Yield (%)	0.0002%	0.0001%	0.0006%	0.0004%

Example 36: Modelling ORF1 and identification of conserved residues and domains

This example describes *in silico* modelling of ORF1 proteins of Betatorqueviruses and defining putative domains based upon structural motifs and amino acid conservation/similarity.

The ORF1 protein is predicted to be the major capsid protein of *Anelloviruses*, based upon the presence of an arginine-rich region and the high presence of beta-sheets in secondary structure prediction using PSIpred (http://bioinf.cs.ucl.ac.uk/psipred/). RaptorX (http://raptorx.uchicago.edu/) was used for structure prediction and contact prediction for the sequences of eight Betatorqueviruses. Betatorquevirus ORF1 sequences were used as they are shorter (~650 amino acids) than Alphatorqueviruses (~750 amino acids) which fewer regions predicted to be unstructured. Five of the predicted structures contained elements of similarity which were used to identify putative domains of ORF1 (Figure 33). ORF1 was divided into five regions – the arginine-rich region, the putative core (jelly-roll domain), the hypervariable region, the N22 region, and the C-terminal domain.

The structural model of the Betatorquevirus strain CBS203 was used to display the residues/structural regions that have some conservation among the Betatorquevirus family. To analyze conserved residues, 110 Betatorquevirus ORF1 sequences were aligned in Geneious using the ClustalW alignment algorithm. Residues were then assessed for conservation by percent identity and similarity using the BLOSUM62 matrix with a threshold of 1. Residues which possessed similarity of greater than 60% of all strains in the alignment were highlighted on the structural model (Figure 34). In total, 26 residues (~4%) possessed amino acid similarity with 100% of aligned sequences. The 80% and 60% cut-offs contained 23.7% and 36.7% of total residues respectively.

A similar alignment algorithm and similarity determination was conducted on 258 strains of Alphatorqueviruses. The similarity and identity were displayed in the consensus sequence from the alignment and putative domains were assigned based upon primary sequence alignment with the Betatorqueviruses (Figure 35). Alphatorqueviruses possessed 29 residues (3.9%) which were 100% similar, remarkably consistent with the observation with Betatorqueviruses. Interestingly, Alphatorqueviruses possess a higher percentage of residues, when compared to Betatorqueviruses with at least 80% (30.9% of residues) or 60% (42.9% of residues) similarity.

Example 37: Production of anellosomes containing chimeric ORF1 with hypervariable domains from different Torque Teno Virus strains

This example describes domain swapping of hypervariable regions of ORF1 to produce chimeric anellosomes containing the ORF1 arginine-rich region, jelly-roll domain, N22, and C-terminal domain of one TTV strain, and the hypervariable domain from an ORF1 protein of a different TTV strain.

The full-length genome LY2 strain of Betatorquevirus has been cloned into expression vectors for expression in mammalian cells. This genome is mutated to remove the hypervariable domain of LY2 and replace it with the hypervariable domain of a distantly related Betatorqueviruses (Figure 36). The plasmid containing the LY2 genome with the swapped hypervariable domain (pTTMV-LY2-HVRa-z) is then

linearized and circularized using previously published methods (Kincaid et al., PLoS Pathogens 2013). HEK293T cells are transfected with the circularized genome and incubated for 5-7 days to allow anellosome production. After the incubation period anellosomes are purified from the supernatant and cell pellet of transfected cells by gradient ultracentrifugation.

To determine if the chimeric anellosomes are still infectious, the isolated viral particles are added to uninfected cells. The cells are incubated for 5-7 days to allow viral replication. After incubation the ability of the chimeric anellosomes to establish infection will be monitored by immunofluorescence, western blot, and qPCR. The structural integrity of the chimeric viruses is assessed by negative stain and cryo-electron microscopy. Chimeric anellosomes can further be tested for ability to infect cells *in vivo*. Establishment of the ability to produce functional chimeric anellosomes through hypervariable domain swapping could allow for engineering of viruses to alter tropism and potentially evade immune detection.

Example 38: Production of chimeric ORF1 containing non-TTV protein/peptides in place of hypervariable domains

This example describes the replacement of the hypervariable regions of ORF1 with other proteins or peptides of interest to produce chimeric ORF1 protein containing the arginine-rich region, jelly-roll domain, N22, and C-terminal domain of one TTV strain, and a non-TTV protein/peptide in place of the hypervariable domain.

As shown in example B, the hypervariable domain of LY2 is deleted from the genome and a protein or peptide of interest may be inserted into this region (Figure 37). Examples of types of sequences that could be introduced into this region include but are not limited to, affinity tags, single chain variable regions (scFv) of antibodies, and antigenic peptides. Mutated genomes in the plasmid (pTTMV-LY2-ΔHVR-POI) are linearized and circularized as described in example B. Circularized genomes are transfected into HEK293T cells and incubated for 5-7 days. Following incubation, the chimeric anellosomes containing the POI are purified from the supernatant and cell pellet via ultracentrifugation and/or affinity chromatography where appropriate.

The ability to produce functional chimeric anellosomes containing POIs is assessed using a variety of techniques. First, purified virus is added to uninfected cells to determine if chimeric anellosomes can replicate and/or deliver payload to naïve cells. Additionally, structural integrity of chimeric anellosomes is assessed using electron microscopy. For chimeric anellosomes that are functional *in vitro*, the ability of replicate/delivery payload *in vivo* is also assessed.

Example 39: Anellosome delivery of secreted enzyme *in vivo*

This example illustrates *in vivo* effector function of a secreted enzyme delivered by Anellosomes after administration.

Anellosomes comprising a transgene encoding ADAMTS13 (a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13) are prepared. Briefly, five constructs are produced: Construct A – TTMV-LY2 vector ± ADAMTS13; Construct B – ADAMTS13 protein and TTMV-LY2 ORFs; Construct C – Plasmids used in the production of TTMV-LY2 vector; Construct D – Plasmids used in the production of the ADAMTS13 protein and TTMV-LY2 ORFs; and Construct E – sterile PBS. Construct A and Construct B are produced in HEK-293T cells and purified via nuclease treatment, ultrafiltration/diafiltration, and sterile filtration. Construct C and Construct D are produced in *E. coli*, purified via MaxiPrep, and then diluted to a target copy number in PBS followed by sterile filtration. Construct E is produced by sterile filtration of PBS. HEK-293T cells are expanded to Passage 4 from thaw in DMEM + 10% FBS on a three- and four-day passage schedule. On Passage 5, cells are seeded at approximately 5 x 10^4 cells/cm^2 for next day transfection. Cells are co-transfected with constructs using Lipofectamine 2000 one day post seeding. After transfection, cells are incubated to permit anellosome production and anellosomes are harvested.

25uL of anellosome preparation or appropriate controls are administered to genetically-engineered VWD-type 2B mice intravenously. Blood draws are performed every day for each animal to determine hemolysis and thrombocytopenia, as well as measure ADAMTS13 secretion in the blood (SensoLyte ® 520 ADAMTS13 Activity Assay, *Fluorimetric* AnaSpec, Inc.)

The presence of ADAMTS13 signal are measured at each of days 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, and 21. Additionally, at the same time points, *in vitro* hemolysis assay protocol is performed. Briefly, blood is centrifuged and the absorbance of the supernatant, which includes plasma and lysed erythrocytes, is measured. Percent lysis is calculated from a standard curve of lysed erythrocytes (Triton X-100). Increased presences of ADAMTS13 and decreased hemolysis will demonstrate *in vivo* expression and activity of ADAMTS13 delivered via anellosome.

Example 40: Design of an anellosome harboring a payload

This example describes the design of an exemplary anellosome genetic element harboring a *trans* gene. The genetic element is composed of the essential *cis* replication and packaging domains from members of the Anelloviridae family along with non-Anellovirus payload, which may include, e.g., protein or non-coding RNA-expressing genes. The anellosome lacks essential *trans* protein elements for replication and packaging, and requires proteins provided by other sources (e.g., helpers, e.g., replicating viruses, expression plasmids, or genome integrations) for rolling circle replication and encapsidation.

In one set of examples, the entire protein-coding DNA sequence was deleted, from the first start codon to the last stop codon (Figure 38). For TTV-tth8, nucleotides 336 through 3015 were deleted, from the ORF2 start codon to the ORF3 stop codon. For TTMV-LY2, 424 through 2813 were deleted, from the ORF2 start codon to the ORF3 stop codon. The resulting DNA retained the viral non-coding region (NCR), including the viral promoter, the 5' UTR conserved domain, the 3' UTR (which encodes miRNAs in some anellovirus strains, such as TTV-tth8), and the GC-rich region. The anellosome NCR harbored essential *cis* domains, including the viral origin of replication and capsid binding domains. However, lacking the anellovirus protein-coding open reading frames, the anellosome was unable to express essential protein factors required for DNA replication and encapsidation, and therefore would not amplify or package unless these elements were provided in *trans*.

Payload DNA, including but not limited to protein-encoding sequences, full *trans* genes (including non-anelloviral promoter sequences), and non-coding RNA genes were incorporated into the anellosome genetic element by insertion into the site of the deleted anelloviral open reading frames (Figure 38). Expression from protein-coding sequences could be driven, for example, by either the native viral promoter or a synthetic promoter incorporated as a *trans* gene.

Replication-deficient or incompetent anellosome genetic elements (e.g., as described herein) may lack the protein-coding sequences for viral replication and/or capsid factors. Therefore, packaged anellosomes were produced by co-transfecting cells with the anellosome DNA described in this example and viral-protein-encoding DNA. The viral proteins were expressed off of replication-competent wild-type viral genomes, non-replicating plasmids harboring the viral proteins under control of the viral promoter, or plasmids harboring the viral proteins under control of a strong constitutive promoter.

Example 41: Anellosomes based on tth8 and LY2 each successfully transduced the EPO gene into lung cancer cells

In this example, a non-small cell lung cancer line (EKVX) was transduced using two different anellosomes carrying the erythropoeitin gene (EPO). The anellosomes were generated by in vitro circularization, as described herein, and included two types of anellosomes based on either an LY2 or tth8 backbone (e.g., as described in Tables 15 and 16, or Tables 5 and 6, respectively). Each of the LY2-EPO and tth8-EPO anellosomes included a genetic element that included the EPO-encoding cassette and non-coding regions of the LY2 or tth8 genome (5' UTR, GC-rich region), respectively, but did not include Anellovirus ORFs, e.g., as described in Example 40. Cells were inoculated with purified anellosomes or a positive control (AAV2-EPO at high dose or at the same dose as the anellosomes) and incubated for 7 days. Anellovirus ORFs were provided in *trans* in a separate in vitro circularized DNA. Culture supernatant was sampled 3, 5.5, and 7 days post-inoculation and assayed using a commercial ELISA kit

to detect EPO. Both LY2-EPO and tth8-EPO anellosomes successfully transduced cells, showing significantly higher EPO titers compared to untreated (negative) control cells ($P < 0.013$ at all time points) (Figure 40).

Example 42: Anellosomes with therapeutic transgenes can be detected *in vivo* after intravenous (i.v.) administration

In this example, anellosomes encoding human growth hormone (hGH) were detected *in vivo* after intravenous (i.v.) administration. Replication-deficient anellosomes, based on a LY2 backbone and encoding an exogenous hGH (LY2-hGH), were generated by in vitro circularization as described herein. The genetic element of the LY2-hGH anellosomes included LY2 non-coding regions (5' UTR, GC-rich region) and the hGH-encoding cassette, but did not include Anellovirus ORFs, e.g., as described in Example 40. LY2-hGH anellosomes were administered to mice intravenously. The Anellovirus ORFs were provided in *trans* in a separate in vitro circularized DNA. Briefly, anellosomes (LY2-hGH) or PBS was injected intravenously at day 0 (n=4 mice/group). Anellosomes were administered to independent animal groups at 4.66E+07 anellosome genomes per mouse.

In a first example, anellosome viral genome DNA copies were detected. At day 7, blood and plasma were collected and analyzed for the hGH DNA amplicon by qPCR. LY2-hGH anellosomes were present in the cellular fraction of whole blood after 7 days post infection in vivo (Figure 41A). Furthermore, the absence of anellosomes in plasma demonstrated the inability of these anellosomes to replicate in vivo (Figure 41B).

In a second example, hGH mRNA transcripts were detected after *in vivo* transduction. At day 7, blood was collected and analyzed for the hGH mRNA transcript amplicon by qRT-PCR. GAPDH was used as a control housekeeping gene. hGH mRNA transcripts in were measured in the cellular fraction of whole blood. mRNA from the anellosome-encoded transgene was detected in vivo (Figure 42).

Example 43: Coding sequence size distribution in Anelloviruses

The coding sequence (CDS) length of all Anelloviruses was assessed utilizing an extensive catalog of wild type strains identified internally. The CDS lengths of Anelloviruses was plotted, comparing virus strains across the three human Anellovirus genera (Alphatorqueviruses, alpha; Betatorqueviruses, beta; and Gammatorqueviruses, gamma) and comparing publicly available genome sequence lengths to those assembled internally (in house) by the present inventors. The mean CDS length of all Anelloviruses is about 2100 nucleotides. TTVs in the Alphatorquevirus genus were larger than Anelloviruses from the Betatorquevirus and Gammatorquevirus genera (TTV minis and TTV midis, respectively). Specifically, an average CDS of 2237 nucleotides was observed in Alphatorquevirus

TTVs, with a range of 1800-2541 nucleotides. An average CDS length of 2011 nucleotides was observed for Betatorqueviruses, with a range of 1803-2229 nucleotides. An average CDS length of 2012 nucleotides was observed for Gammatorqueviruses, with a range of 1812-2379 nucleotides.

Example 44: A highly conserved motif to characterize ORF2

Anellovirus ORF2, as shown in an exemplary genome in Figure 43A, likely encodes a non-structural protein with possible phosphatase activity and roles in viral replication and regulation of host immunity. An extensive viral sequence repository was examined for the presence of a conserved ORF2 amino acid motif (Figure 43B). This motif was then used to identify over 1,000 Anellovirus ORF2 sequences among in-house and public sequences. This ORF2 motif was found to remain conserved across a vast catalog of human Anellovirus strains, as well as all non-human Anelloviruses examined (rodent, pig, and primate Anelloviruses, as well as chicken anemia virus), making it the most highly conserved Anellovirus motif identified to date. ORF2 structural modelling was also performed, which revealed that the conserved residues in the ORF2 motif was maintained in a helix-turn-helix structure, with an orientation that suggests a possible metalbinding domain (Figure 43C). Interestingly, phylogenetic trees of ORF1 compared to ORF2 (Figure 43D) showed a similar genus-level breakdown by Alphatorqueviruses, Betatorqueviruses, and Gammatorqueviruses, indicating that ORF2s are genus-specific.

Example 45: Evidence for full-length Anellovirus ORF1 mRNA in humans

Anelloviruses express at least three alternatively spliced mRNAs *in vitro*, the longest of which (~2.2kb) is predicted to encode full-length ORF1. In this example, ORF1 mRNA transcription was assessed *in vivo*.

To do this, publicly available RNA Seq tissue data from the GTEx (Genotype-Tissue Expression) project was examined. The goal was to identify human tissue samples that contained enough Anellovirus RNA reads to categorize viral transcripts. 104 tissue samples with Anellovirus RNA reads were identified (2.4% of all tissues, 19% of blood samples); 7 of these samples had greater than 20 Anellovirus RNA reads, permitting viral transcriptome analysis. 3 of these 7 Anellovirus-positive samples also had matched WGS data, from which could be assembled the corresponding Anellovirus DNA genome for precise read mapping (Figure 44A). Absent corresponding viral reference genomes, Anellovirus diversity prohibits informative RNA read mapping. RNA reads that map to the ORF1 region were detected in three donors (two blood samples and one lung tissue sample). In one donor blood sample, Anellovirus RNA reads were identified that covered the full length ORF1 region (Figure 44B, grey bars depict read pairs). This is the first confirmation of full-length Anellovirus transcripts *in vivo* using RNA Seq data.

Example 46: In vitro circularized genome as input material for producing anellosomes in vitro

This example demonstrates that in vitro circularized (IVC) double stranded anellovirus DNA, as source material for an anellosome genetic element as described herein, is more robust than an anellovirus genome DNA in a plasmid to yield packaged anellosome genomes of the expected density.

1.2E+07 HEK293T cells (human embryonic kidney cell line) in T75 flasks were transfected with 11.25 ug of either, (i) in vitro circularized double stranded TTV-tth8 genome (IVC TTV-tth8), (ii) TTV-tth8 genome in a plasmid backbone, or (iii) plasmid containing just the ORF1 sequence of TTV-tth8 (non-replicating TTV-tth8). Cells were harvested 7 days post transfection, lysed with 0.1% Triton, and treated with 100 units per ml of Benzonase. The lysates were used for cesium chloride density analysis; density was measured and TTV-tth8 copy quantification was performed for each fraction of the cesium chloride linear gradient. As shown in Figure 45, IVC TTV-tth8 yielded dramatically more viral genome copies at the expected density of 1.33 as compared to TTV-tth8 plasmid.

1E+07 Jurkat cells (human T lymphocyte cell line) were nucleofected with either in-vitro circularized LY2 genome (LY2 IVC) or LY2 genome in plasmid. Cells were harvested 4 days post transfection and lysed using a buffer containing 0.5% triton and 300 mM sodium chloride, followed by two rounds of instant freeze-thaw. The lysates were treated with 100 units/ ml benzonase, followed by cesium chloride density analysis. Density measurement and LY2 genome quantification was performed on each fraction of the cesium chloride linear gradient. As shown in Figure 46, transfection of in vitro circularized LY2 genome in Jurkat cells led to a sharp peak at the expected density, as compared to the transfection of plasmid containing the LY2 genome, which showed no detectable peak in Figure 46.

Example 47: Identification of conserved secondary structural motifs in Anellovirus ORF1

In this example, computational modelling was used to identify conserved motifs in the secondary structure of the Anellovirus ORF1 protein. Secondary structure predictions were conducted on single strains using the program JPred.

Generally, the jelly-roll domain of human TTVs are approximately 200 amino acids (AA) ± 3 AA in length. The secondary structure of an exemplary jelly-roll domain begins with a beta strand of 5-7 AA, followed by a 3-5 AA random coil, a 15-16 AA beta strand, a 26-28 AA random coil, a 15-17 AA alpha helix, a 2 AA random coil, a 3-4 AA beta strand, an 8 AA random coil, a 10-11 AA beta strand, a 5-6 AA random coil, a 6-7 AA beta strand, a 8-14 AA random coil, a 8-14 AA alpha-helix (which may be broken into 2 smaller helices in some instances), a 3-4 AA random coil, a 4-5 AA beta strand, a 10 AA random coil, a 5-6 AA beta strand, a 20-21 AA random coil, a 7-9 AA beta strand, a 14-16 AA random coil, a 5-7

AA beta strand. An alignment of exemplary Anellovirus ORF1 secondary structures from the Alphatorquevirus, Betatorquevirus, and Gammatorquevirus clades is shown in Figure 47.

The secondary structure of the YNPX^{2}DXGX^{2}N (SEQ ID NO: 829) motif in the N22 domain of ORF1 also has a conserved secondary structure surrounding it. Starting with a 5-6 AA beta strand that breaks after the tyrosine (Y) at position 1 of the motif, most of the motif lines in an 8-9 AA random coil, until the terminal asparagine (N) at which point another beta strand of 7-8 AA originates. An alignment of exemplary Anellovirus ORF1 N22 motif sequences is shown in Figure 48. The tyrosine in the motif breaks a beta strand, and a second beta strand starts on the terminal asparagine of the motif.

What is claimed is:

1. A synthetic anellosome comprising:

(I) a genetic element comprising:

(a) a promoter element,

(b) a nucleic acid sequence encoding an exogenous effector, wherein the nucleic acid sequence is operably linked to the promoter element, and wherein the exogenous effector is erythropoietin (EPO), and

(c) a 5' UTR domain comprising a nucleic acid sequence of nucleotides 323 – 393 of SEQ ID NO: 54, or a nucleic acid sequence at least 90% identical thereto;

(II) a proteinaceous exterior comprising a polypeptide encoded by *Anellovirus* ORF1 nucleic acid, e.g., an ORF1 molecule;

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the synthetic anellosome is capable of delivering the genetic element into a human cell.

2. The synthetic anellosome of claim 1, wherein the ORF1 molecule comprises the amino acid sequence of SEQ ID NO: 217, or an amino acid sequence having least 90% identity thereto.

3. The synthetic anellosome of any of the preceding claims, wherein the ORF1 molecule is encoded by nucleotides 612-2612 of SEQ ID NO: 54.

4. The synthetic anellosome of any of the preceding claims, wherein the genetic element comprises the nucleic acid sequence of nucleotides 2868-2929 of SEQ ID NO: 54, or a nucleic acid sequence having at least 85% sequence identity thereto.

5. The synthetic anellosome of any of the preceding claims, wherein the ORF1 molecule comprises an amino acid sequence comprising one or more of the amino acid sequences of an arg-rich region, jelly-roll domain, hypervariable domain, N22 domain, and/or C-terminal domain as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

6. The synthetic anellosome of any of the preceding claims, wherein the ORF1 molecule comprises the amino acid sequence of SEQ ID NO: 58, or a nucleic acid sequence having at least 85% sequence identity thereto.

7. The synthetic anellosome of any of the preceding claims, further comprising a polypeptide comprising the amino acid sequence of an ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

8. The synthetic anellosome of any of the preceding claims, wherein the genetic element encodes the amino acid sequence of an ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

9. The synthetic anellosome of any of the preceding claims, wherein the synthetic anellosome does not comprise a polypeptide comprising the amino acid sequence of an ORF2, ORF2/2, ORF2/3, TAIP, ORF1/1, or ORF1/2 as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

10. The synthetic anellosome of any of the preceding claims, wherein the genetic element does not encode the amino acid sequence of an ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 as listed in Table 16, or an amino acid sequence having at least 85% identity thereto.

11. The synthetic anellosome of any of the preceding claims, wherein the ORF1 molecule comprises the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829), wherein X^n is each independently a contiguous sequence of any n amino acids.

12. The synthetic anellosome of claim 11, wherein the ORF1 molecule further comprises a first beta strand and a second beta strand flanking the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829), e.g., wherein the first beta strand comprises the tyrosine (Y) residue of the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829) and/or wherein the second beta strand comprises the second asparagine (N) residue (from N to C) of the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829).

13. The synthetic anellosome of any of the preceding claims, wherein the ORF1 molecule comprises, 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.

14. The synthetic anellosome of any of the preceding claims, wherein the genetic element is capable of being amplified by rolling circle replication in a host cell, e.g., to produce at least 8 copies.

15. The synthetic anellosome of any of the preceding claims, wherein the genetic element is single-stranded.

16. The synthetic anellosome of any of the preceding claims, wherein the genetic element is circular.

17. The synthetic anellosome of any of the preceding claims, wherein the genetic element is DNA.

18. The synthetic anellosome of any of the preceding claims, wherein the genetic element is a negative strand DNA.

19. The synthetic anellosome of any of the preceding claims, wherein the genetic element integrates at a frequency of less than 10%, 8%, 6%, 4%, 3%, 2%, 1%, 0.5%, 0.2%, 0.1% of the anellosomes that enters the cell, e.g., wherein the synthetic anellosome is non-integrating.

20. The synthetic anellosome of any of the preceding claims, wherein the genetic element comprises a sequence of the Consensus 5' UTR nucleic acid sequence shown in Table 16-1.

21. The synthetic anellosome of any of the preceding claims, wherein the genetic element comprises a sequence of the Consensus GC-rich region shown in Table 16-2.

22. The synthetic anellosome of any of the preceding claims, wherein the genetic element comprises a sequence of at least 100 nucleotides in length, which consists of G or C at at least 70% (e.g., about 70- 100%, 75-95%, 80-95%, 85-95%, or 85-90%) of the positions.

23. The synthetic anellosome of any of the preceding claims, wherein the genetic element comprises the nucleic acid sequence of SEQ ID NO: 120.

24. The synthetic anellosome of any of the preceding claims, wherein the promoter element is exogenous to wild-type Anellovirus.

25. The synthetic anellosome of any of the preceding claims, wherein the promoter element is endogenous to wild-type Anellovirus.

26. The synthetic anellosome of any of the preceding claims, wherein 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.

27. The synthetic anellosome of any of the preceding claims, wherein the synthetic anellosome is capable of infecting human cells, e.g., neurons (e.g., motor neurons), myocytes, retinal cell, liver cells, blood cells, skin cells, nerve cells, adipose cells, or endothelial cells, e.g., in vitro.

28. The synthetic anellosome of any of the preceding claims, which is substantially non-immunogenic, e.g., does not induce a detectable and/or unwanted immune response, e.g., as detected according to the method described in Example 4.

29. The synthetic anellosome of claim 28, wherein the substantially non-immunogenic anellosome has an efficacy in a subject that is a least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of the efficacy in a reference subject lacking an immune response.

30. The synthetic anellosome of any of the preceding claims, wherein a population of at least 1000 of the anellosomes is capable of delivering at least about 100 copies (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 copies) of the genetic element into one or more human cells.

31. A synthetic anellosome comprising:

(I) a genetic element comprising:

(a) a promoter element,

(b) a nucleic acid sequence encoding an exogenous effector, wherein the nucleic acid sequence is operably linked to the promoter element,

wherein the exogenous effector comprises a polypeptide that, when mutated, causes a human disease, or a functional variant of said polypeptide, chosen from:

(i) a therapeutic enzyme,

(ii) a structural protein, or

(iii) a complement protein; and

(c) a 5' UTR domain comprising a nucleic acid sequence of nucleotides 323 – 393 of SEQ ID NO: 54, or a nucleic acid sequence at least 85% identical thereto; or

(II) a proteinaceous exterior comprising a polypeptide encoded by an ORF1 nucleic acid, e.g., an ORF1 molecule;

wherein the genetic element is enclosed within the proteinaceous exterior; and

wherein the synthetic anellosome is capable of delivering the genetic element into a human cell.

32. A pharmaceutical composition comprising the synthetic anellosome of any of the preceding claims, and a pharmaceutically acceptable carrier or excipient.

33. The pharmaceutical composition of claim 32, which comprises at least 10^3, 10^4, 105, 10^6, 10^7, 10^8, or 10^9 synthetic anellosomes.

34. The pharmaceutical composition of claim 32 or 33, wherein the pharmaceutical composition has a predetermined ratio of particles:infectious units (e.g., <300:1, < 200:1, <100:1, or <50:1).

35. A reaction mixture comprising:

(i) a first nucleic acid (e.g., a double-stranded or single-stranded circular DNA) comprising the sequence of the genetic element of the synthetic anellosome of any of the preceding claims, and

(ii) a second nucleic acid sequence 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 listed in Table 16, or an amino acid sequence having at least 85% sequence identity thereto.

36. The reaction mixture of claim 35, wherein the first nucleic acid and second nucleic acid are in the same nucleic acid molecule.

37. The reaction mixture of claim 35, wherein the first nucleic acid and second nucleic acid are different nucleic acid molecules.

38. The reaction mixture of claim 35, wherein the first nucleic acid and second nucleic acid are different nucleic acid molecules and wherein the second nucleic acid is provided as double-stranded circular DNA.

39. The reaction mixture of claim 35, wherein the first nucleic acid and second nucleic acid are different nucleic acid molecules and wherein the first and the second nucleic acid are provided as double-stranded circular DNA.

40. The reaction mixture of claim 37, wherein the second nucleic acid sequence is comprised by a helper cell or helper virus.

41. A method of making a synthetic anellosome, the method comprising:

a) providing a host cell comprising:

(i) a first nucleic acid molecule comprising the nucleic acid sequence of a genetic element of a synthetic anellosome of any of the preceding claims, and

(ii) 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 listed in any of Table 16, or an amino acid sequence having at least 85% sequence identity thereto; and

b) incubating the host cell under conditions suitable to make a synthetic anellosome;

thereby making the synthetic anellosome.

42. The method of claim 41, further comprising, prior to step (a), introducing the first nucleic acid molecule and/or the second nucleic acid molecule into the cell.

43. The method of claim 42, wherein the second nucleic acid molecule is introduced into the host cell prior to, concurrently with, or after the first nucleic acid molecule.

44. The method of any of claims 41 or 42, wherein the second nucleic acid molecule is integrated into the genome of the host cell.

45. The method of any of claims 41-44, wherein the second nucleic acid molecule is a helper (e.g., a helper plasmid or the genome of a helper virus).

46. The method of any of claims 41-44, wherein second nucleic acid molecule encodes an ORF2 molecule comprising 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.

47. A method of manufacturing a synthetic anellosome preparation, the method comprising:

a) providing a plurality of synthetic anellosomes according to claims 1-31, a pharmaceutical composition of any of claims 32-34, or a reaction mixture of any of claims 35-40;

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); and

c) formulating the plurality of synthetic anellosomes, 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.

48. A host cell comprising:

(i) a first nucleic acid molecule comprising the nucleic acid sequence of a genetic element of a synthetic anellosome of any of the preceding claims, 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 as listed in any of Table 16, or an amino acid sequence having at least 85% sequence identity thereto.

49. A method of delivering an exogenous effector (e.g., a therapeutic exogenous effector) to a mammalian cell, comprising:

(a) providing a synthetic anellosome of any of the preceding claims; and

(b) contacting a mammalian cell with the synthetic anellosome;

wherein the synthetic anellosome is capable of delivering the genetic element into the mammalian cell; and

optionally wherein the synthetic anellosome is produced by introducing the genetic element into a host cell, under conditions suitable for enclosing the genetic element within the proteinaceous exterior in the host cell;

thereby delivering the therapeutic exogenous effector to the mammalian cell.

50. Use of a synthetic anellosome of any of the claims 1-31 or the pharmaceutical composition of any of claims 32-34 for delivering the genetic element to a host cell.

51. Use of a synthetic anellosome of any of the claims 1-31 or the pharmaceutical composition of any of claims 32-34 for treating a disease or disorder in a subject.

52. A synthetic anellosome of any of claims 1-31 or the pharmaceutical composition of any of claims 32-34, for use in treating a disease or disorder in a subject.

53. A method of treating a disease or disorder in a subject, the method comprising administering a synthetic anellosome of any of claims 1-31 or the pharmaceutical composition of any of claims 32-34 to the subject.

54. Use of the synthetic anellosome of any of claims 1-31 or the pharmaceutical composition of any of claims 32-34, in the manufacture of a medicament for treating a disease or disorder in a subject.

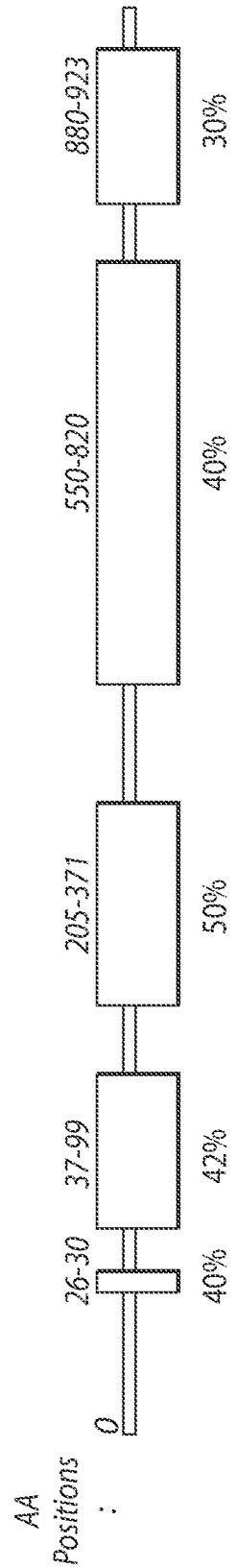


Figure 1A

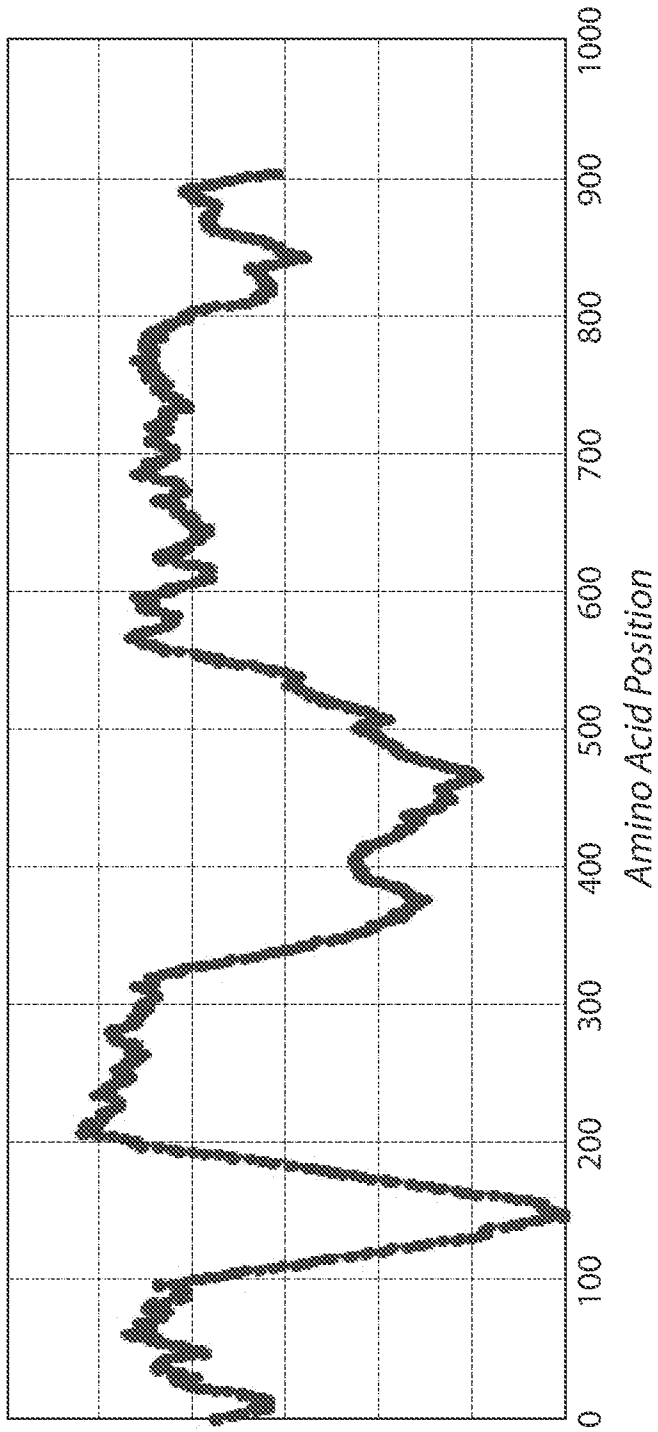


Figure 1B

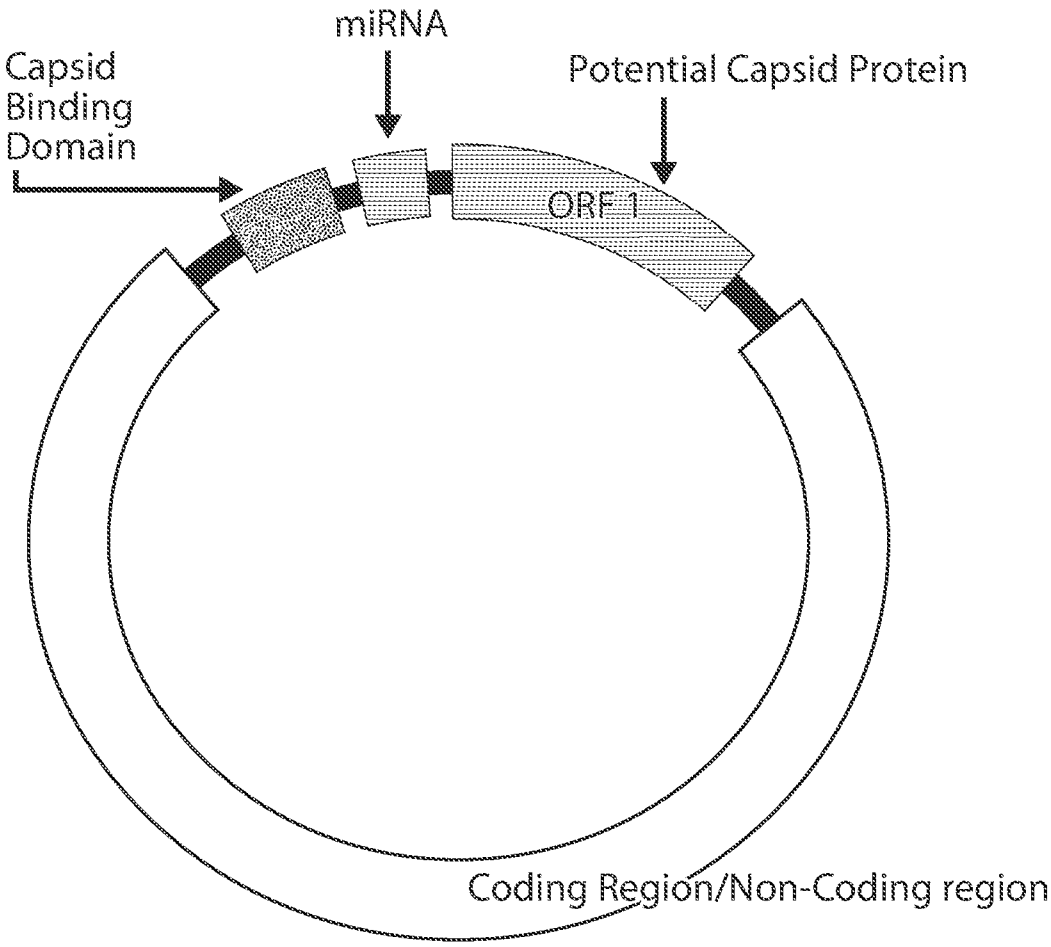
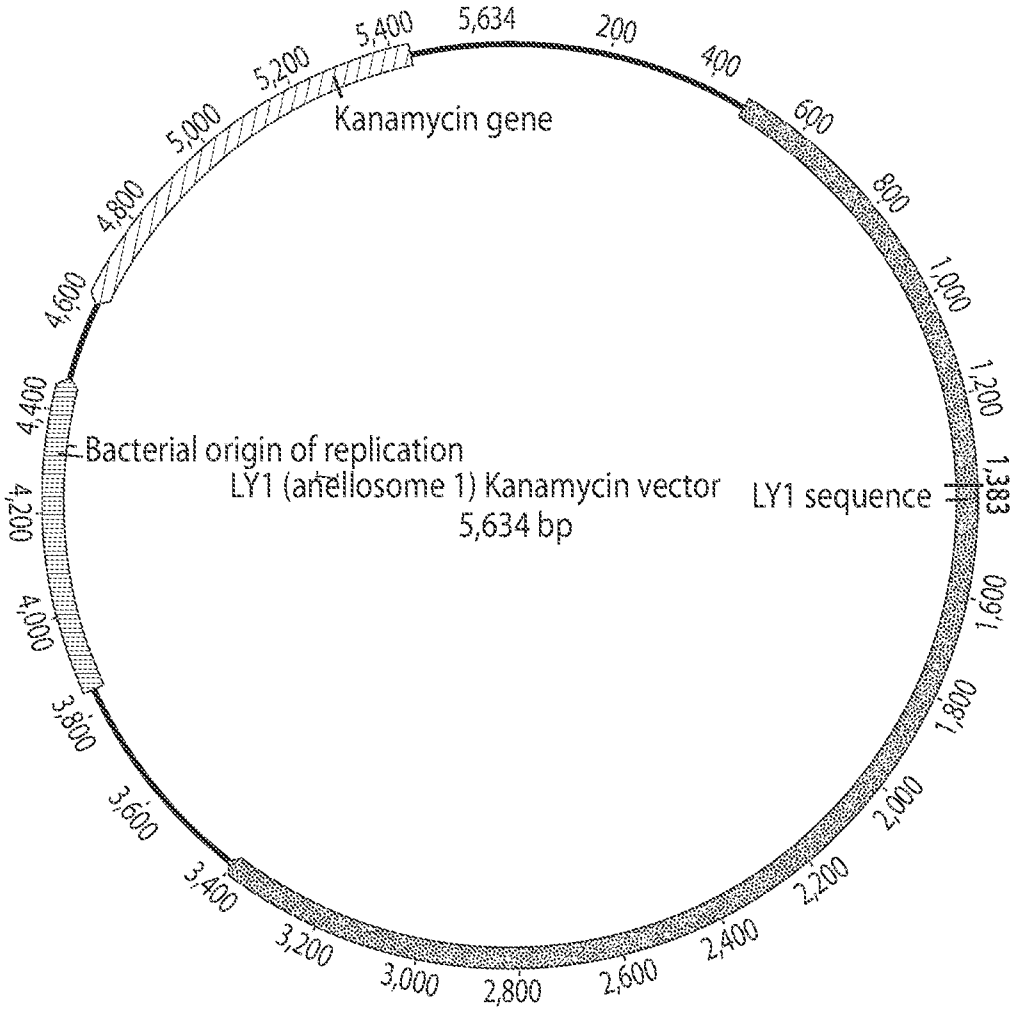
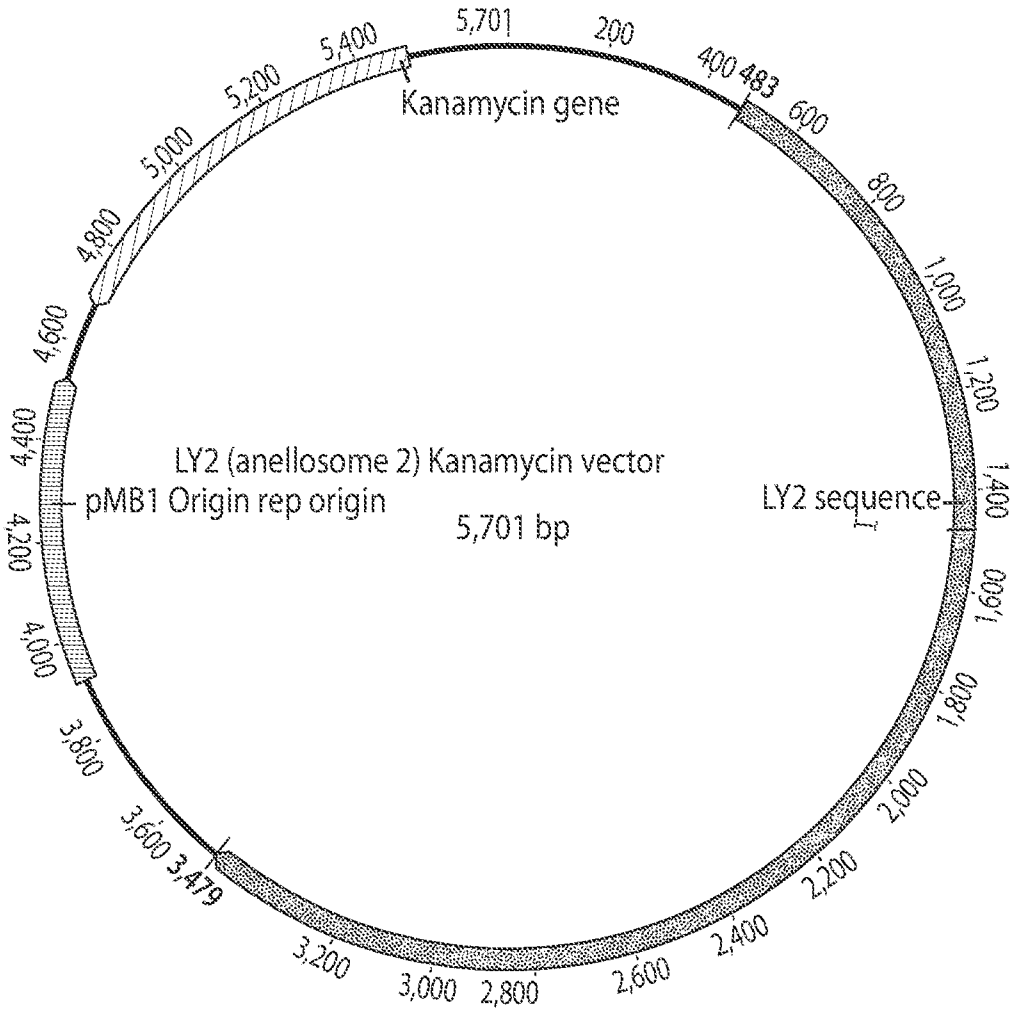

miRNA
Capsid
Binding
Domain
Potential Capsid Protein
ORF 1
Coding Region/Non-Coding region

Figure 2



Schematic of the kanamycin vector encoding the LY1 strain of TTMiniV (Anellosome 1)

Figure 3



Schematic of the kanamycin vector encoding the LY2 strain of TTMiniV (Anellosome 2)

Figure 4

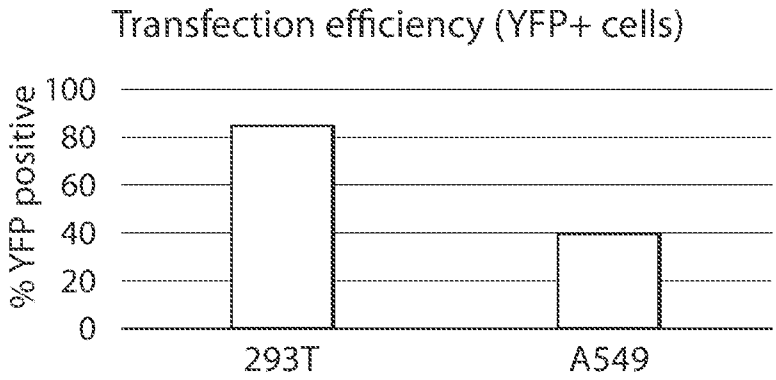

Transfection efficiency (YFP+ cells)
% YFP positive
100
80
60
40
20
0
293T
A549

Figure 5

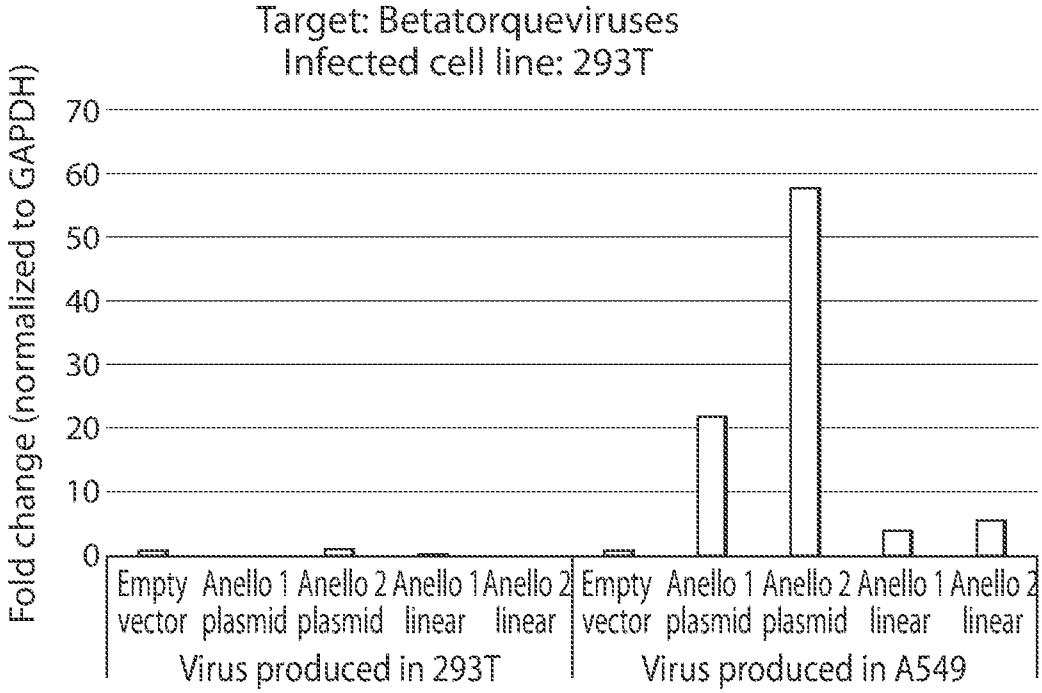

Target: Betatorqueviruses
Infected cell line: 293T
Fold change (normalized to GAPDH)
70
60
50
40
30
20
10
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 6A

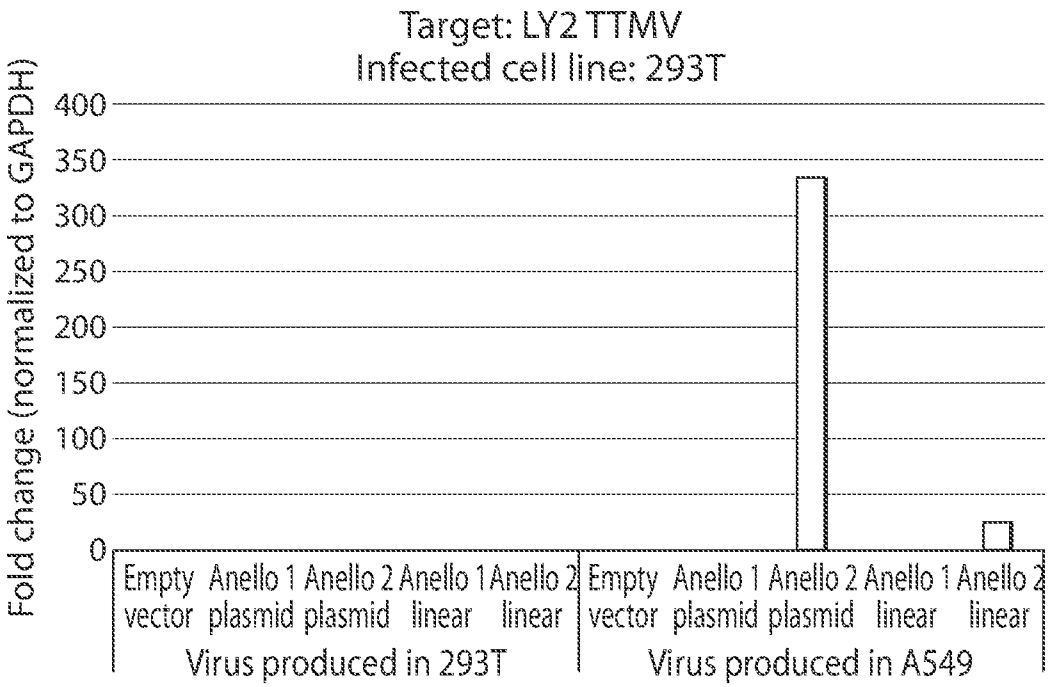

Target: LY2 TTMV
Infected cell line: 293T
Fold change (normalized to GAPDH)
400
350
300
250
200
150
100
50
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 6B

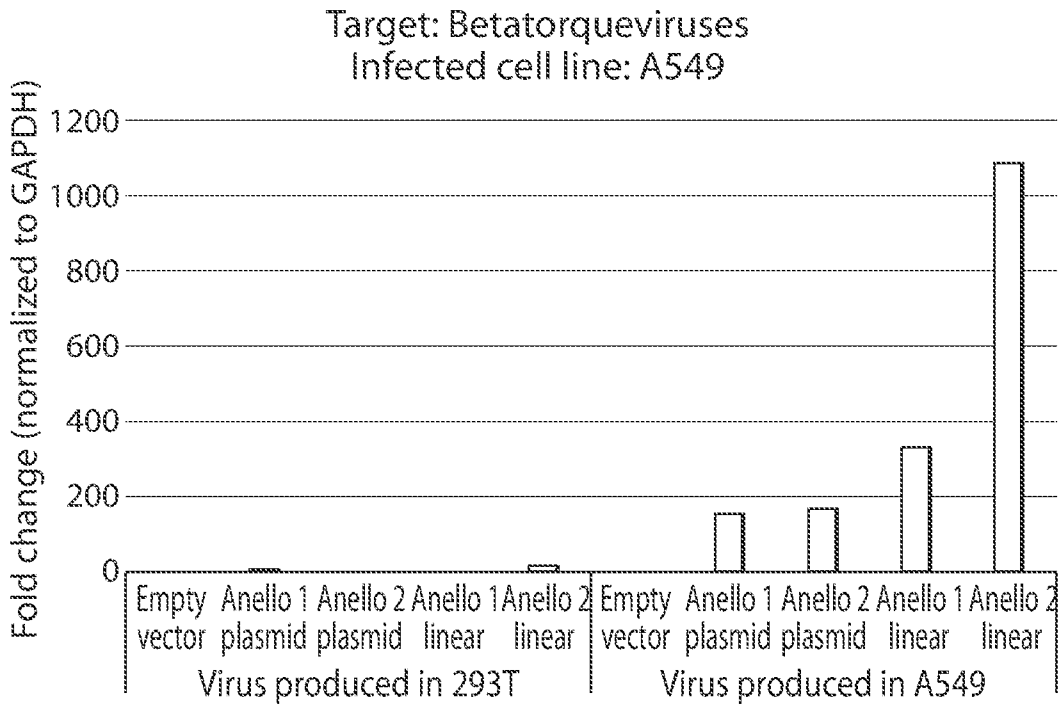

Target: Betatorqueviruses
Infected cell line: A549
Fold change (normalized to GAPDH)
1200
1000
800
600
400
200
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 7A

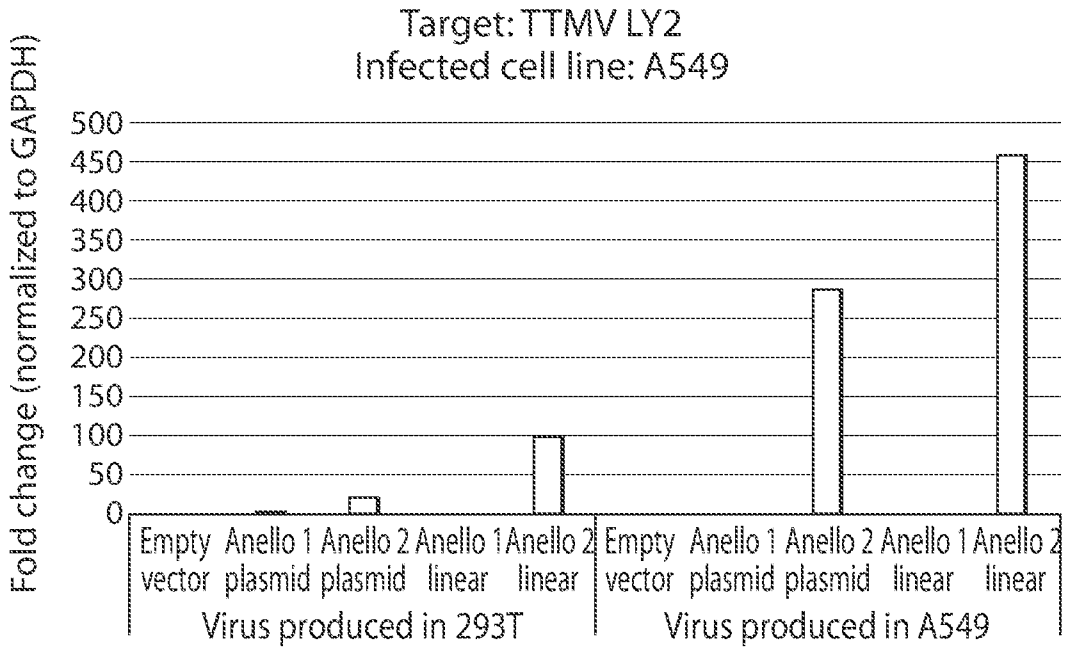

Target: TTMV LY2
Infected cell line: A549
Fold change (normalized to GAPDH)
500
450
400
350
300
250
200
150
100
50
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 7B

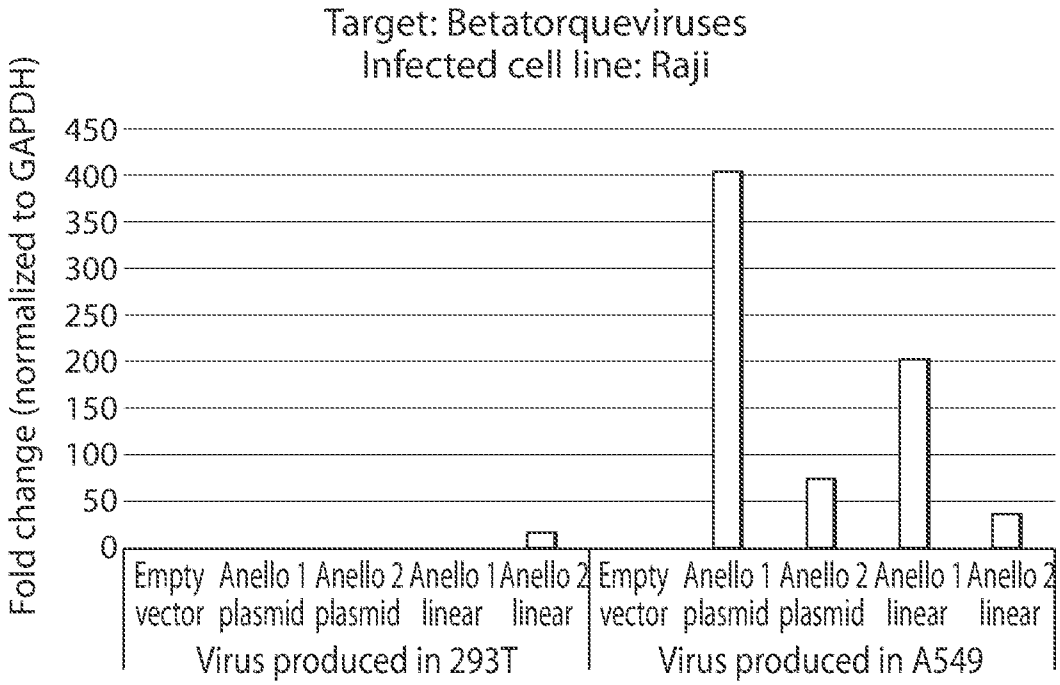

Target: Betatorqueviruses
Infected cell line: Raji
Fold change (normalized to GAPDH)
450
400
350
300
250
200
150
100
50
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 8A

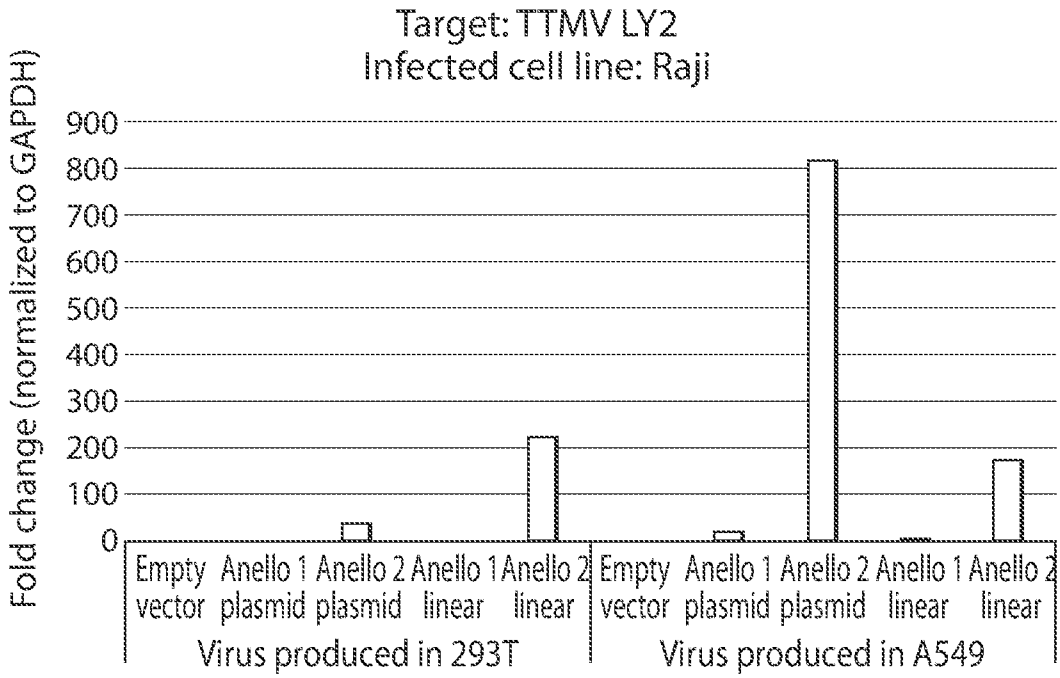

Target: TTMV LY2
Infected cell line: Raji
Fold change (normalized to GAPDH)
900
800
700
600
500
400
300
200
100
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 8B

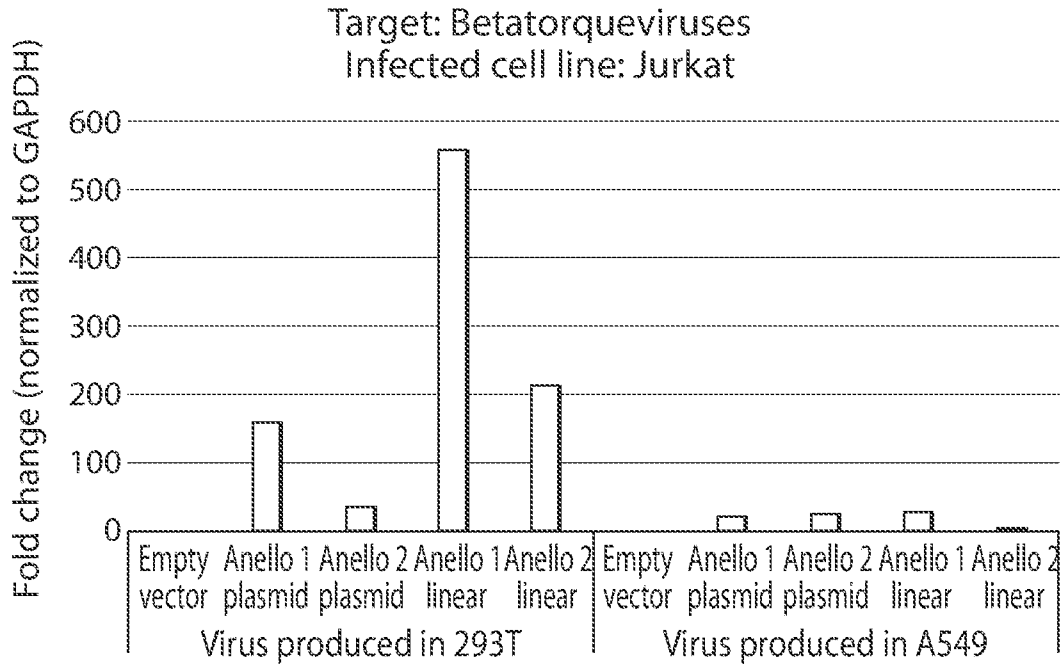

Target: Betatorqueviruses
Infected cell line: Jurkat
Fold change (normalized to GAPDH)
600
500
400
300
200
100
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 9A

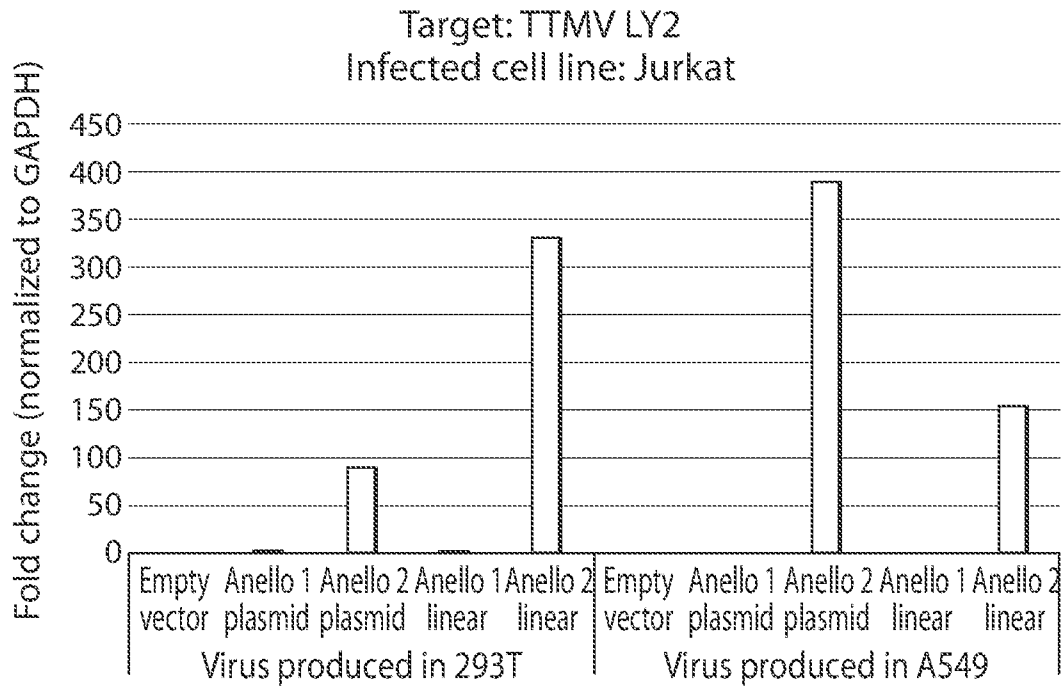

Target: TTMV LY2
Infected cell line: Jurkat
Fold change (normalized to GAPDH)
450
400
350
300
250
200
150
100
50
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 9B

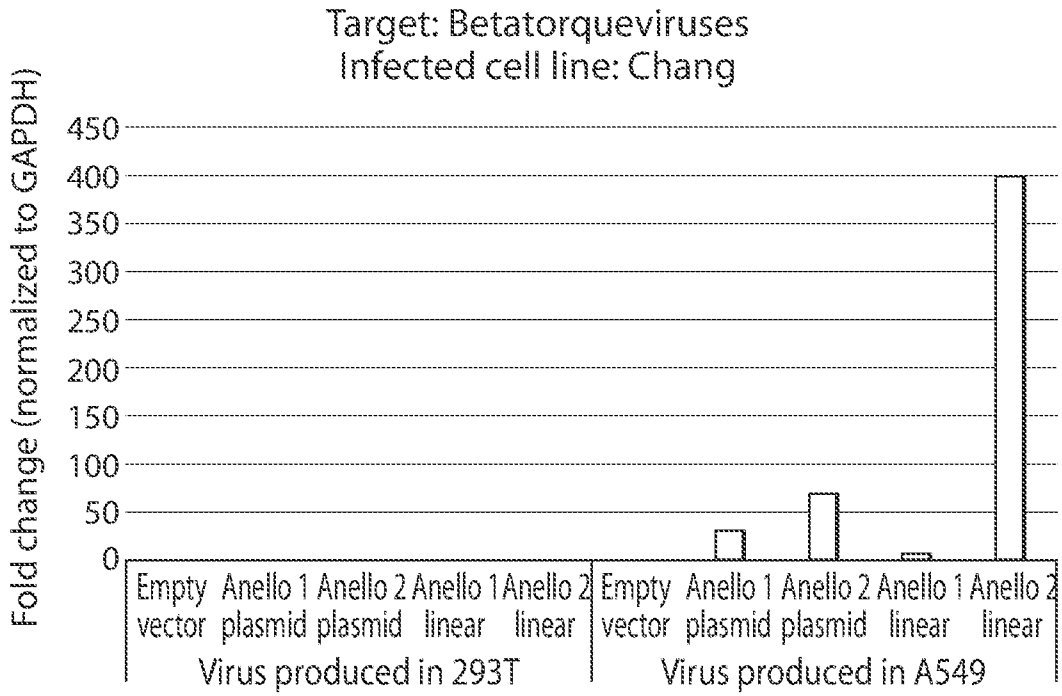

Target: Betatorqueviruses
Infected cell line: Chang
Fold change (normalized to GAPDH)
450
400
350
300
250
200
150
100
50
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 10A

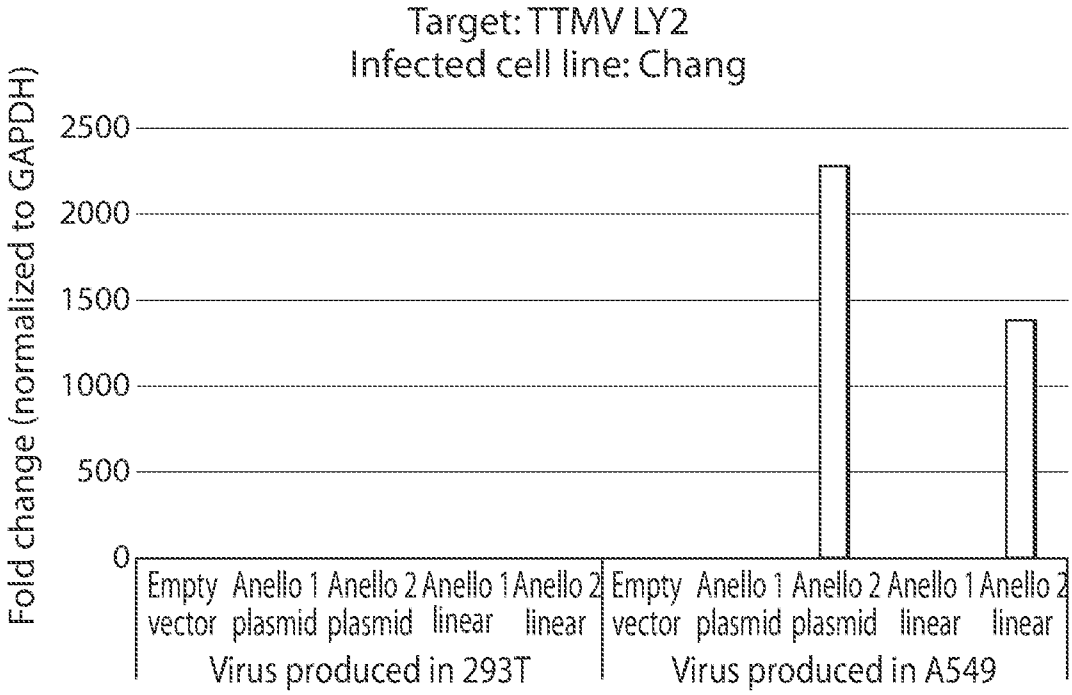

Target: TTMV LY2
Infected cell line: Chang
Fold change (normalized to GAPDH)
2500
2000
1500
1000
500
0
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Empty vector
Anello 1 plasmid
Anello 2 plasmid
Anello 1 linear
Anello 2 linear
Virus produced in 293T
Virus produced in A549

Figure 10B

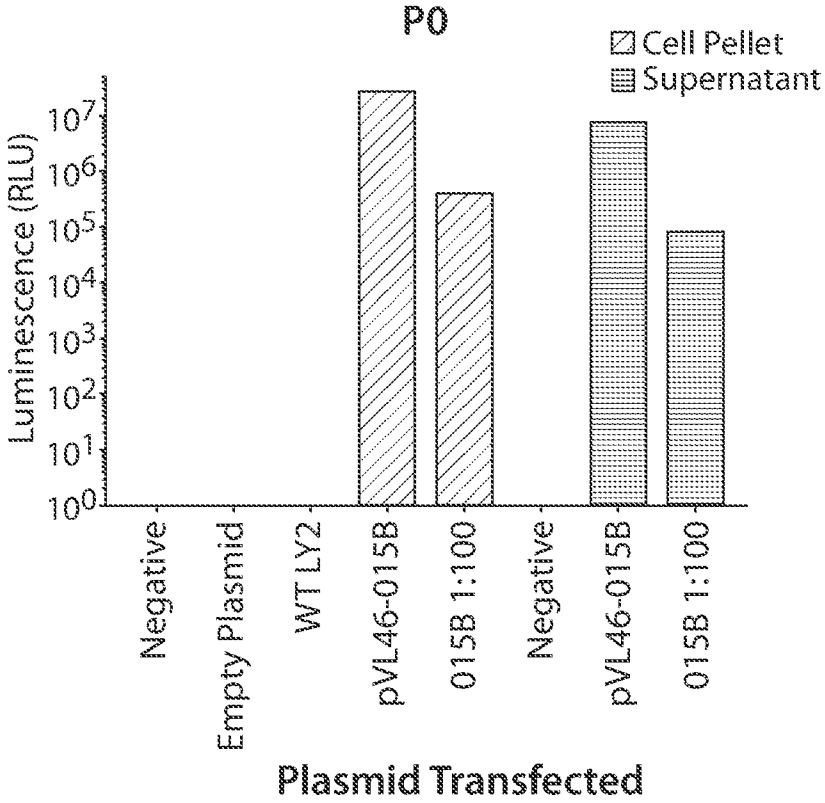


Figure 11A

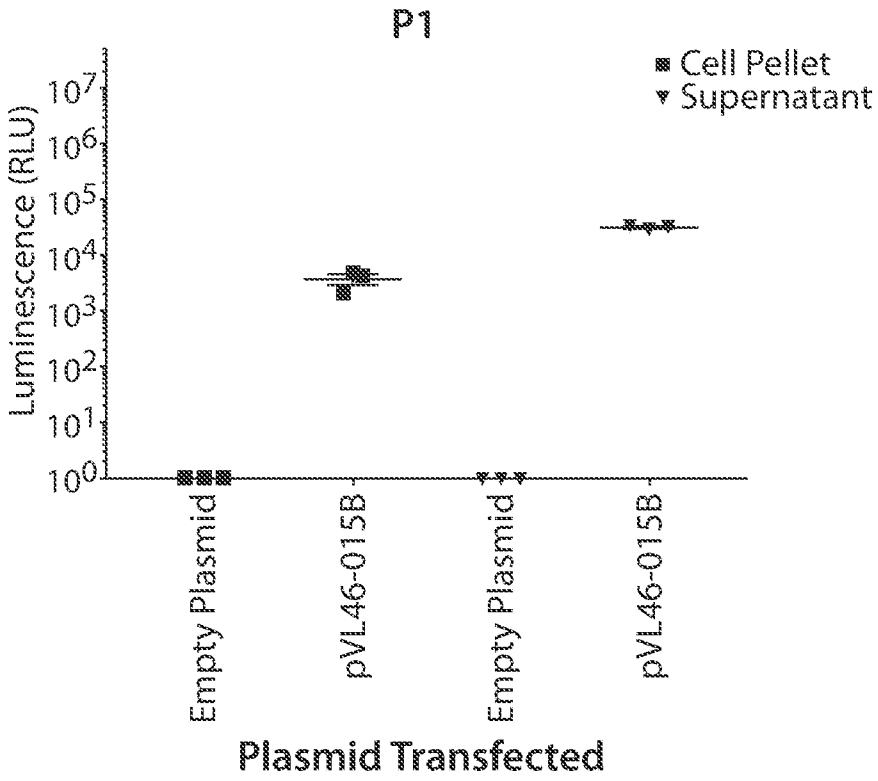


Figure 11B

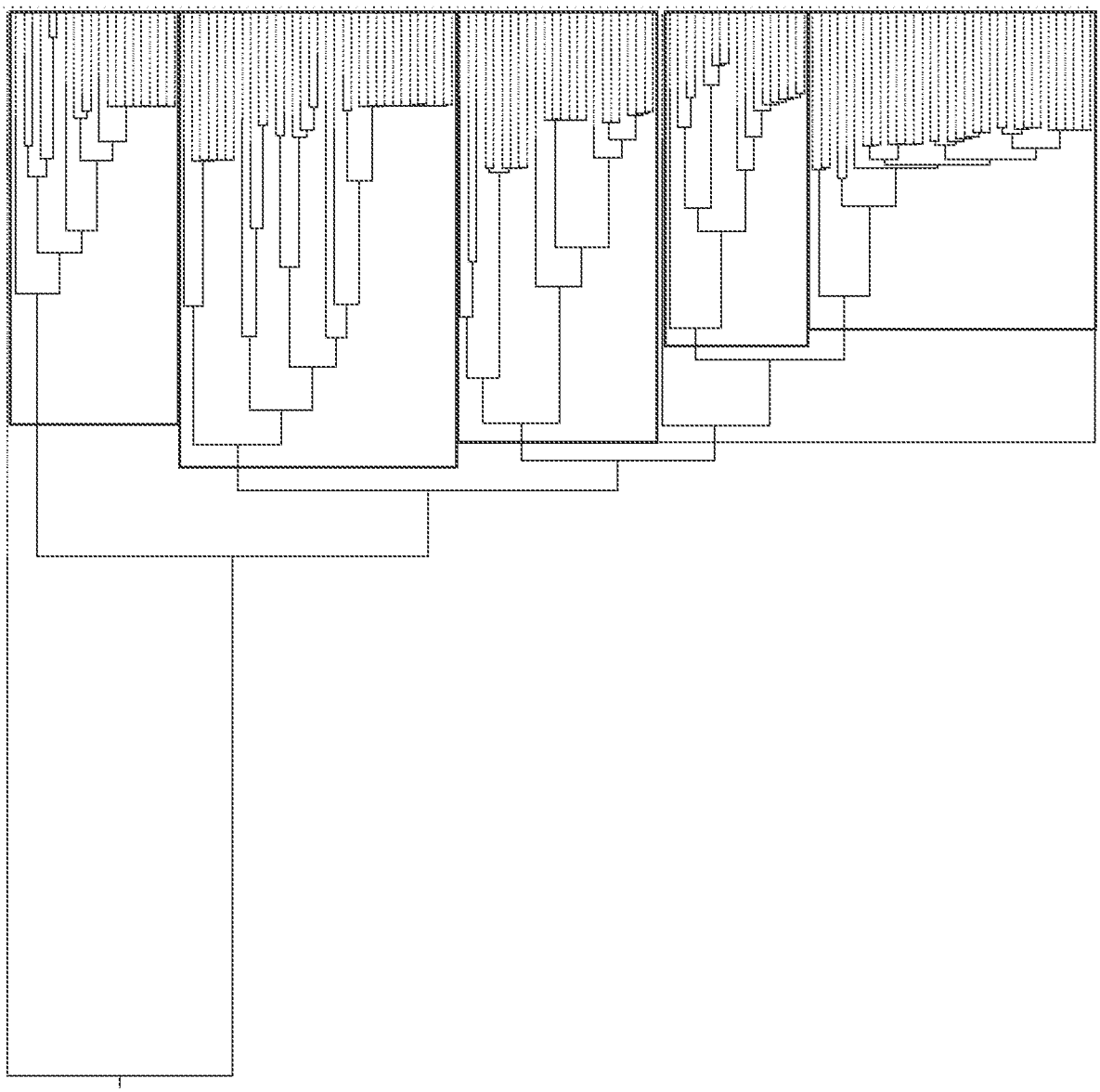


Figure 11C

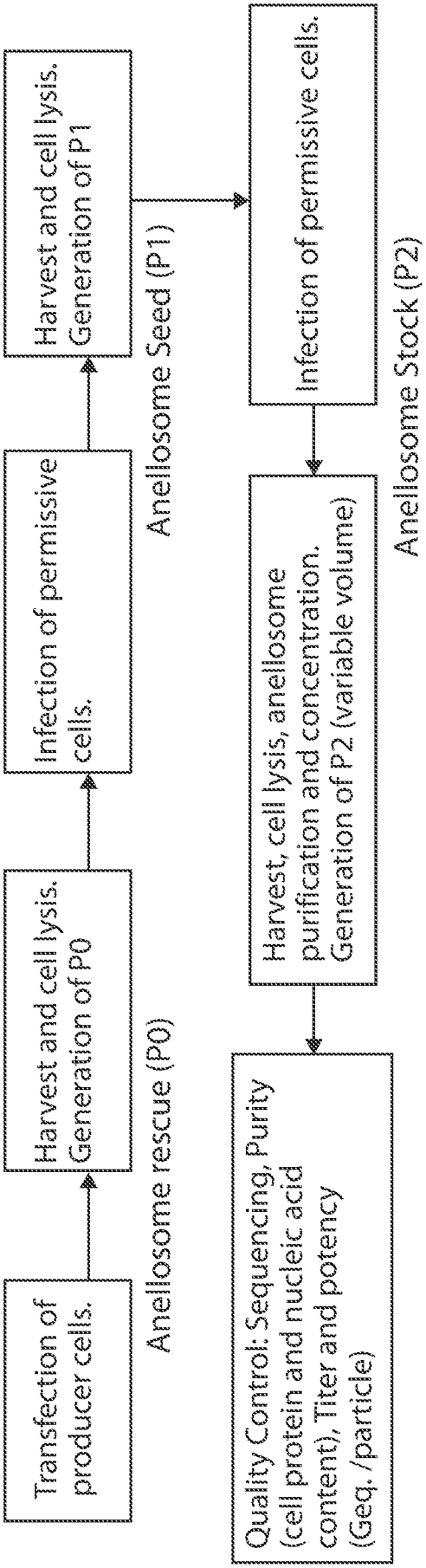

Transfection of producer cells.
Harvest and cell lysis. Generation of P0
Anellosome rescue (P0)
Infection of permissive cells.
Harvest and cell lysis. Generation of P1
Anellosome Seed (P1)
Infection of permissive cells.
Harvest, cell lysis, anellosome purification and concentration. Generation of P2 (variable volume)
Anellosome Stock (P2)
Quality Control: Sequencing, Purity (cell protein and nucleic acid content), Titer and potency (Geq. /particle)

Figure 12

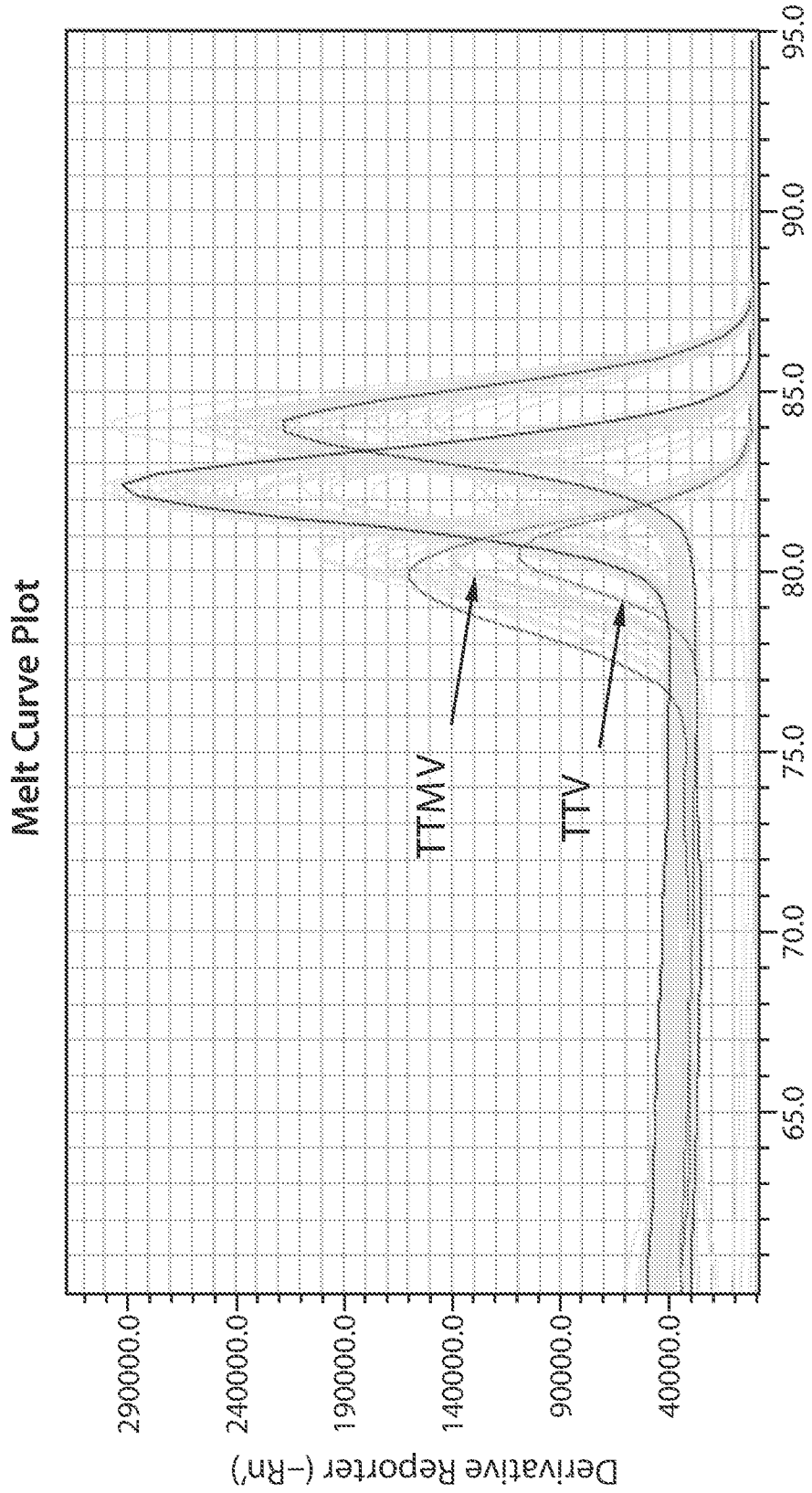

Melt Curve Plot
Derivative Reporter (-Rn')
290000.0
240000.0
190000.0
140000.0
90000.0
40000.0
65.0
70.0
75.0
80.0
85.0
90.0
95.0
Temperature (°C)
TTMV
TTV

Figure 13

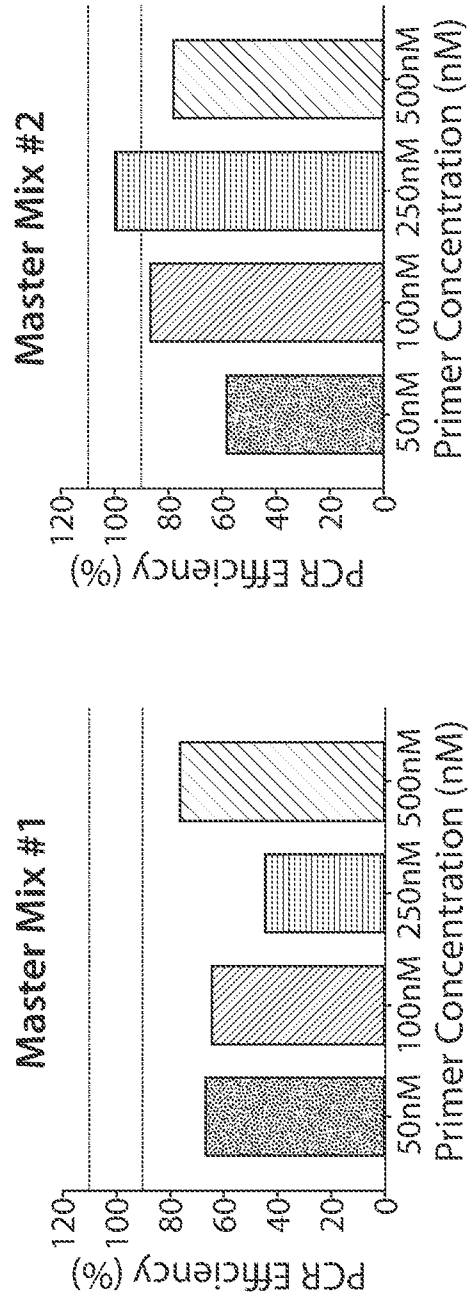
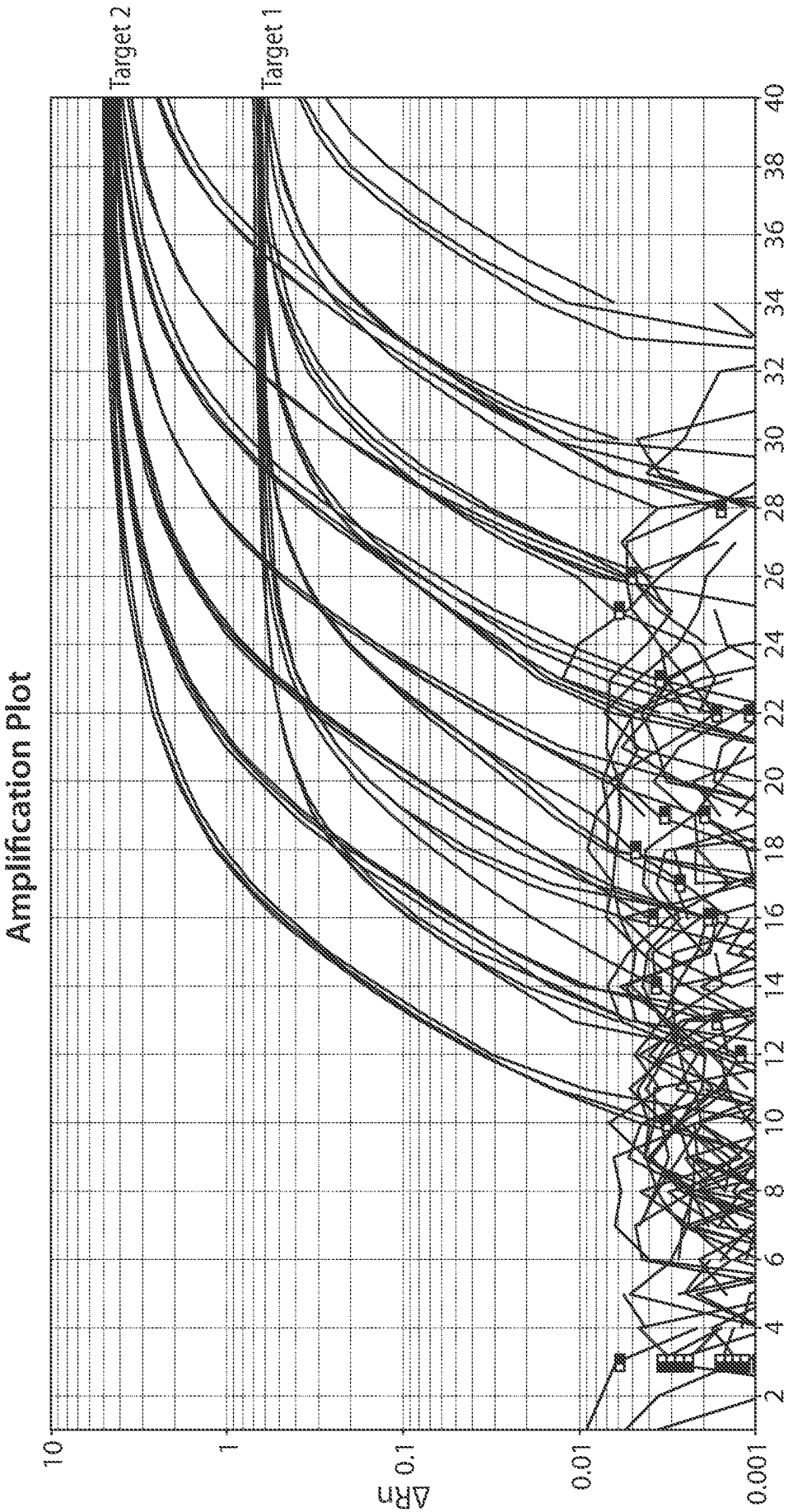

Master Mix #1
PCR Efficiency (%)
120
100
80
60
40
20
0
50nM 100nM 250nM 500nM
Primer Concentration (nM)
Master Mix #2
PCR Efficiency (%)
120
100
80
60
40
20
0
50nM 100nM 250nM 500nM
Primer Concentration (nM)

Figure 14



Target 1 : TTMV. Target 2: TTV.

Figure 15

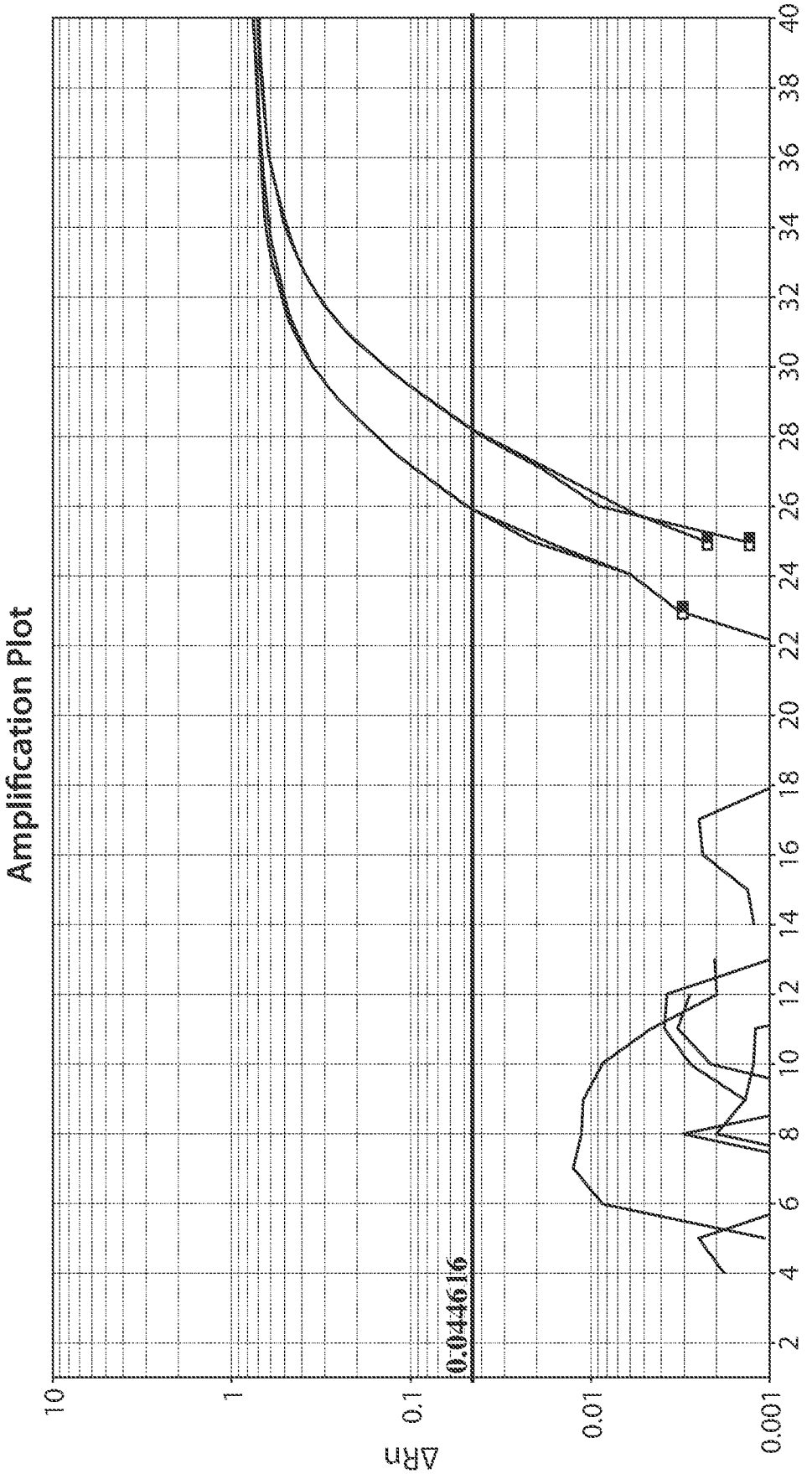

Amplification Plot
ΔRn
10
1
0.1
0.01
0.001
0.044616
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
Cycle
Target 1

Figure 16A

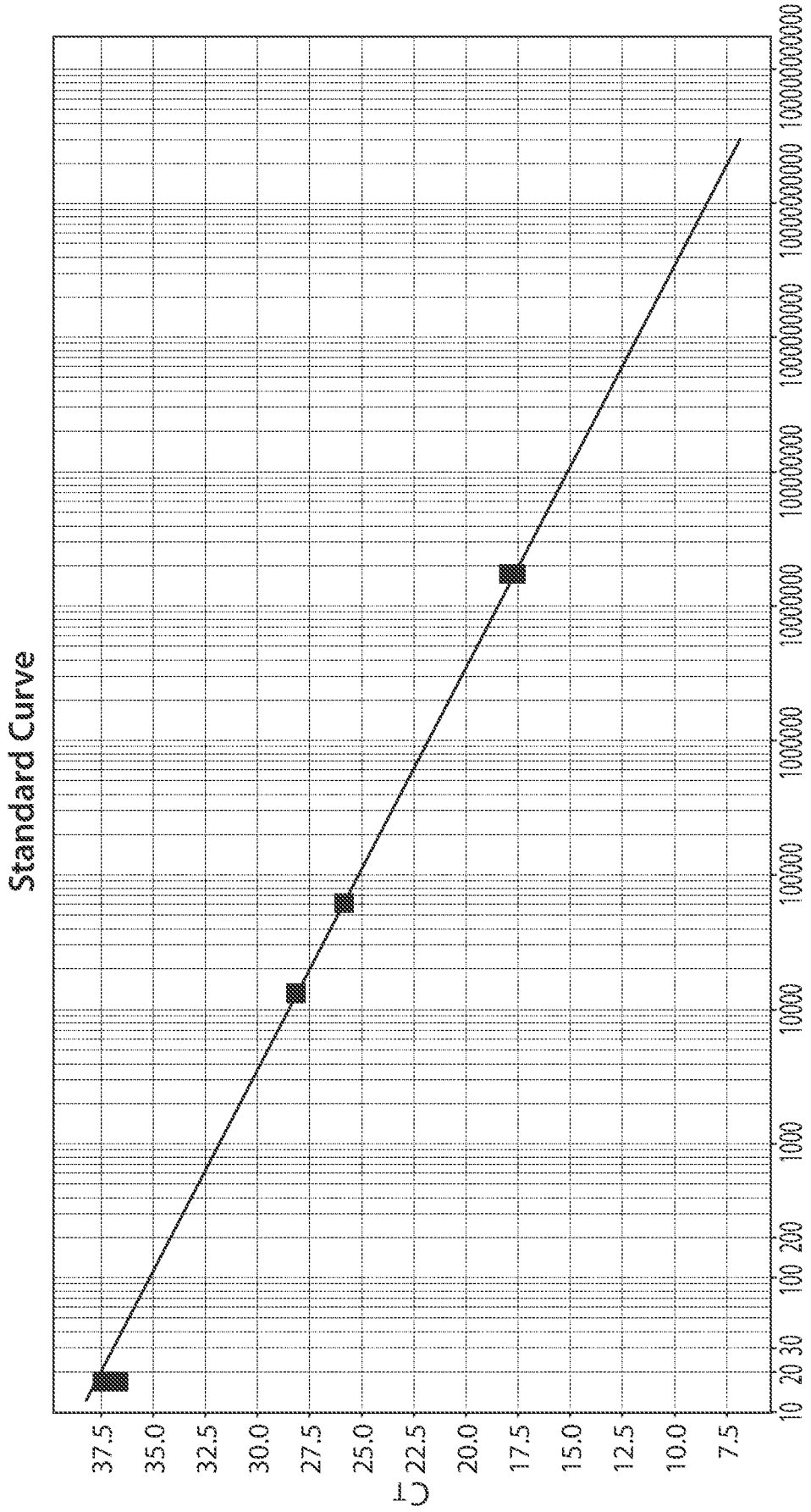

Standard Curve
Ct
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
10 20 30 100 200 1000 10000 100000 1000000 10000000 100000000 1000000000 10000000000
Quantity
Target 1

Figure 16B

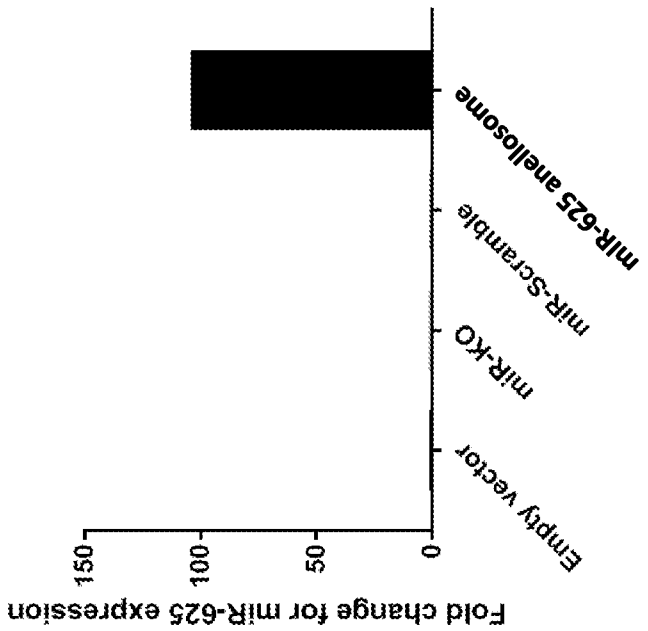

Fold change for miR-625 expression
150
100
50
0
Empty vector
miR-KO
miR-Scramble
miR-625 anellosome

FIG. 17

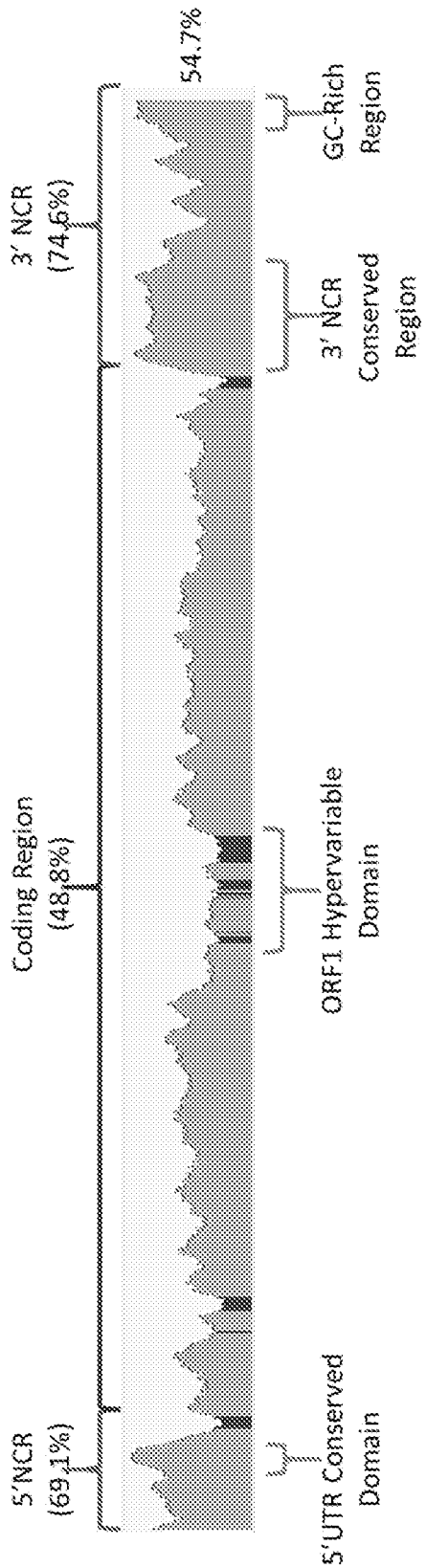

5'NCR
(69.1%)
5'UTR Conserved
Domain
Coding Region
(48.8%)
ORF1 Hypervariable
Domain
3' NCR
(74.6%)
3' NCR
Conserved
Region
54.7%
GC-Rich
Region

FIG. 18

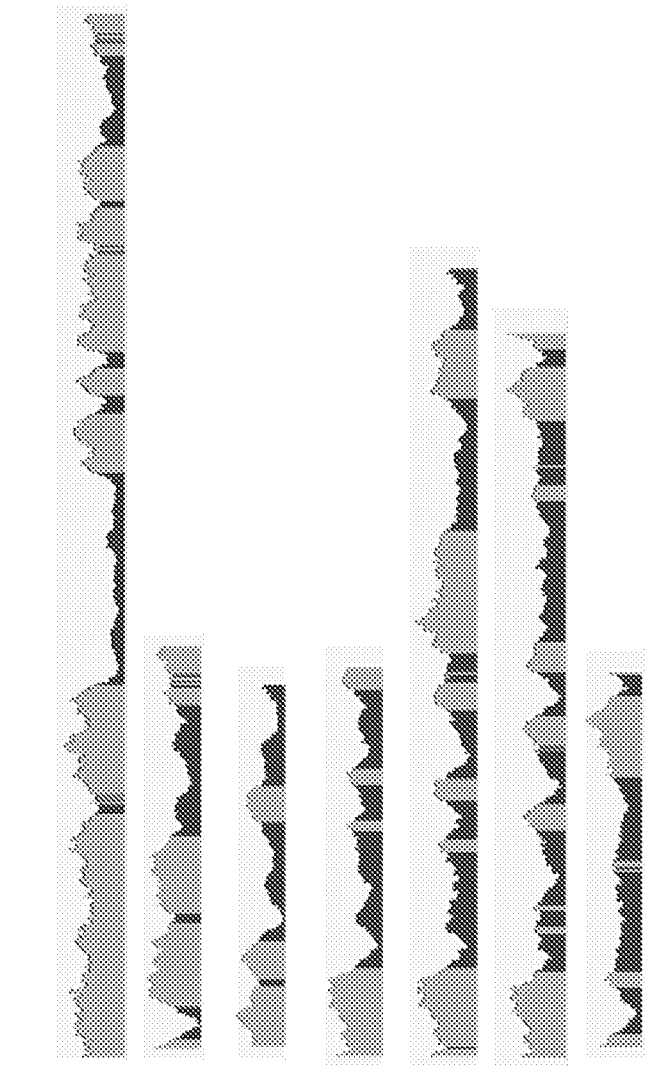

DNA
AA
ORF 1
49.6%
34.8%
ORF 1/1
49.6%
34.2%
ORF 1/2
49.4%
30.2%
ORF 2
47.8%
34.2%
ORF 2/2
48.9%
32.0%
ORF 2/3
48.8%
32.9%
ORF 2t/3*
48.9%
37.5%

FIG. 19

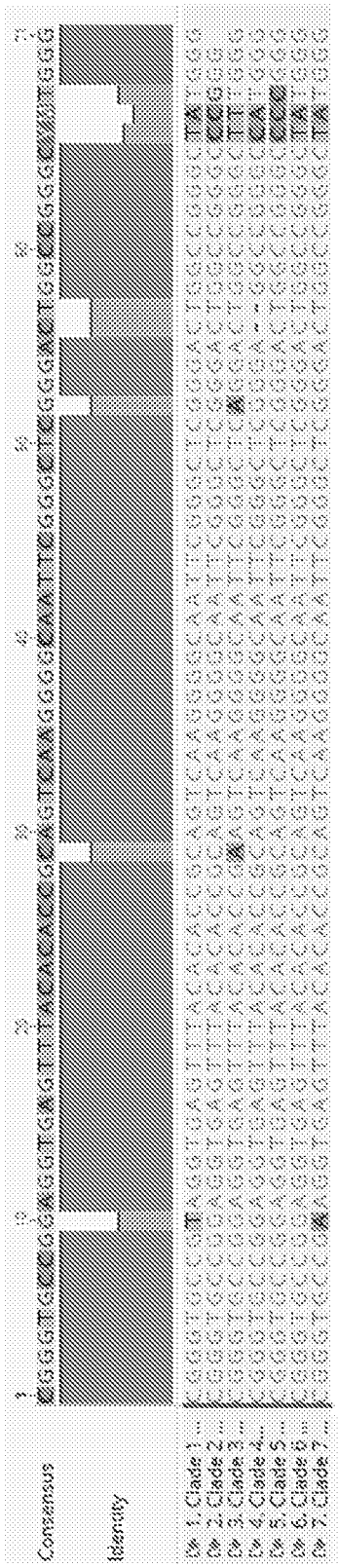

Consensus
Identity
1. Clade 1 ...
2. Clade 2 ...
3. Clade 3 ...
4. Clade 4 ...
5. Clade 5 ...
6. Clade 6 ...
7. Clade 7 ...

FIG. 20

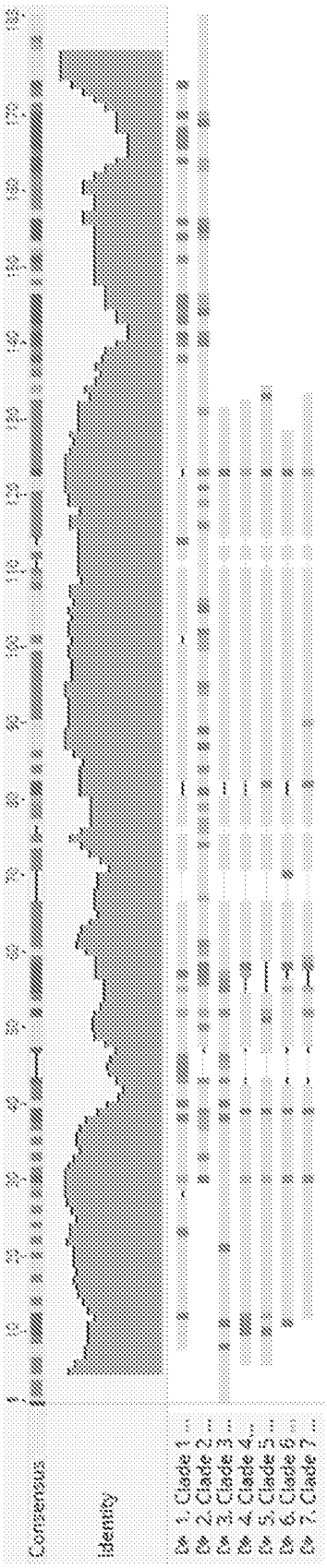

Consensus
Identity
1. Clade 1 ...
2. Clade 2 ...
3. Clade 3 ...
4. Clade 4 ...
5. Clade 5 ...
6. Clade 6 ...
7. Clade 7 ...

FIG. 21

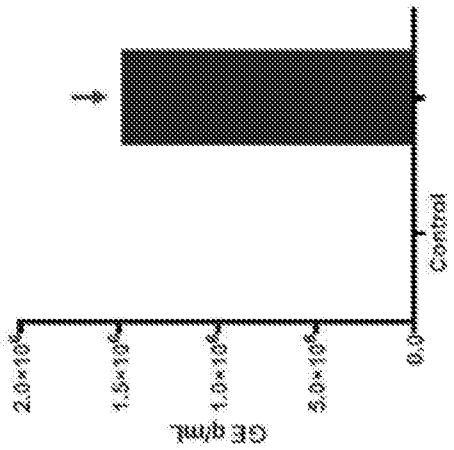


FIG. 22

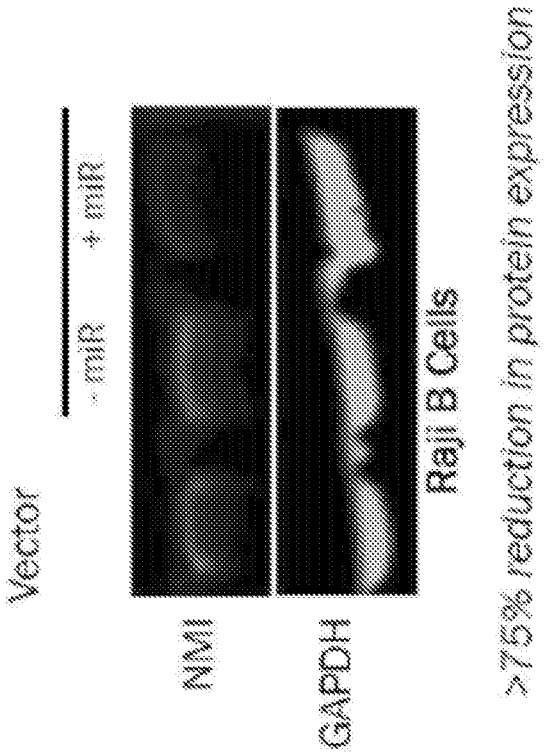

Vector
- miR
+ miR
NMI
GAPDH
Raji B Cells
>75% reduction in protein expression

FIG. 23

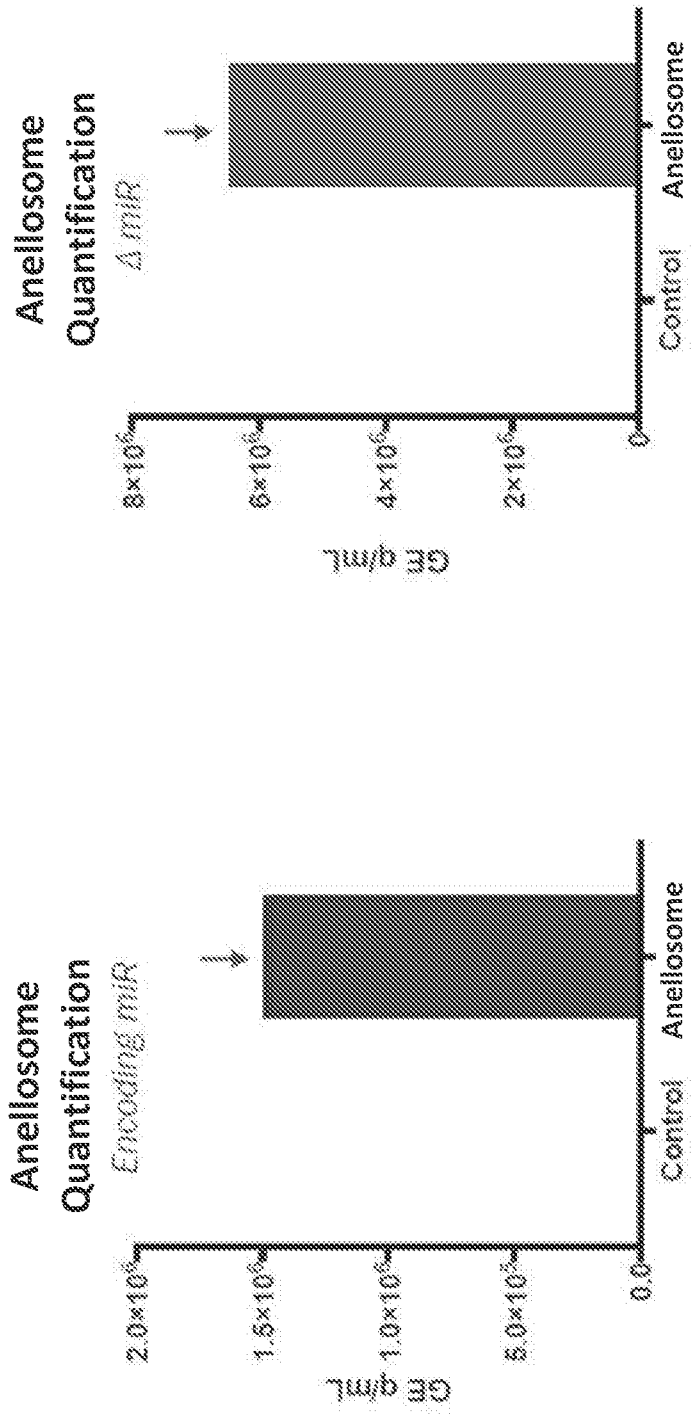

Anellosome Quantification
Δ miR
GE q/mL
8×10^6
6×10^6
4×10^6
2×10^6
0
Control
Anellosome
Anellosome Quantification
Encoding miR
GE q/mL
2.0×10^6
1.5×10^6
1.0×10^6
5.0×10^5
0.0
Control
Anellosome

FIG. 24

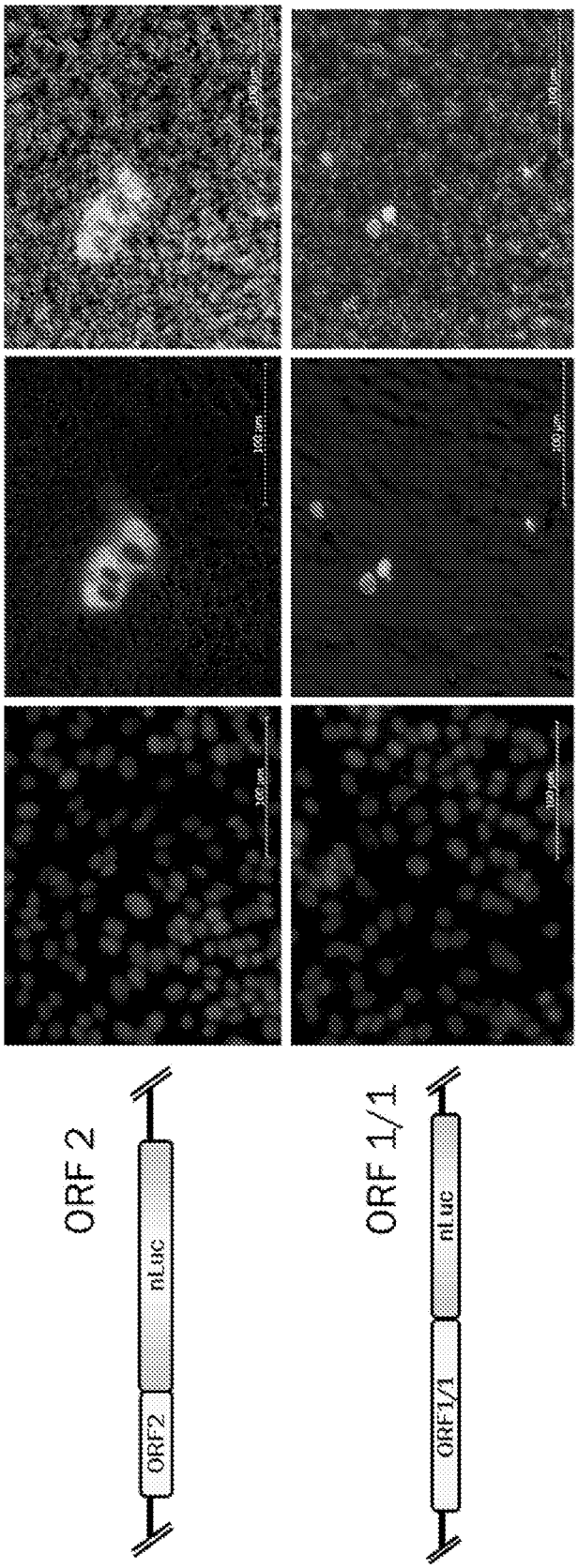


FIG. 25A

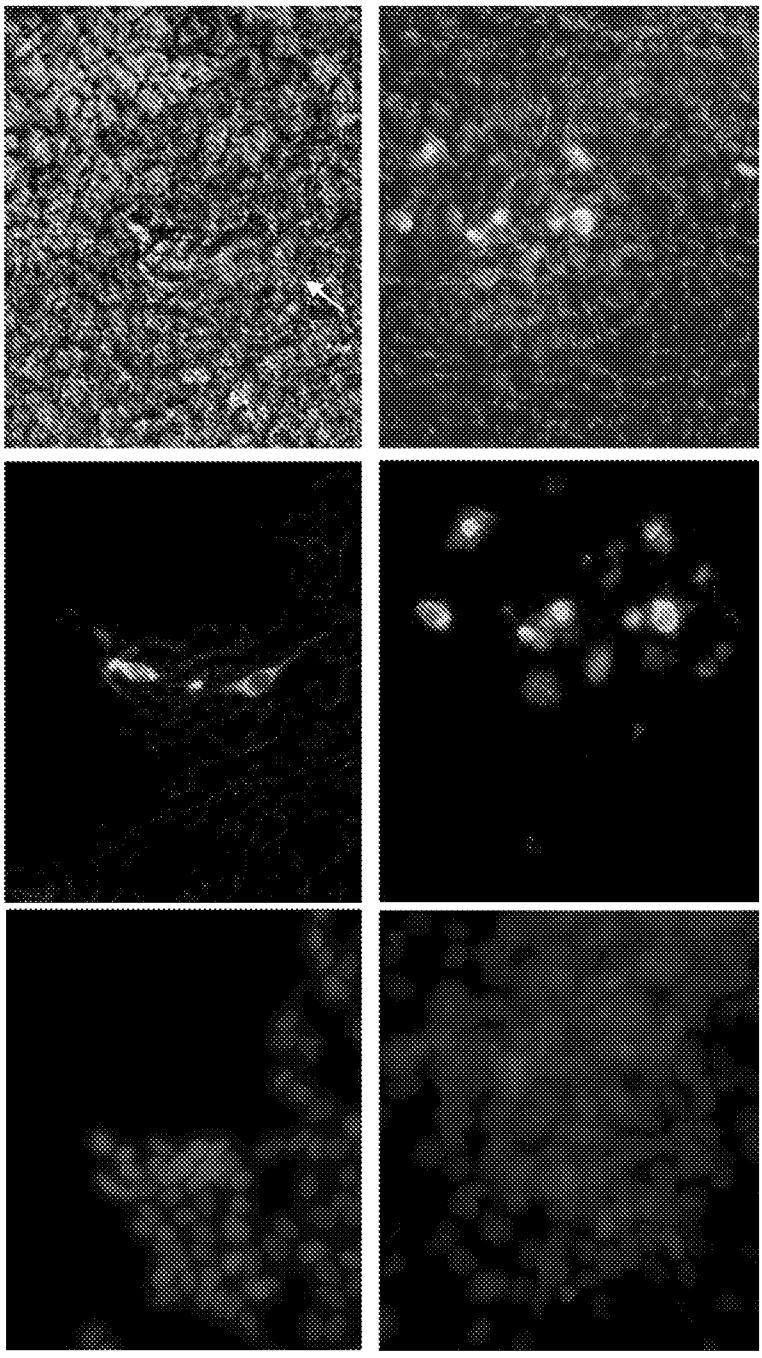


FIG. 25B

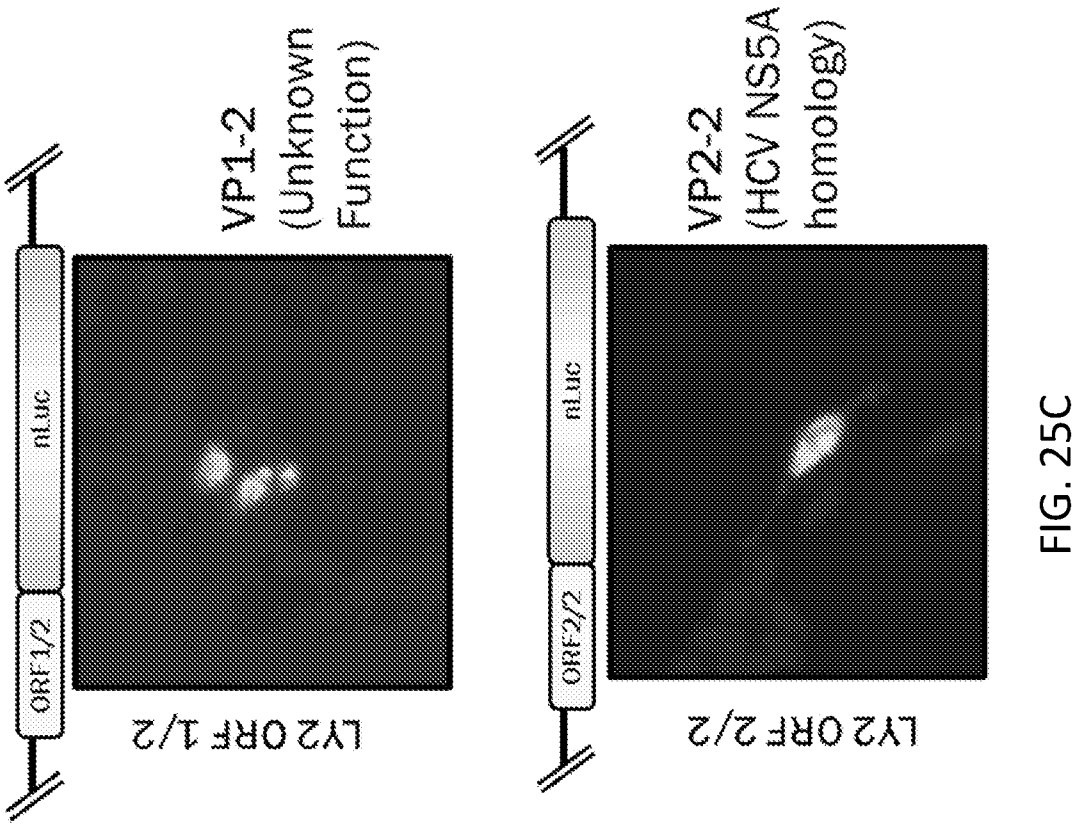

LY2 ORF 1/2
ORF1/2
nLuc
VP1-2
(Unknown
Function)
LY2 ORF 2/2
ORF2/2
nLuc
VP2-2
(HCV NS5A
homology)

FIG. 25C

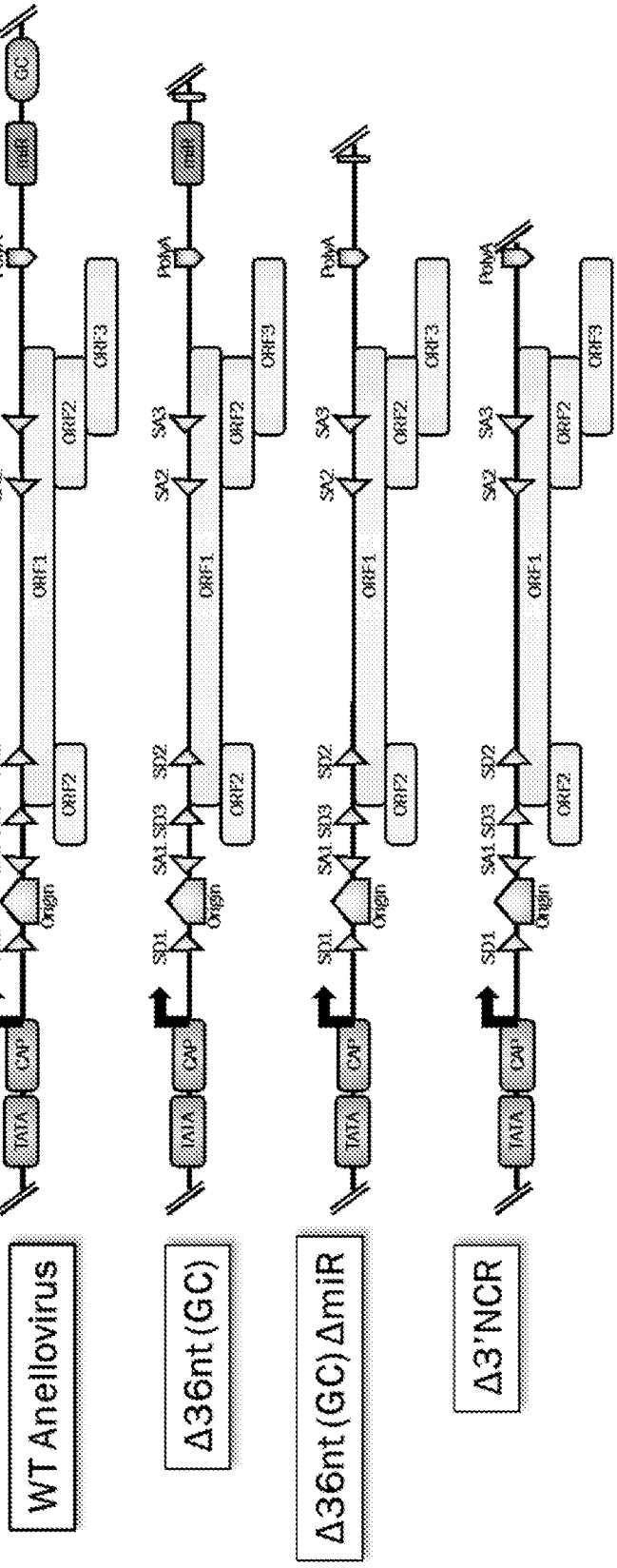

WT Anellovirus
Δ36nt (GC)
Δ36nt (GC) ΔmiR
Δ3'NCR
TATA
CAP
SD1
Origin
SA1
SD3
SD2
ORF2
SA2
ORF1
SA3
ORF3
PolyA
GC

FIG. 26

FIG. 27A

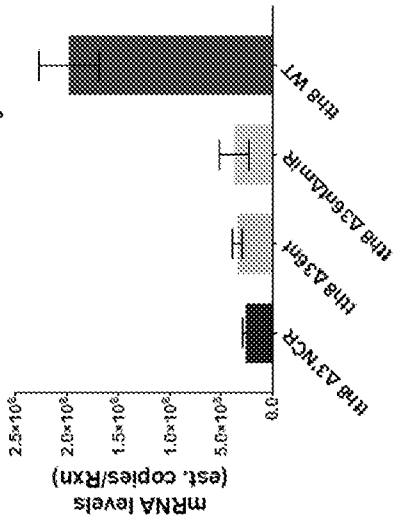


FIG. 27B

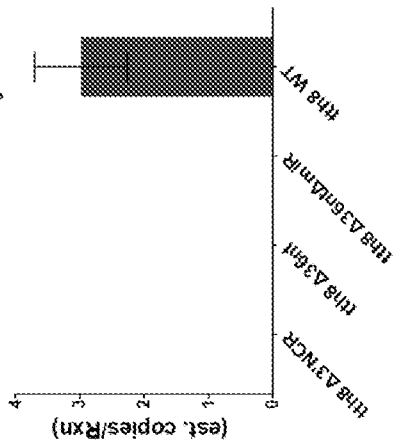


FIG. 27C

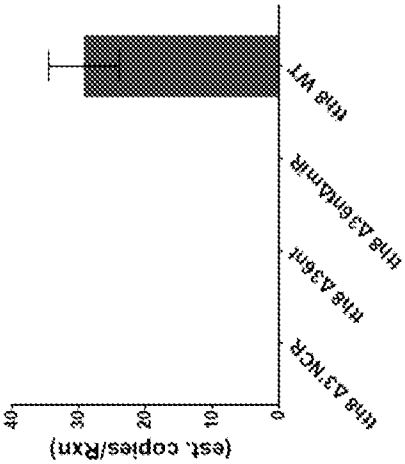
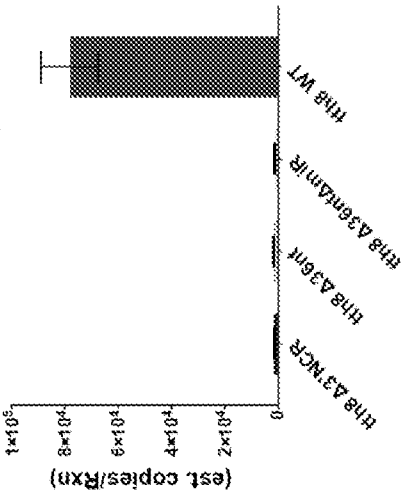


FIG. 27D



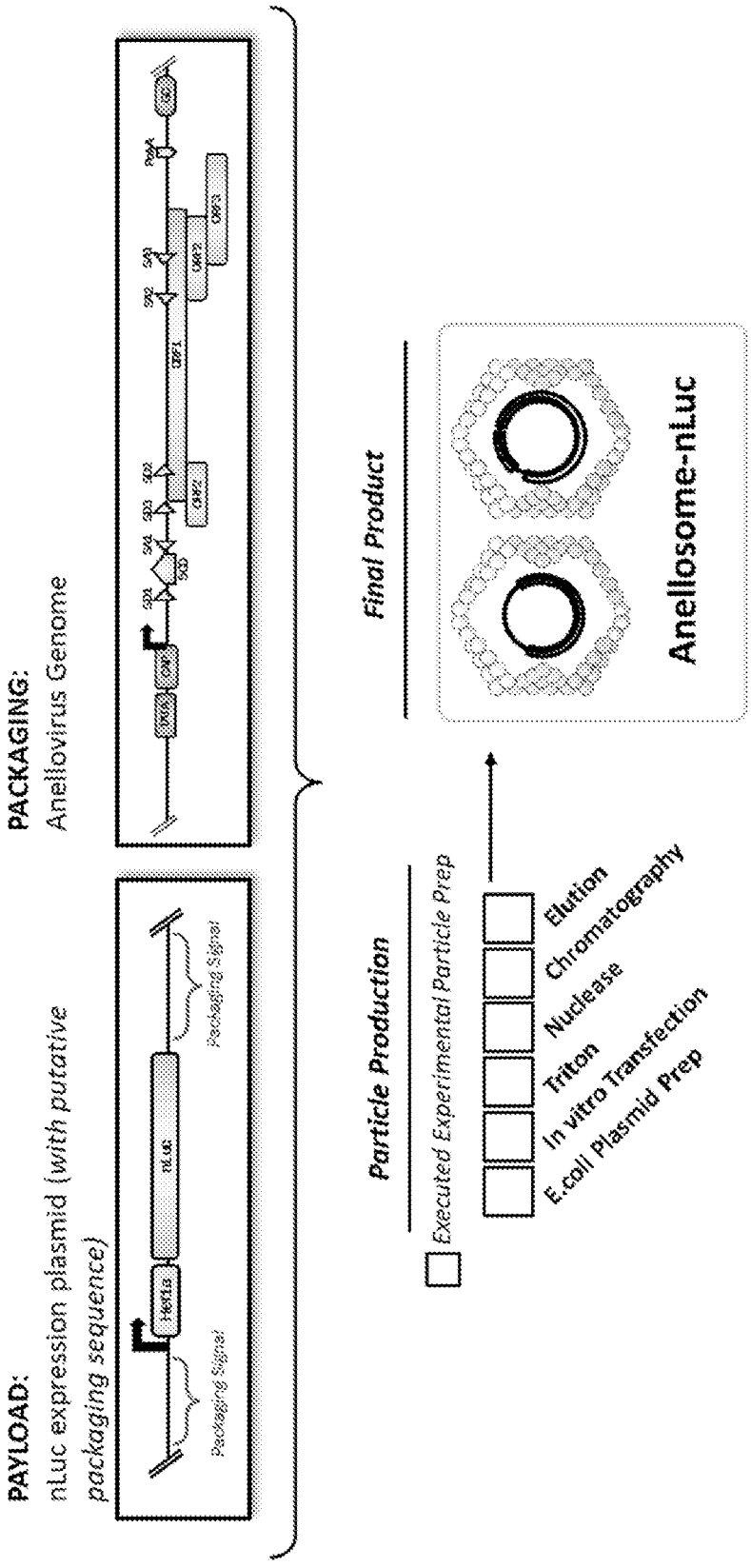

PAYLOAD:
nLuc expression plasmid (with putative packaging sequence)
Packaging Signal
Packaging Signal
PACKAGING:
Anellovirus Genome
Particle Production
Executed Experimental Particle Prep
Elution Chromatography
Nuclease
Triton
In vitro Transfection
E.coli Plasmid Prep
Final Product
Anellosome-nLuc

FIG. 28A

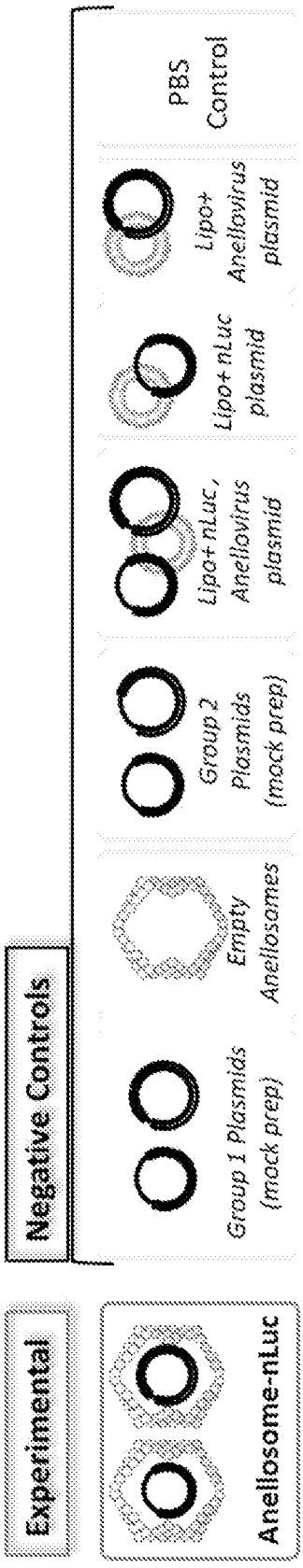

Experimental
Anellosome-nLuc
Negative Controls
Group 1 Plasmids (mock prep)
Empty Anellosomes
Group 2 Plasmids (mock prep)
Lipo+ nLuc, Anellovirus plasmid
Lipo+ nLuc plasmid
Lipo+ Anellovirus plasmid
PBS Control

FIG. 28B

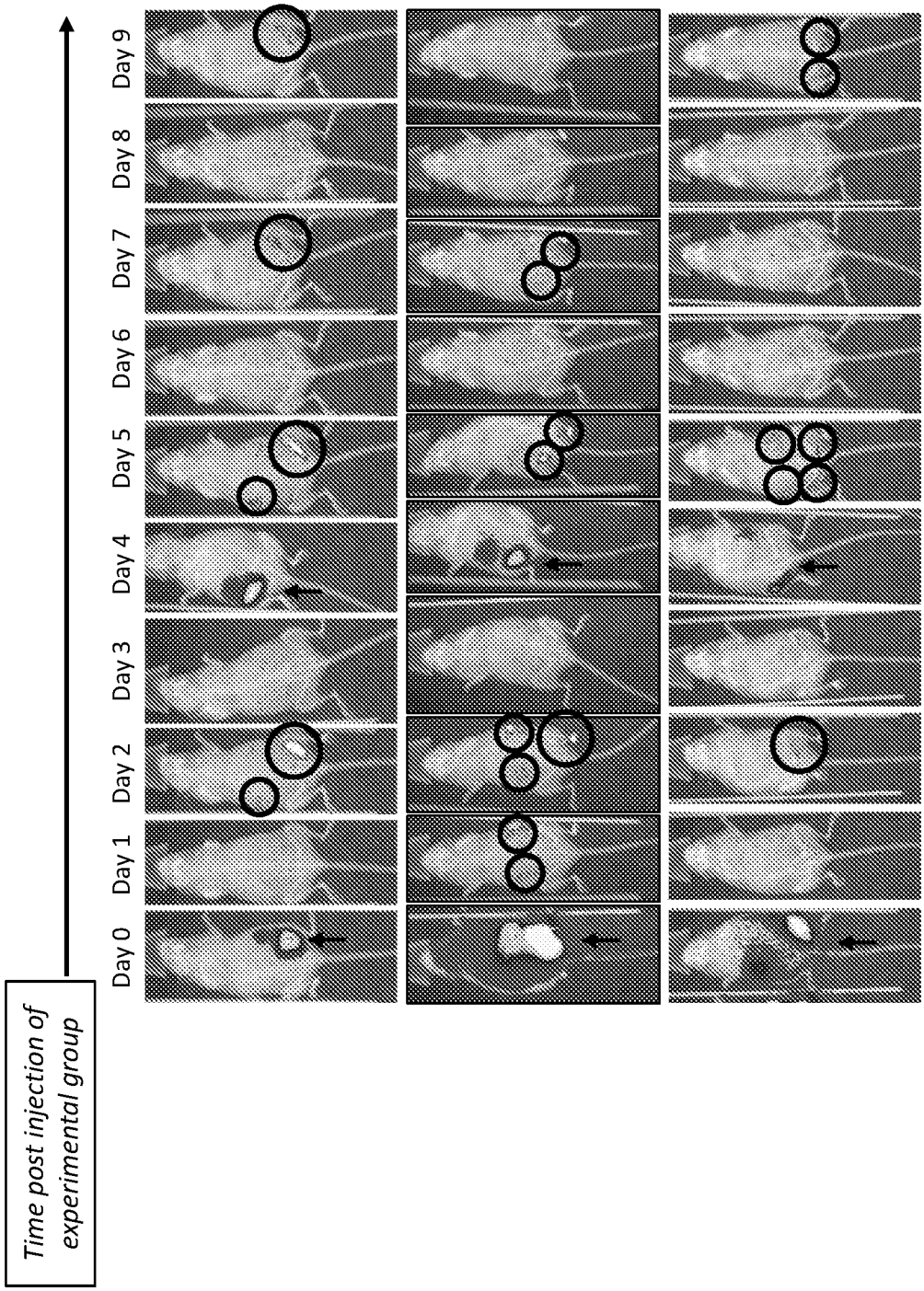

Time post injection of experimental group
Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Day 9

FIG. 29A

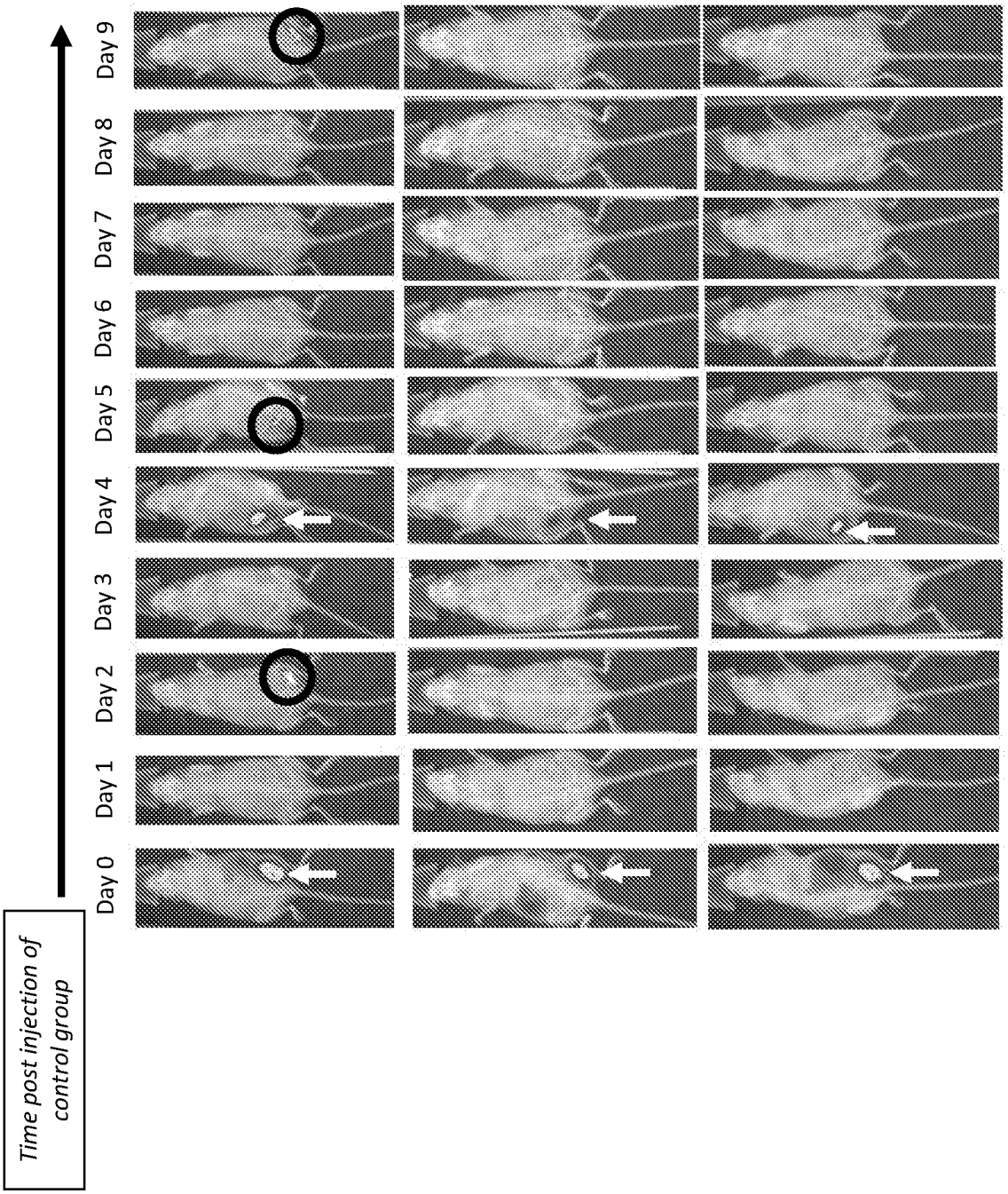

Time post injection of control group
Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Day 9

FIG. 29B

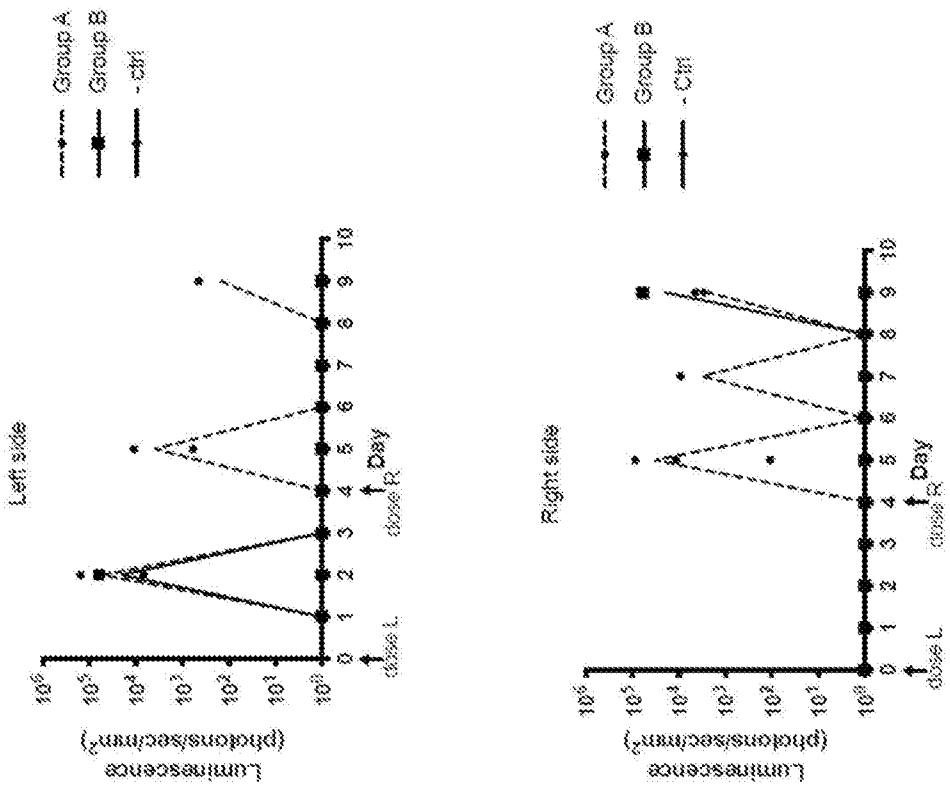

Left side
Right side
Group A
Group B
- Ctrl
Luminescence
(photons/sec/mm²)
Day
dose L
dose R

FIG. 29C

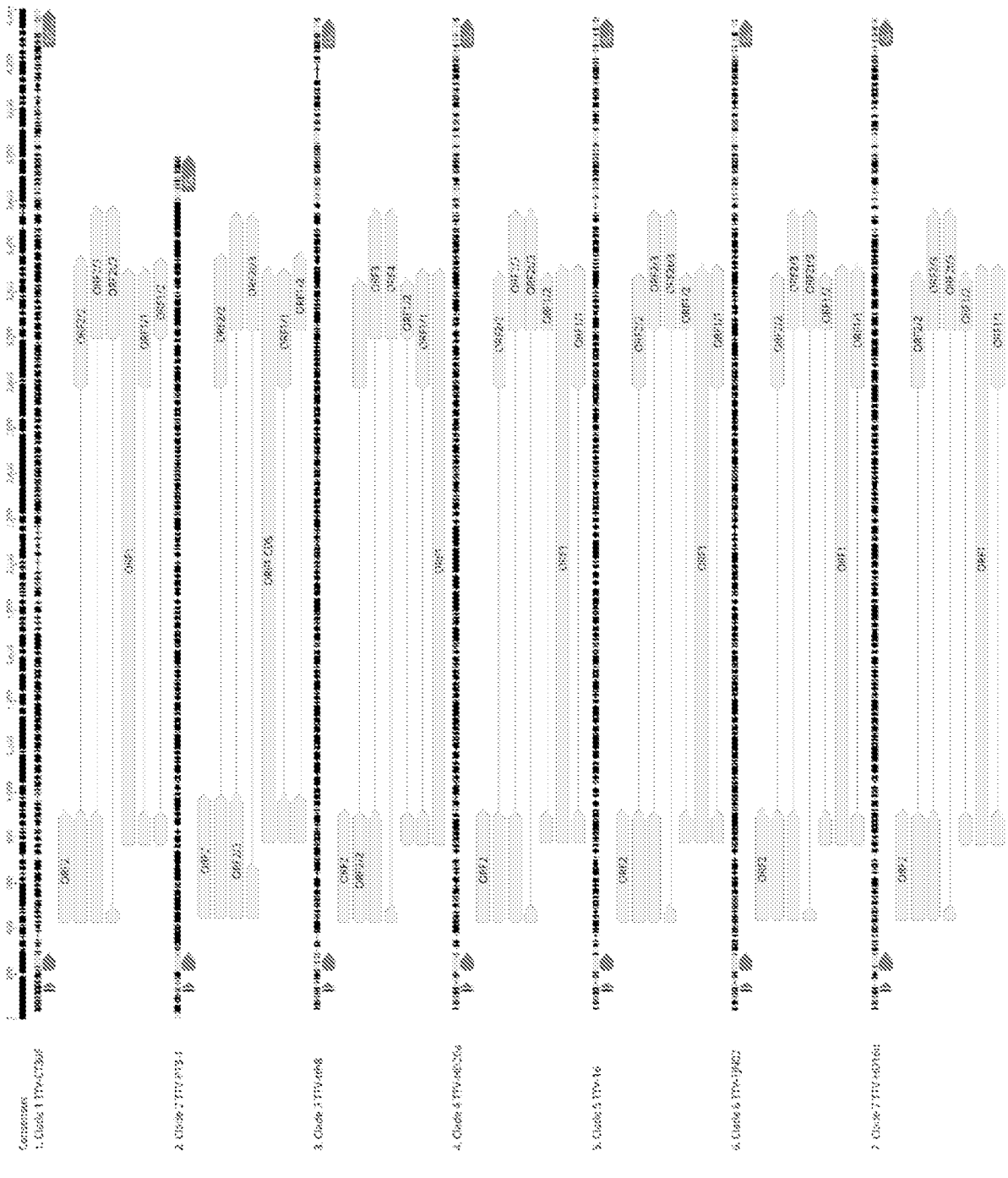


FIG. 29D

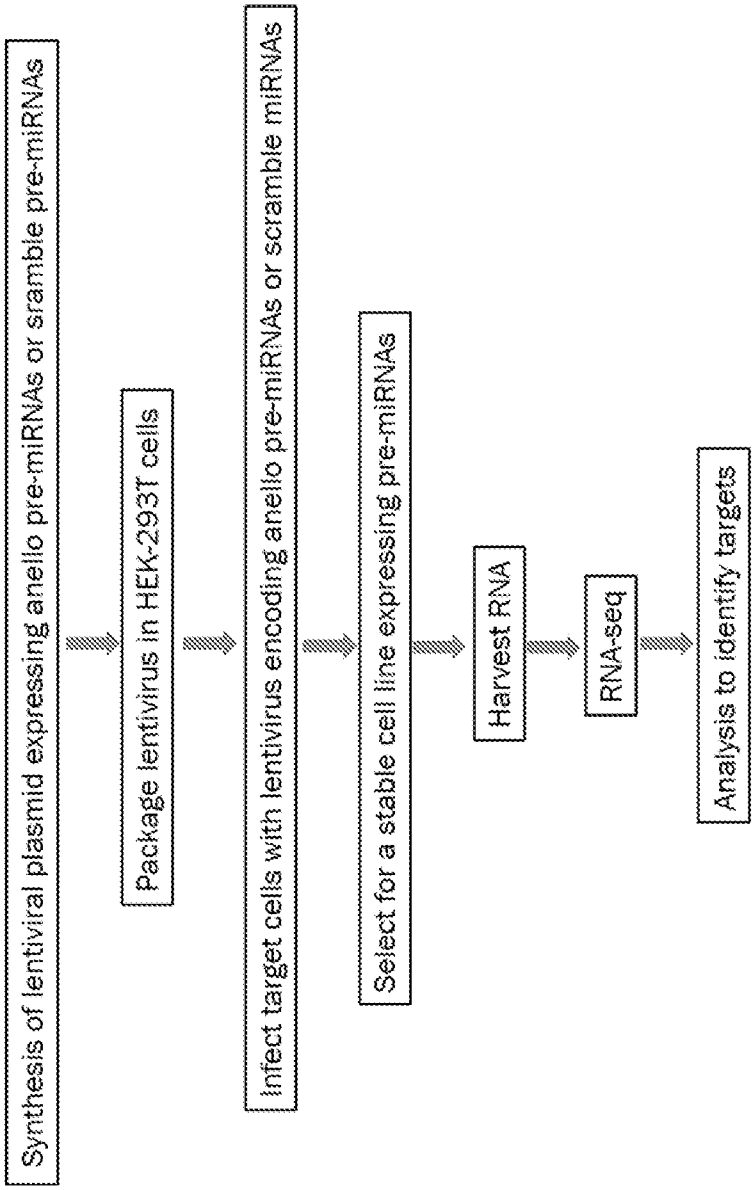

Synthesis of lentiviral plasmid expressing anello pre-miRNAs or srambie pre-miRNAs
Package lentivirus in HEK-293T cells
Infect target cells with lentivirus encoding anello pre-miRNAs or scramble miRNAs
Select for a stable cell line expressing pre-miRNAs
Harvest RNA
RNA-seq
Analysis to identify targets

FIG. 30

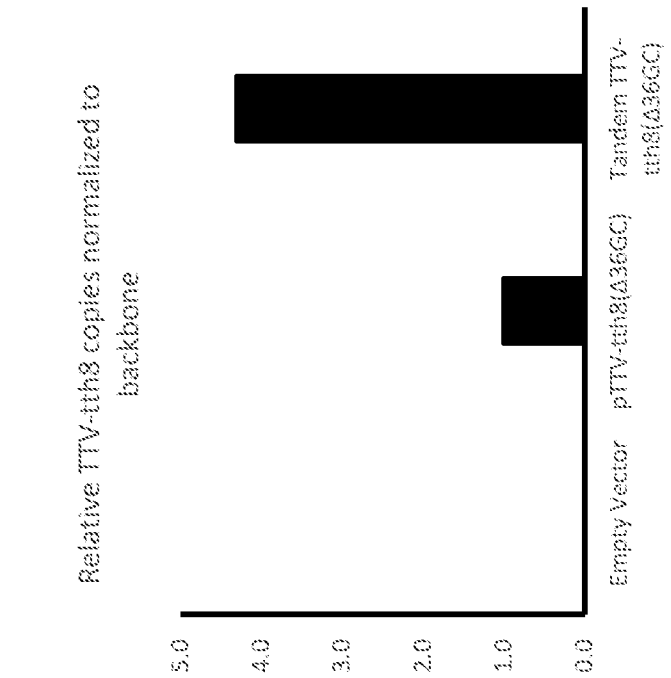


FIG. 31B

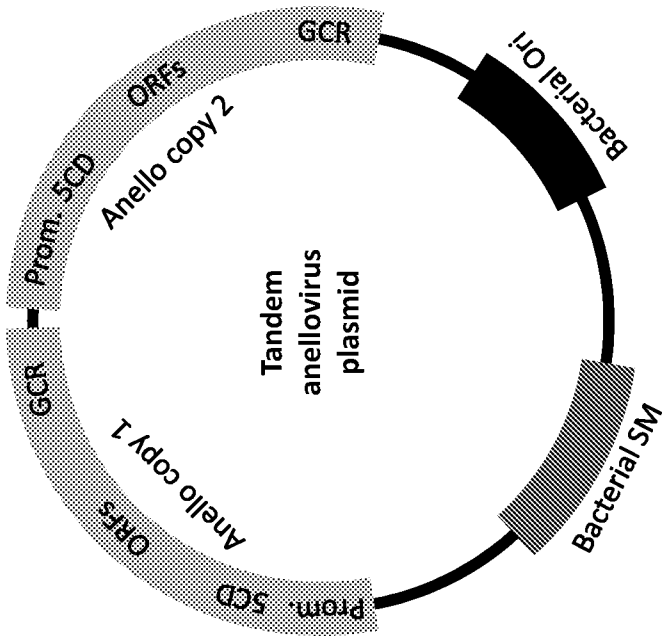


FIG. 31A

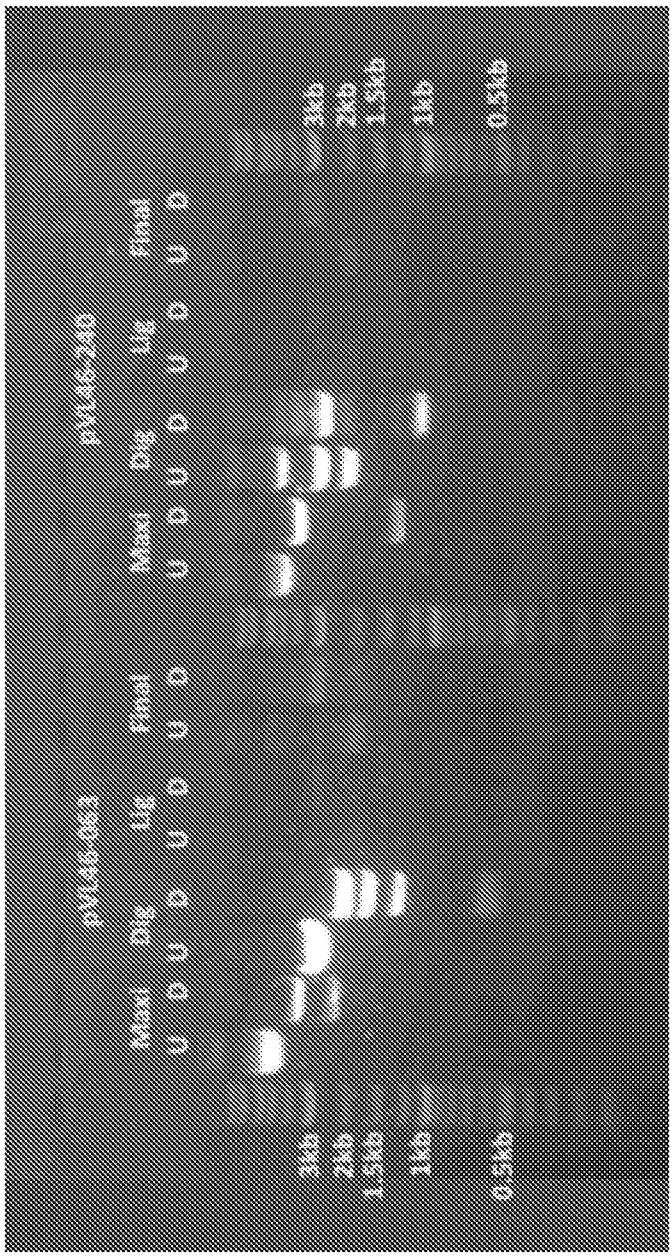


FIG. 31C

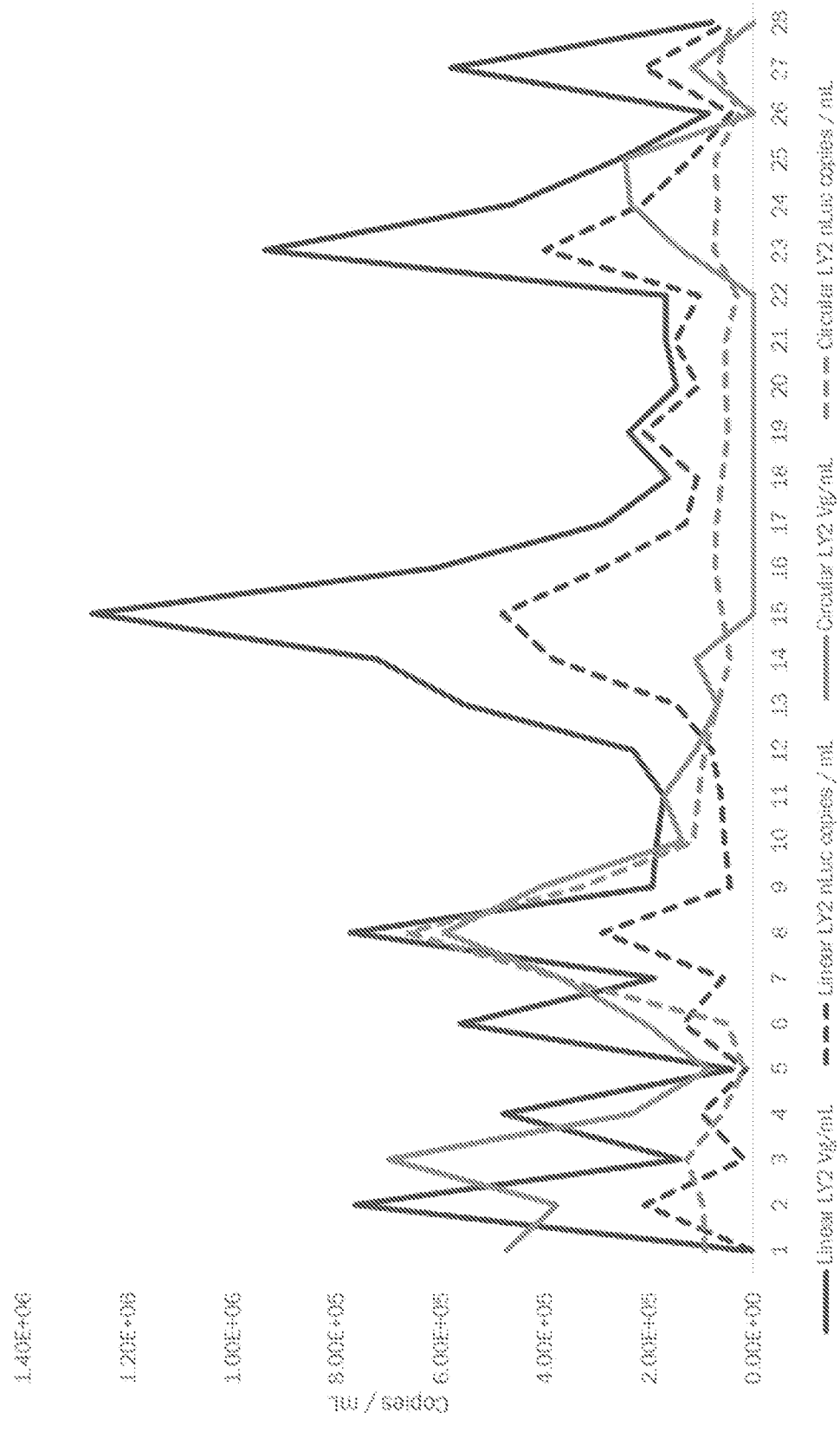

1.40E+06
1.20E+06
1.00E+06
8.00E+05
6.00E+05
4.00E+05
2.00E+05
0.00E+00
copies / mL
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Linear LY2 Vg/mL
Linear LY2 nluc copies / mL
Circular LY2 Vg/mL
Circular LY2 nluc copies / mL

FIG. 31D

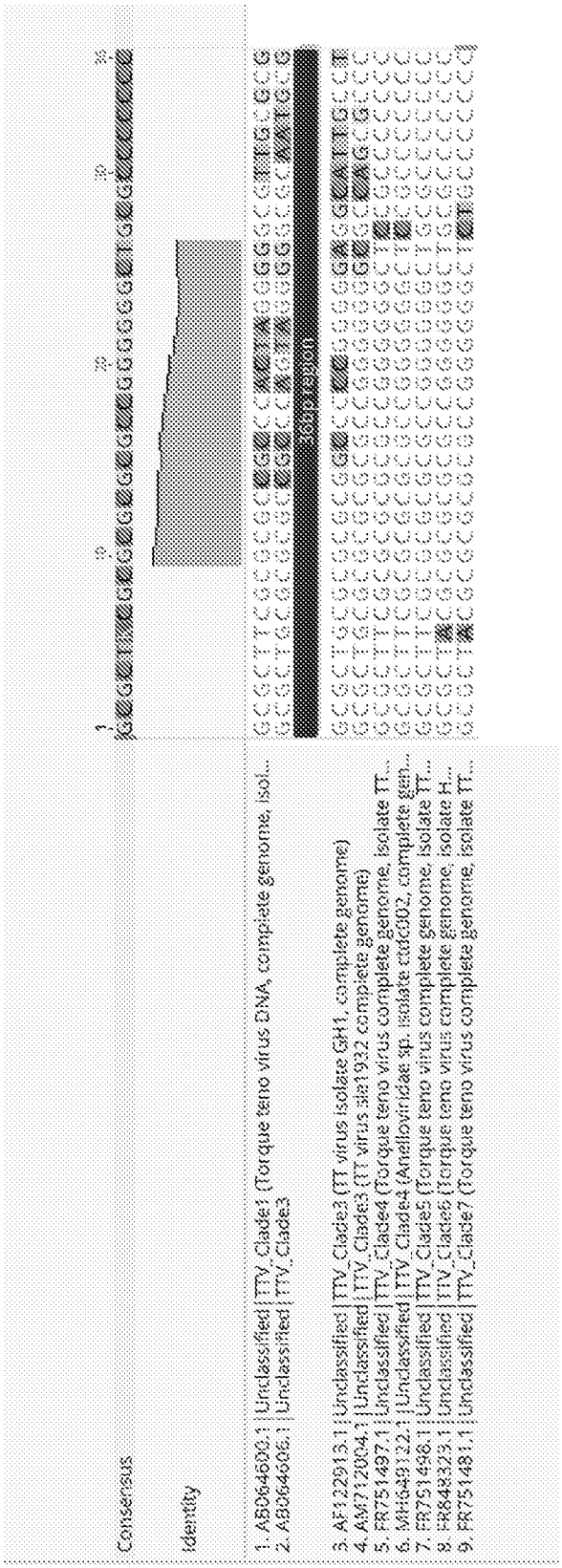

Consensus
Identity
1. AB064600.1 | Unclassified | TTV_Clade1 (Torque teno virus DNA, complete genome, isol...
2. AB064606.1 | Unclassified | TTV_Clade3
3. AF122913.1 | Unclassified | TTV_Clade3 (TT virus isolate GH1, complete genome)
4. AM712004.1 | Unclassified | TTV_Clade3 (TT virus sle1932 complete genome)
5. FR751497.1 | Unclassified | TTV_Clade4 (Torque teno virus complete genome, isolate TT...
6. MH649122.1 | Unclassified | TTV_Clade4 (Anelloviridae sp. isolate ctdc002, complete gen...
7. FR751496.1 | Unclassified | TTV_Clade5 (Torque teno virus complete genome, isolate TT...
8. FR648323.1 | Unclassified | TTV_Clade6 (Torque teno virus complete genome, isolate H...
9. FR751481.1 | Unclassified | TTV_Clade7 (Torque teno virus complete genome, isolate TT...

FIG. 32

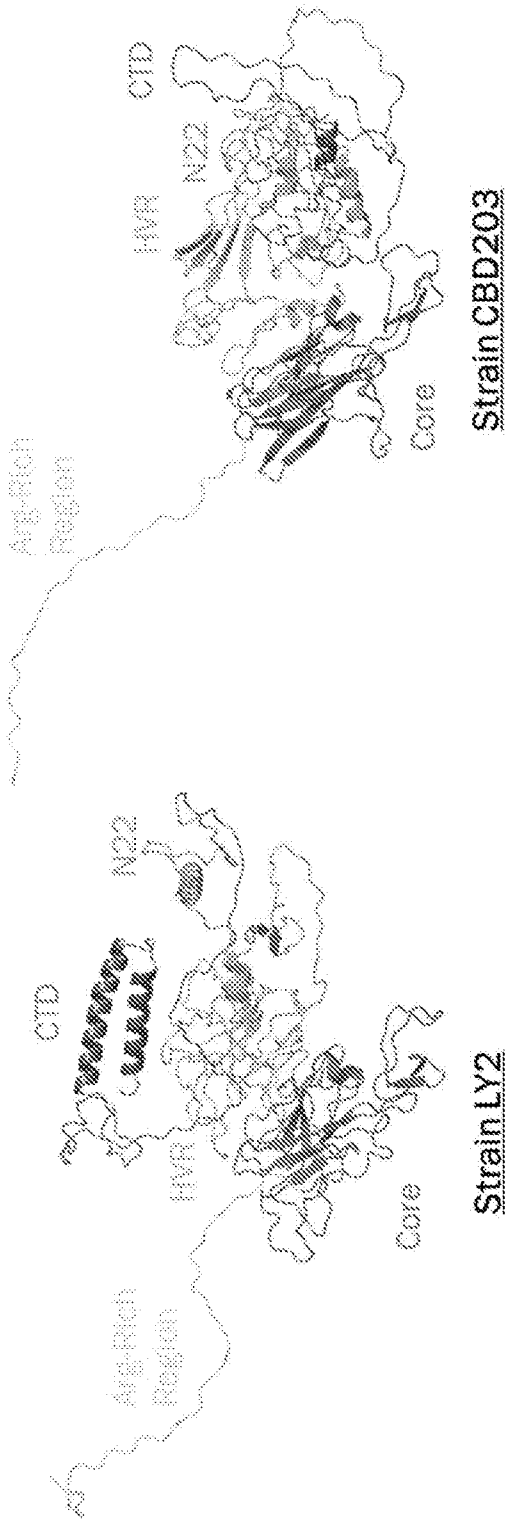

CTD
N22
HVR
Core
Arg-Rich Region
Strain CBD203
CTD
N22
HVR
Core
Arg-Rich Region
Strain LY2

FIG. 33

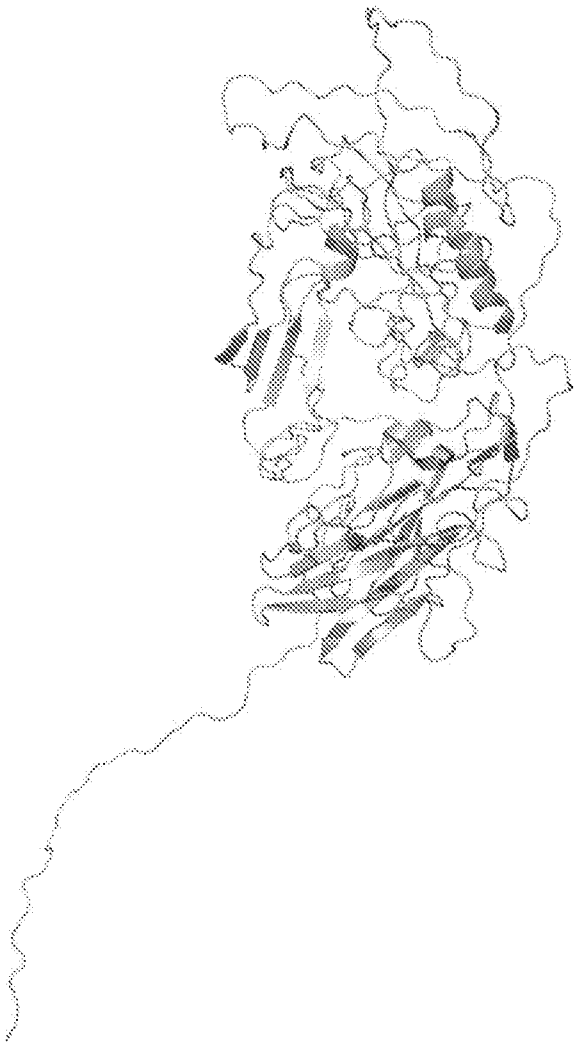


FIG. 34

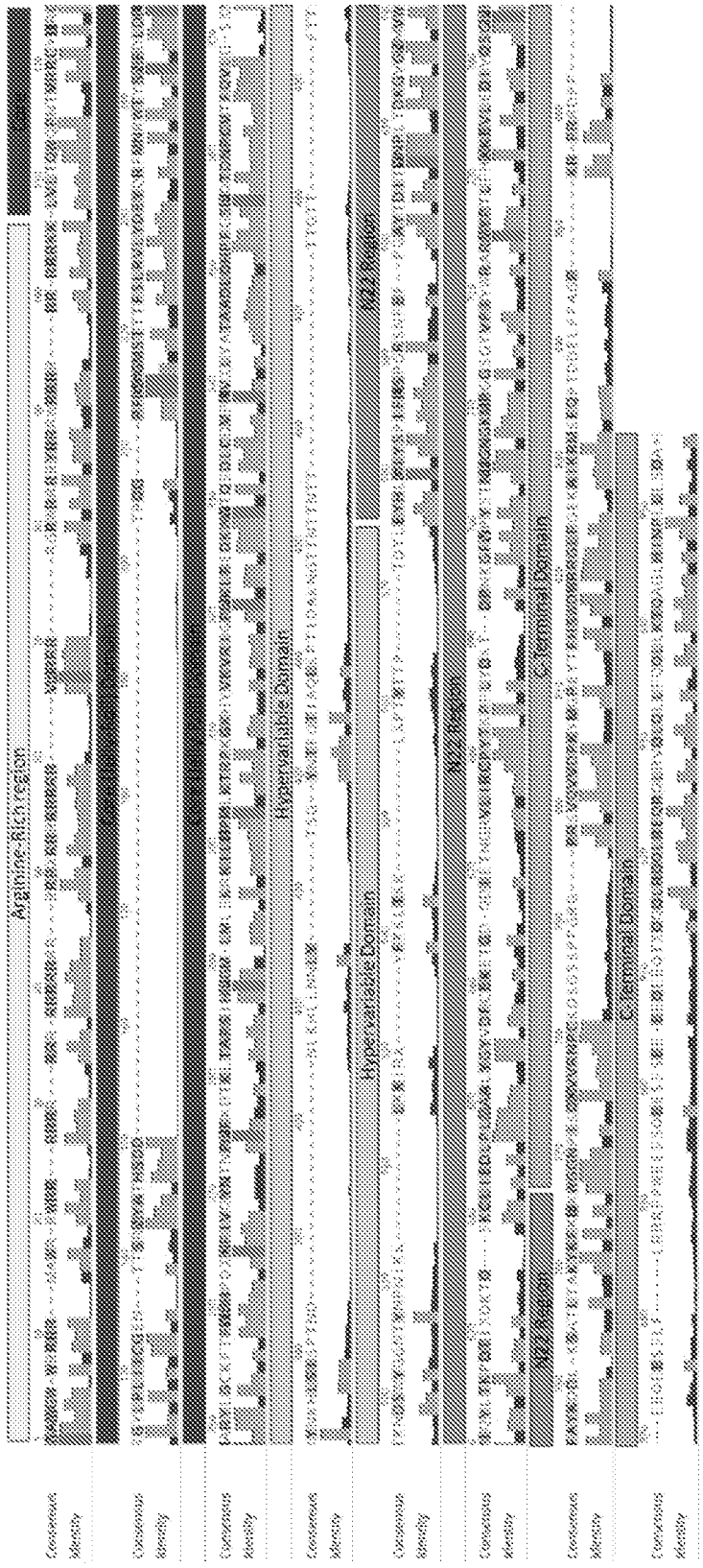

Arginine-Rich region
Hypervariable Domain
Hypervariable Domain
N-Terminal Domain
C-terminal Domain
C-terminal Domain

FIG. 35

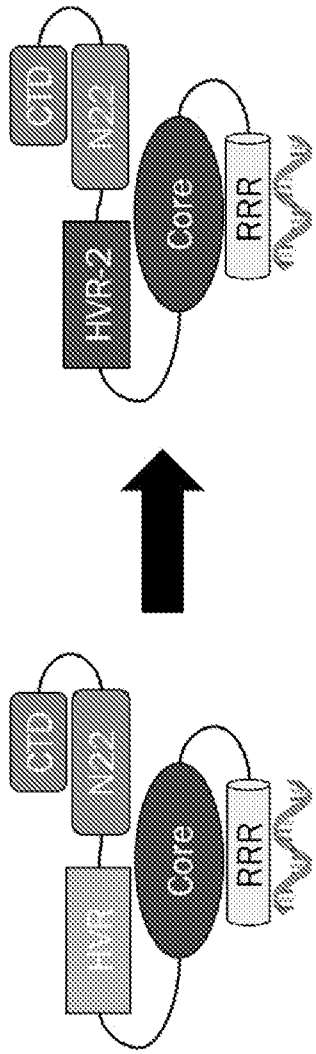


FIG. 36

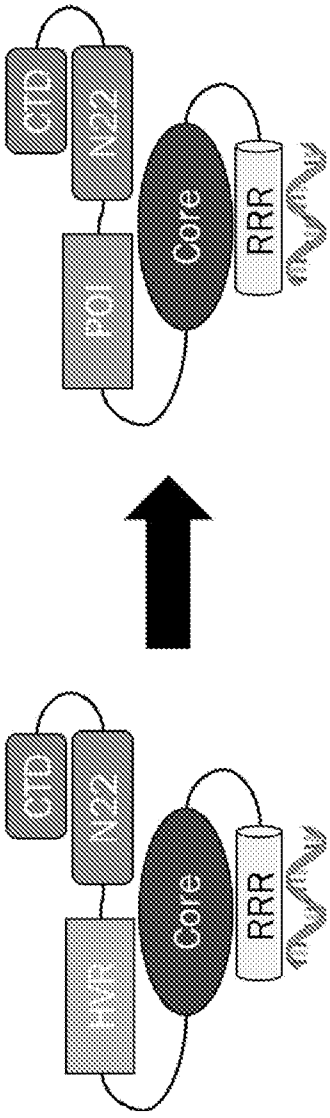


FIG. 37

FIG. 38

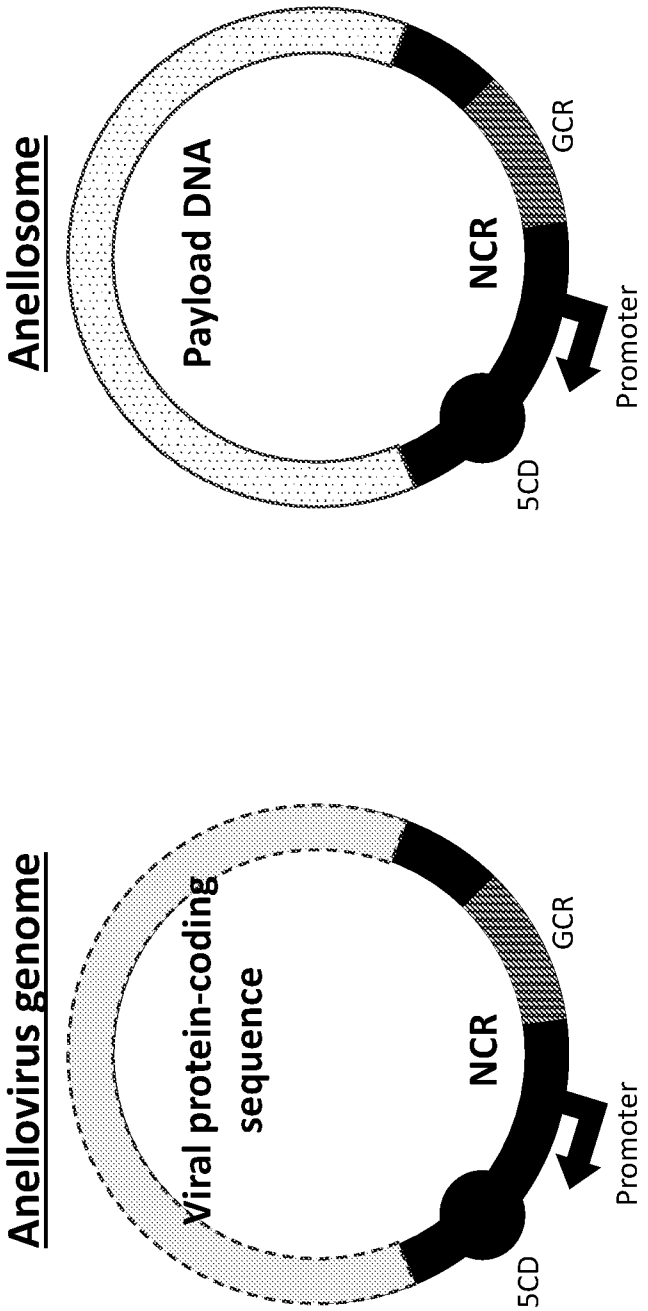

Anellosome
Payload DNA
NCR
GCR
Promoter
5CD
Anellovirus genome
Viral protein-coding sequence
NCR
GCR
Promoter
5CD

FIG. 39

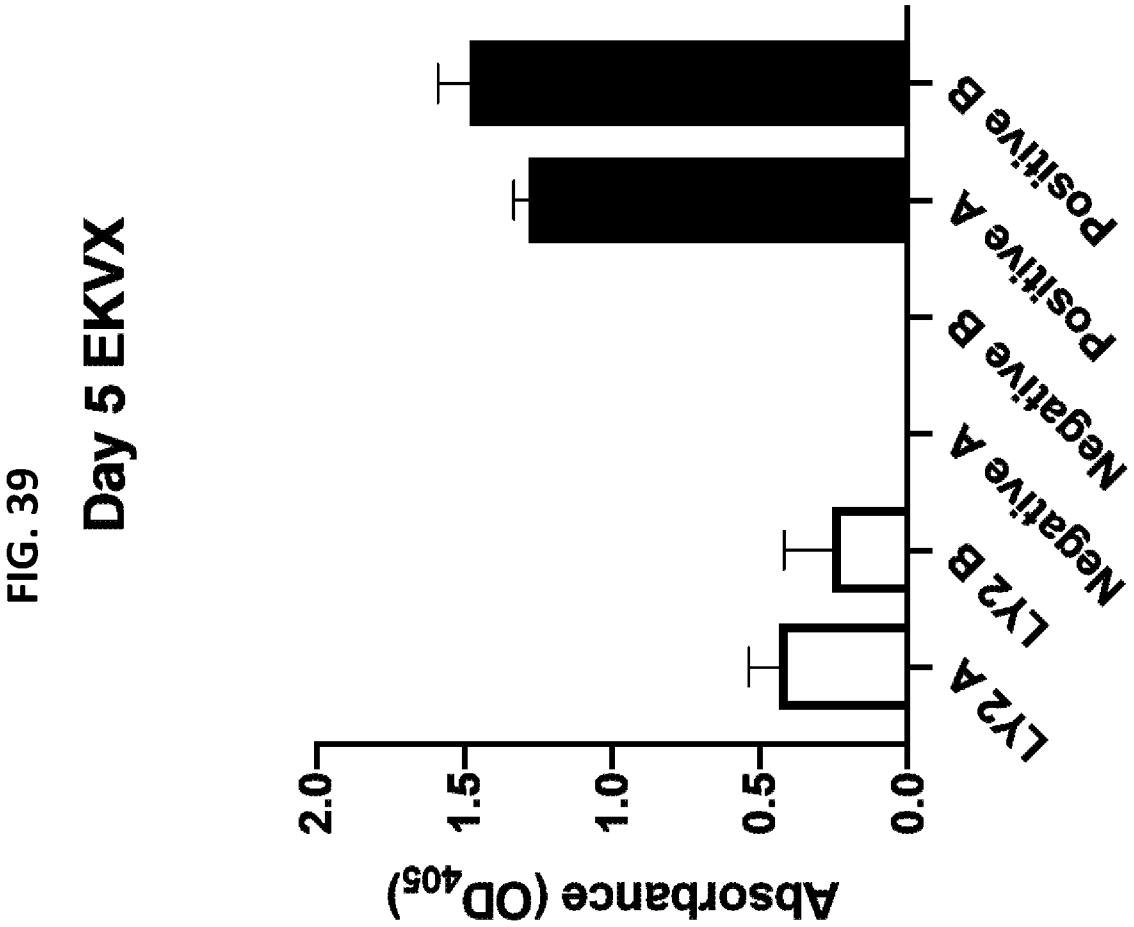

Day 5 EKVX
Absorbance (OD_{405})
2.0
1.5
1.0
0.5
0.0
LY2 A
LY2 B
Negative A
Negative B
Positive A
Positive B

FIG. 40

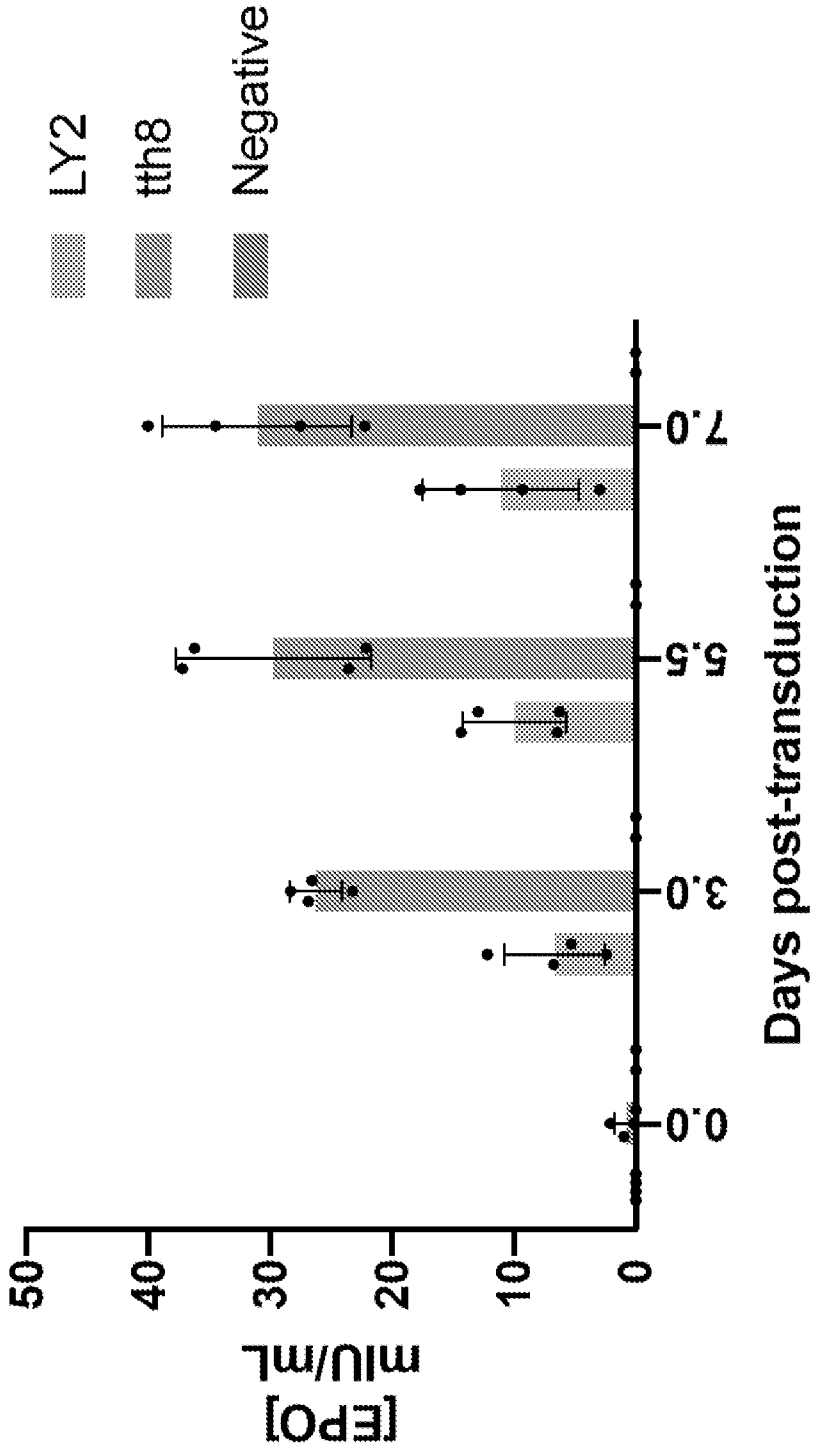

hEpo
LY2
tth8
Negative
[EPO] mIU/mL
0
10
20
30
40
50
0.0
3.0
5.5
7.0
Days post-transduction

FIG. 41A

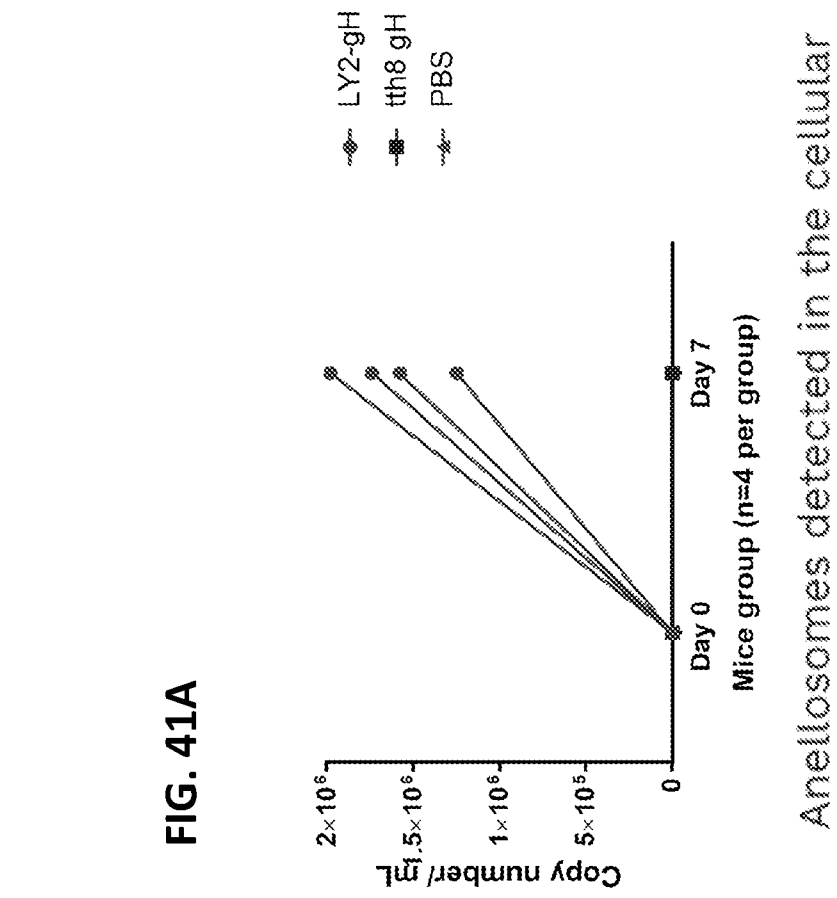


FIG. 41B

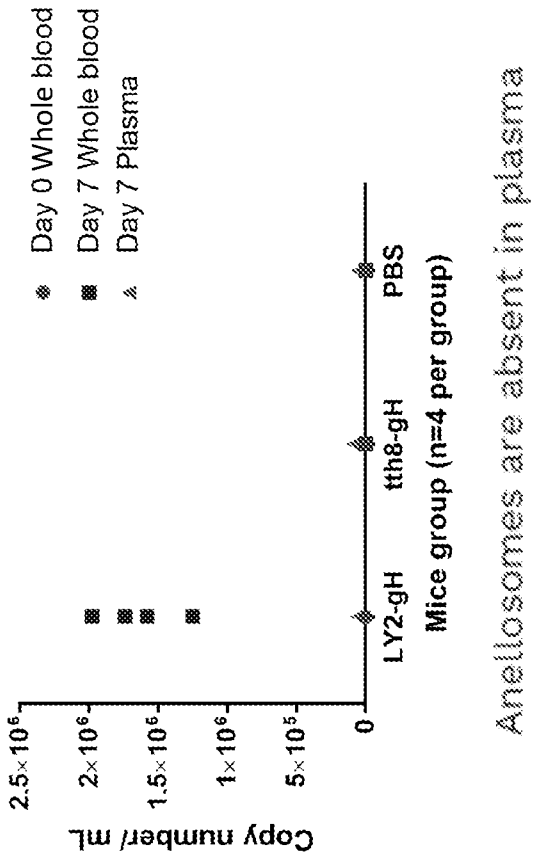


FIG. 42

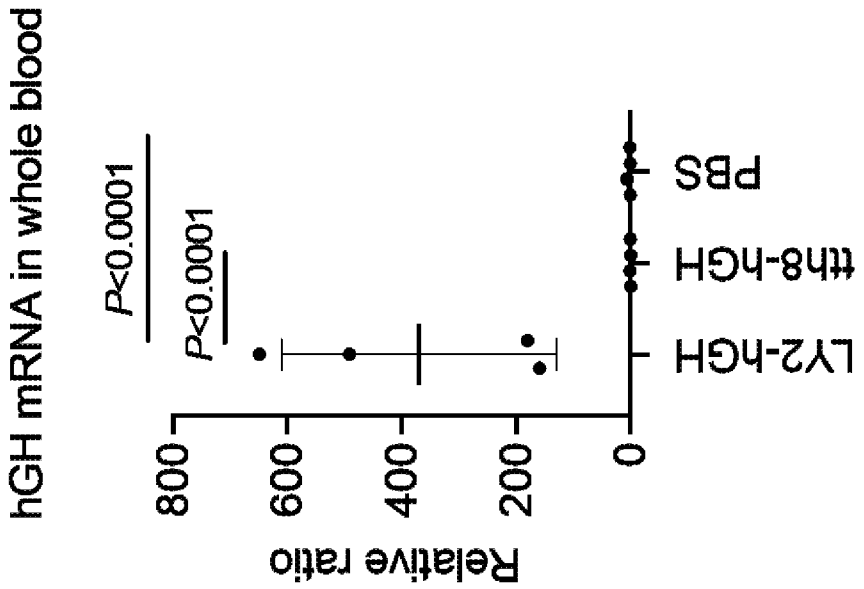

hGH mRNA in whole blood
P<0.0001
P<0.0001
Relative ratio
800
600
400
200
0
LY2-hGH
ttn8-hGH
PBS

FIG. 43A

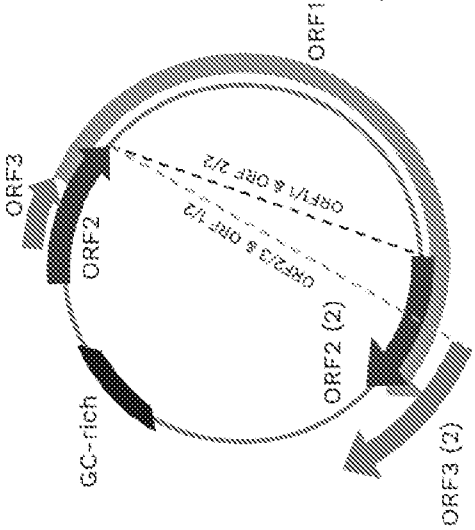


FIG. 43B

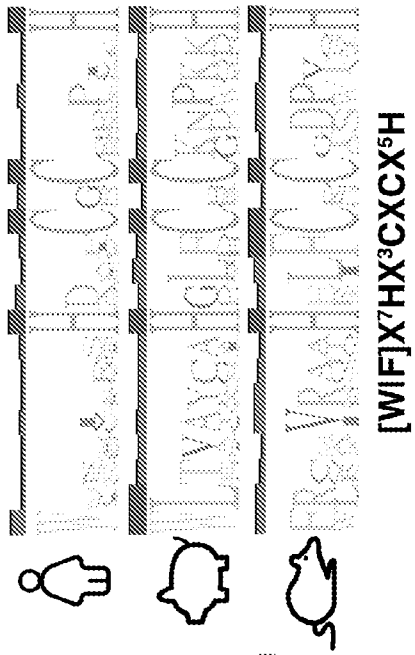


FIG. 43C

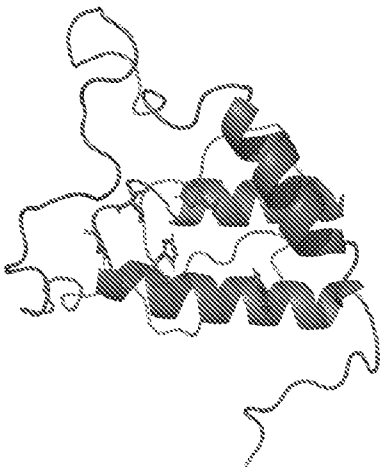


FIG. 43D

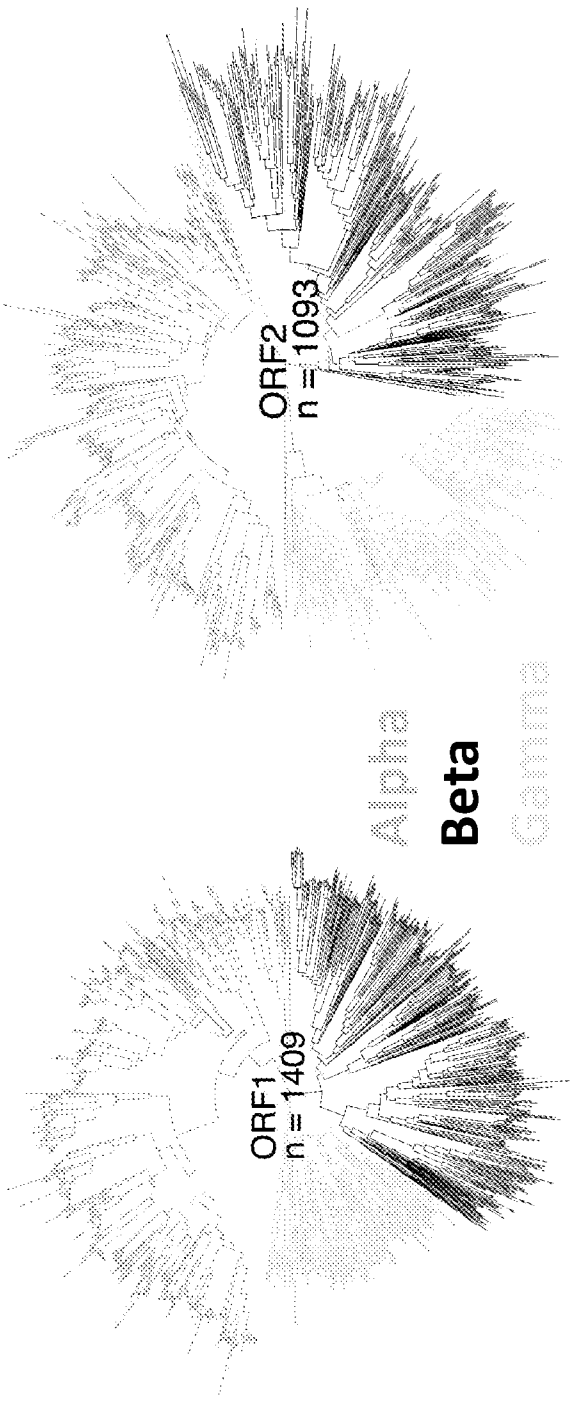


FIG. 44A

Anello read counts

Donor	Gender	Tissue	Anello RNA reads	Anello DNA reads
1	female	Blood	660	109
2	male	Blood	68	171
3	male	Lung	33	113

FIG. 44B

Anello RNA reads mapped to corresponding genome

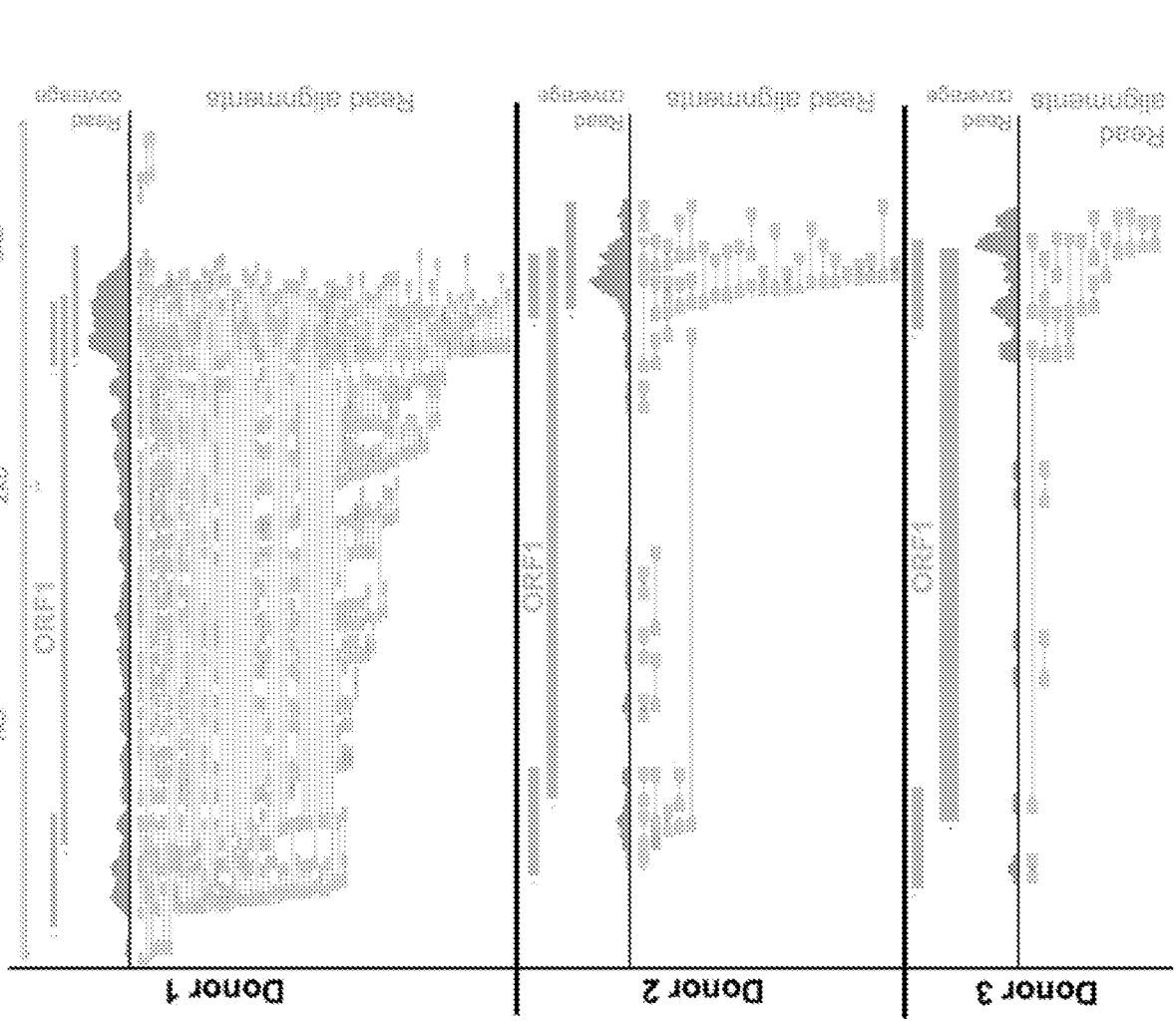


FIG. 45

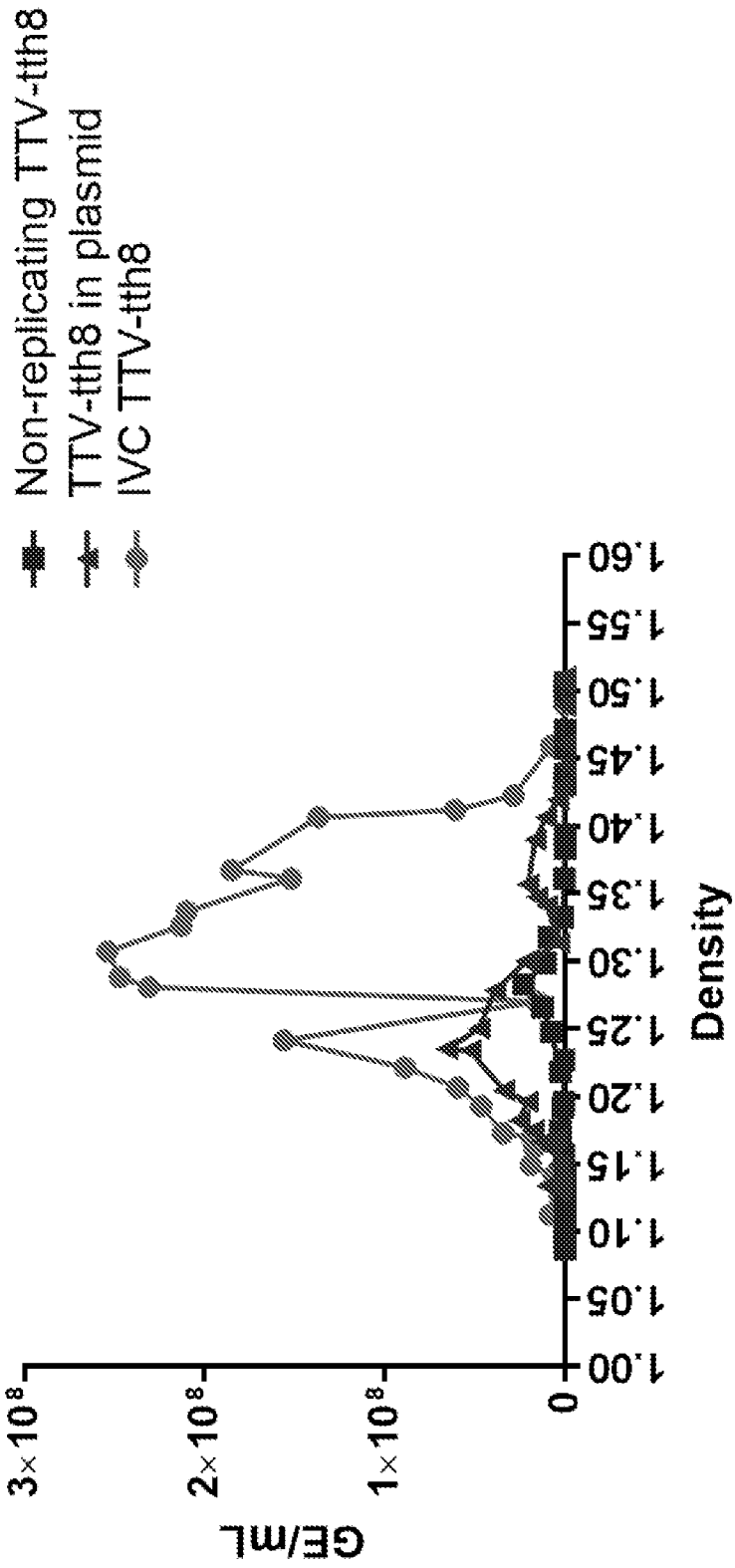

Non-replicating TTV-tth8
TTV-tth8 in plasmid
IVC TTV-tth8
GE/mL
3×10^8
2×10^8
1×10^8
0
1.00
1.05
1.10
1.15
1.20
1.25
1.30
1.35
1.40
1.45
1.50
1.55
1.60
Density

FIG. 46

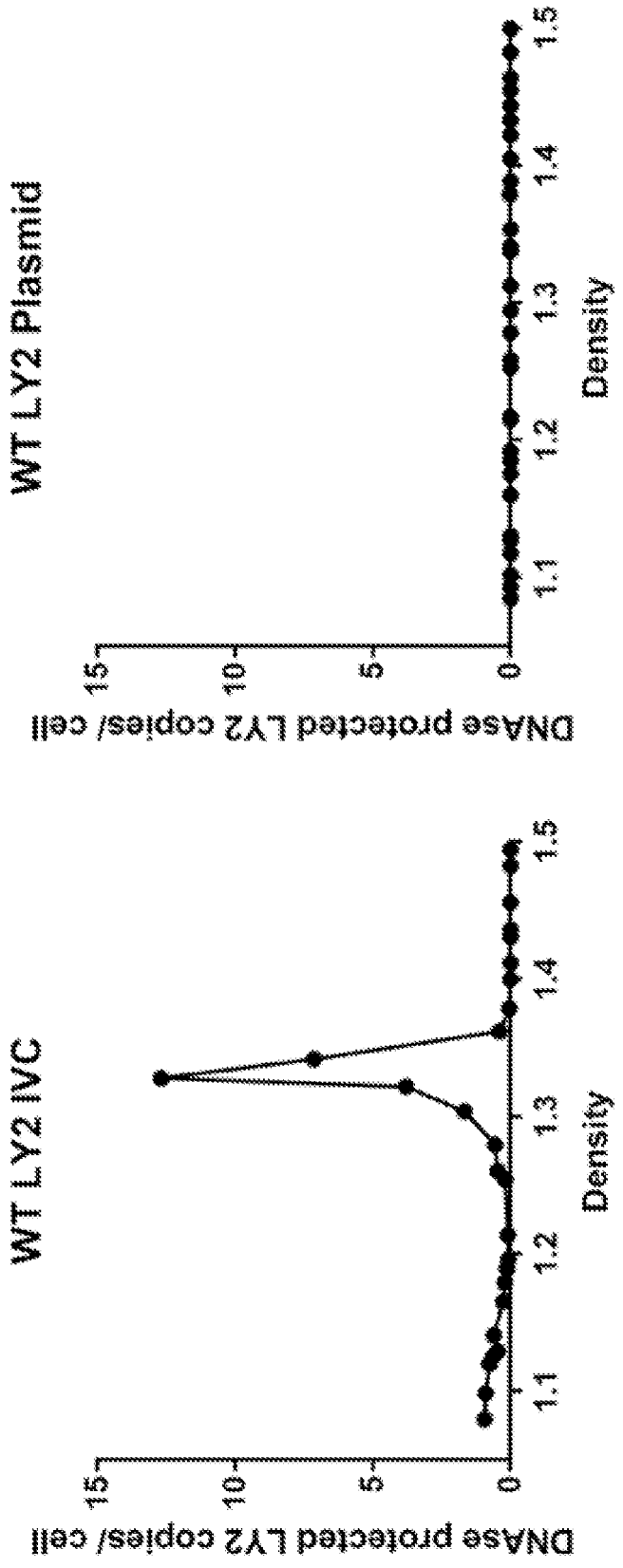

WT LY2 IVC
DNAse protected LY2 copies/ cell
15
10
5
0
1.1
1.2
1.3
1.4
1.5
Density
WT LY2 Plasmid
DNAse protected LY2 copies/ cell
15
10
5
0
1.1
1.2
1.3
1.4
1.5
Density

FIG. 47

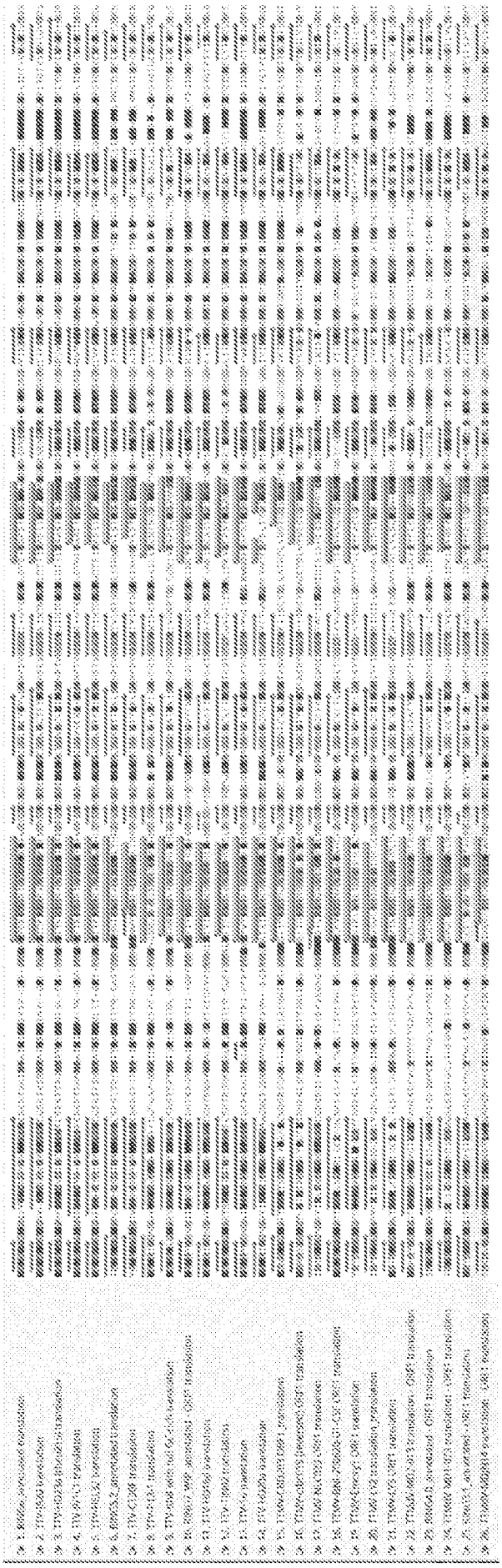


FIG. 48

SEQUENCE LISTING

<110> FLAGSHIP PIONEERING INNOVATIONS V, INC.

<120> ANELLOSOMES FOR DELIVERING PROTEIN REPLACEMENT THERAPEUTIC MODALITIES

<130> V2057-7003WO

<140>
<141>

<150> 62/778,861
<151> 2018-12-12

<150> 62/778,866
<151> 2018-12-12

<160> 1000

<170> PatentIn version 3.5

<210> 1
<211> 3570
<212> DNA
<213> Torque teno virus

<400> 1
attttgtgca gcccgccaat tctcgttcaa acaggccaat caggaggctc tacgtacact 60

tcctggggtg tgtcttcgaa gagtatataa gcagaggcgg tgacgaatgg tagagttttt 120

cctggcccgt ccgcggcgag agcgcgagcg gagcgagcga tcgagcgtcc cgtgggcggg 180

tgccgtaggt gagtttacac accgcagtca aggggcaatt cgggctcggg actggccggg 240

ctatgggcaa gattcttaaa aaattccccc gatccctctg tcgccaggac ataaaaacat 300

gccgtggaga ccgccggtgc atagtgtcca ggggcgagag gatcagtggt tcgcgagctt 360

ttttcacggc cacgcttcat tttgcggttg cggtgacgct gttggccatc ttaatagcat 420

tgctcctcgc tttcctcgcg ccggtccacc aaggcccccc ccggggctag agcagcctaa 480

ccccccgcag cagggcccgg ccgggcccgg agggccgccc gccatcttgg cgctgccggc 540

tccgcccgcg gagcctgacg acccgcagcc acggcgtggt ggtggggacg gtggcgccgc 600

cgctggcgcc gcaggcgacc gtggagaccg agactacgac gaagaagagc tagacgagct 660

tttccgcgcc gccgccgaag acgatttgta agtaggagat ggcgccggcc ttacaggcgc 720

aggaggagac gcgggcgacg cagacgcaga cgcagacgca gacataagcc caccctagta 780

ctcagacagt ggcaacctga cgttatcaga cactgtaaga taacaggacg gatgcccctc 840

attatctgtg gaaaggggtc cacccagttc aactacatca cccacgcgga cgacatcacc 900

cccaggggag cctcctacgg gggcaacttc acaaacatga ctttctccct ggaggcaata 960

tacgaacagt ttctgtacca cagaaacagg tggtcagcct ccaaccacga cctcgaactc 1020

tgcagataca agggtaccac cctaaaactg tacaggcacc cagatgtaga ctacatagtc 1080

acctacagca gaacgggacc ctttgagatc agccacatga cctacctcag cactcacccc 1140

cttctcatgc tgctaaacaa acaccacata gtggtgccca gcctaaagac taagcccagg 1200

ggcagaaagg ccataaaagt cagaataaga ccccccaaac tcatgaacaa caagtggtac 1260

ttcaccagag acttctgtaa cataggcctc ttccagctct gggccacagg cttagaactc 1320

agaaacccct ggctcagaat gagcaccctg agcccctgca taggcttcaa tgtccttaaa 1380

aacagcattt acacaaacct cagcaaccta cctcagcaca gagaagacag acttaacatt 1440

attaacaaca cattacaccc acatgacata acaggaccaa acaataaaaa atggcagtac 1500

acatatacca aactcatggc ccccatttac tattcagcaa acagggccag cacctatgac 1560

ttactacgag agtatggcct ctacagtcca tactacctaa accccacaag gataaacctt 1620

gactggatga ccccctacac acacgtcagg tacaatccac tagtagacaa gggcttcgga 1680

aacagaatat acatacagtg gtgctcagag gcagatgtaa gctacaacag gactaaatcc 1740

aagtgtctct tacaagacat gcccctgttt ttcatgtgct atggctacat agactgggca 1800

attaaaaaca caggggtctc ctcactagcg agagacgcca gaatctgcat caggtgtccc 1860

tacacagagc cacagctggt gggctccaca gaagacatag ggttcgtacc catcacagag 1920

accttcatga ggggcgacat gccggtactt gcaccataca taccgttgag ctggttttgc 1980

aagtggtatc ccaacatagc tcaccagaag gaagtacttg aggcaatcat ttcctgcagc 2040

cccttcatgc cccgtgacca gggcatgaac ggttgggata ttacaatagg ttacaaaatg 2100

gacttcttat ggggcggttc ccctctcccc tcacagccaa tcgacgaccc ctgccagcag 2160

ggaacccacc cgattcccga ccccgataag caccctcgcc tcctacaagt gtcgaacccg 2220

aaactgctcg gaccgaggac agtgttccac aagtgggaca tcagacgtgg gcagtttagc 2280

aaaagaagta ttaaaagagt gtcagaatac tcatcggatg atgaatctct tgcgccaggt 2340

ctcccatcaa agcgaaacaa gctcgactcg gccttcagag gagaaaaccc agagcaaaaa 2400

gaatgctatt ctctcctcaa agcactcgag gaagaagaga ccccagaaga agaagaacca 2460

gcaccccaag aaaaagccca gaaagaggag ctactccacc agctccagct ccagagacgc 2520

caccagcgag tcctcagacg agggctcaag ctcgtcttta cagacatcct ccgactccgc 2580

cagggagtcc actggaaccc cgagctcaca tagagccccc accttacata ccagacctac 2640

tttttcccaa tactggtaaa aaaaaaaaat tctctccctt cgactgggaa acggaggccc 2700

agctagcagg gatattcaag cgtcctatgc gcttctatcc ctcagacacc cctcactacc 2760

cgtggttacc ccccaagcgc gatatcccga aaatatgtaa cataaacttc aaaataaagc 2820

tgcaagagtg agtgattcga ggccctcctc tgttcactta gcggtgtcta cctcttaaag 2880

tcaccaagca ctccgagcgt cagcgaggag tgcgaccctc caccaagggg caacttcctc 2940

ggggtccggc gctacgcgct tcgcgctgcg ccggacgcct cggacccccc cccgacccga 3000

atcgctcgcg cgattcggac ctgcggcctc gggggggggtc gggggcttta ctaaacagac 3060

tccgagttgc cactggactc aggagctgtg aatcagtaac gaaagtgagt ggggccagac 3120

ttcgccatag ggcctttaac ttggggtcgt ctgtcggtgg cttccgggtc cgcctgggcg 3180

ccgccatttt agctttagac gccattttag gccctcgcgg gcacccgtag gcgcgtttta 3240

atgacgtcac ggcagccatt ttgtcgtgac gtttgagaca cgtgatgggg gcgtgcctaa 3300

acccggaagc atccctggtc acgtgactct gacgtcacgg cggccatttt gtgctgtccg 3360

ccatcttgtg acttccttcc gctttttcaa aaaaaaagag gaagtatgac agtagcggcg 3420

gggggggcggc cgcgttcgcg cgccgcccac cagggggtgc tgcgcgcccc cccccgcgca 3480

tgcgcggggc ccccccccgg gggggctccg ccccccggc ccccccccgt gctaaaccca 3540

ccgcgcatgc gcgaccacgc ccccgccgcc 3570

<210> 2
<211> 130
<212> PRT
<213> Torque teno virus

<400> 2
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 Ala Ser Phe Cys Gly Cys Gly
20 25 30

```
Asp Ala Val Gly His Leu Asn Ser Ile Ala Pro Arg Phe Pro Arg Ala
        35                  40                  45

Gly Pro Pro Arg Pro Pro Pro Gly Leu Glu Gln Pro Asn Pro Pro Gln
    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 Gly Asp Arg Gly Asp 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> 3
<211> 303
<212> PRT
<213> Torque teno virus

<400> 3

```
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 Ala Ser Phe Cys Gly Cys Gly
            20                  25                  30

Asp Ala Val Gly His Leu Asn Ser Ile Ala Pro Arg Phe Pro Arg Ala
        35                  40                  45

Gly Pro Pro Arg Pro Pro Pro Gly Leu Glu Gln Pro Asn Pro Pro Gln
    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 Gly Asp Arg Gly Asp 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 Ser Thr Leu Ala Ser Tyr Lys Cys Arg Thr Arg Asn Cys
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 Lys Glu Cys Gln Asn Thr His Arg Met Met
            180                 185                 190

Asn Leu Leu Arg Gln Val Ser His Gln Ser Glu Thr Ser Ser Thr Arg
        195                 200                 205

Pro Ser Glu Glu Lys Thr Gln Ser Lys Lys Asn Ala Ile Leu Ser Ser
    210                 215                 220

Lys His 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 Pro Ser Ser His
        275                 280                 285

Arg Ala Pro Thr Leu His Thr Arg Pro Thr Phe Ser Gln Tyr Trp
```

290 295 300

<210> 4
<211> 293
<212> PRT
<213> Torque teno virus

<400> 4
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 Ala Ser Phe Cys Gly Cys Gly
20 25 30

Asp Ala Val Gly His Leu Asn Ser Ile Ala Pro Arg Phe Pro Arg Ala
35 40 45

Gly Pro Pro Arg Pro Pro Pro Gly Leu Glu Gln Pro Asn Pro Pro Gln
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 Gly Asp Arg Gly Asp 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 Leu Gln Arg
130 135 140

Arg Lys Pro Arg Ala Lys Arg Met Leu Phe Ser Pro Gln Ser Thr 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 Arg Ala His Ile Glu 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 Leu Ala Gly Ile Phe
                245                 250                 255

Lys Arg Pro Met Arg Phe Tyr Pro Ser Asp Thr Pro His Tyr Pro Trp
            260                 265                 270

Leu Pro Pro Lys Arg Asp Ile Pro Lys Ile Cys Asn Ile Asn Phe Lys
        275                 280                 285

Ile Lys Leu Gln Glu
    290
```

<210> 5
<211> 180
<212> PRT
<213> Torque teno virus

<400> 5

```
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 Leu Gln Arg Arg
            20                  25                  30

Lys Pro Arg Ala Lys Arg Met Leu Phe Ser Pro Gln Ser Thr 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 Arg Ala His Ile Glu 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 Leu Ala Gly Ile Phe Lys
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 Lys Arg Asp Ile Pro Lys Ile Cys Asn Ile Asn Phe Lys Ile
165 170 175

Lys Leu Gln Glu
180

<210> 6
<211> 680
<212> PRT
<213> Torque teno virus

<400> 6
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1 5 10 15

Pro Trp Arg Pro 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 Val Leu Arg Gln Trp Gln Pro Asp Val Ile Arg

```
65                  70                  75                  80

His Cys Lys Ile Thr Gly Arg 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 Asp 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 Leu 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 His Arg Glu Asp Arg Leu
        275                 280                 285
```

```
Asn Ile Ile Asn Asn Thr Leu His Pro His Asp Ile Thr Gly Pro 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 Leu Leu Arg Glu Tyr Gly
                325                 330                 335

Leu Tyr Ser Pro Tyr Tyr Leu Asn Pro Thr Arg Ile Asn Leu Asp Trp
            340                 345                 350

Met Thr Pro Tyr Thr His Val Arg Tyr Asn Pro Leu Val Asp Lys Gly
        355                 360                 365

Phe Gly Asn Arg Ile Tyr Ile Gln Trp Cys Ser Glu Ala Asp Val Ser
    370                 375                 380

Tyr Asn Arg Thr Lys Ser Lys Cys Leu Leu Gln Asp Met Pro Leu Phe
385                 390                 395                 400

Phe Met Cys Tyr Gly Tyr Ile Asp Trp Ala Ile Lys Asn Thr Gly Val
                405                 410                 415

Ser Ser Leu Ala Arg Asp Ala Arg Ile Cys Ile Arg Cys Pro Tyr Thr
            420                 425                 430

Glu Pro Gln Leu Val Gly Ser Thr Glu Asp Ile Gly Phe Val Pro Ile
        435                 440                 445

Thr Glu Thr Phe Met Arg Gly Asp Met Pro Val Leu Ala Pro Tyr Ile
    450                 455                 460

Pro Leu Ser Trp Phe Cys Lys Trp Tyr Pro Asn Ile Ala His Gln Lys
465                 470                 475                 480

Glu Val Leu Glu Ala Ile Ile Ser Cys Ser Pro Phe Met Pro Arg Asp
                485                 490                 495

Gln Gly Met Asn Gly Trp Asp Ile Thr Ile Gly Tyr Lys Met Asp Phe
```

```
            500                 505                 510

Leu Trp Gly Gly Ser Pro Leu Pro Ser Gln Pro Ile Asp Asp Pro Cys
        515                 520                 525

Gln Gln Gly Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu
    530                 535                 540

Leu Gln Val Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His
545                 550                 555                 560

Lys Trp Asp Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg
                565                 570                 575

Val Ser Glu Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro
            580                 585                 590

Ser Lys Arg Asn Lys Leu Asp Ser Ala Phe Arg Gly Glu Asn Pro Glu
        595                 600                 605

Gln Lys Glu Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr
    610                 615                 620

Pro Glu Glu Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu
625                 630                 635                 640

Leu Leu His Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg
                645                 650                 655

Arg Gly Leu Lys Leu Val Phe Thr Asp Ile Leu Arg Leu Arg Gln Gly
            660                 665                 670

Val His Trp Asn Pro Glu Leu Thr
        675                 680


<210> 7
<211> 197
<212> PRT
<213> Torque teno virus

<400> 7
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
```

```
1               5                   10                  15

Pro Trp Arg Pro 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 Pro 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> 8
<211> 145
<212> PRT
```

<213> Torque teno virus

<400> 8
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1   5   10   15

Pro Trp Arg Pro 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 Pro Ser Glu Glu Lys Thr Gln Ser Lys Lys Asn Ala Ile Leu
    50   55   60

Ser Ser Lys His 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 Pro Ser
        115   120   125

Ser His Arg Ala Pro Thr Leu His Thr Arg Pro Thr Phe Ser Gln Tyr
    130   135   140

Trp
145

<210> 9
<211> 3451
<212> DNA
<213> Torque teno virus

<400> 9
aattttgcta aacagactcc gaggtgctct tggacactga gtgggcgtac agcaacgaaa 60

gtgagtgggg ccagacttcg ccataaggcc tttatcttcg ggtctacatc ataatataaa 120

gatgtgcact tccgaatggc tgagtttttc acgccattcc gcagcggtgg agcagcgcag 180

ccacgacccc cgcgtcccga gggcgggtgc cggaggtgag tttacacacc gcagtcaagg 240

ggcaattcgg gctcgggact ggccgggccc gggcaaggct cttaaagcga aaccatgttc 300

ctcggcaggc cctaccgcca cagaaagcgg caccaggccg gcaagaaagg gccactgcca 360

ctgccaaatc tgcaacctgc acaggagaaa cgggctggtg gtccgtcctt gatggcctcc 420

ggacgcaggg gatggatgcc cccggacctg acggtccagg agagggagga tgcctggtgg 480

accagcttct gcgctagcca ccgcagcttt tgtagctgcg acgatcctgt gggccatatt 540

aatactctcg cccgcgataa tagtcctctg gcccagactc ctactacaac ttcaggccag 600

gggccgccgc cgccgcctac gcctccgcgg acgccggggc cgcgccctgg gtctgctccg 660

gaccaggggg gaaggatcag ggcctcctgg acctaccccc tagcccccgg aggtcccggt 720

agcacgccat ggcctactgg tggggccgga gacgccggtg gcgccgctgg aggaggcgcc 780

ggcgtcctct ccgccgccgc cggcggtggc ggagaaggcg acgctggccc agaaggcgcc 840

ggtggaggcg aaggagacga cgtgcgagac ctgctcgccg ctatcgaagg agacgtgggc 900

gcagacgggt aaggagacgc cgtcgccccc agaaactagt actgactcag tggaatcccc 960

agactgtgag aaagtgtgtt attagggggt ttctgcccct gttcttctgc ggacaggggg 1020

cctaccacag aaactttaca gaccactatg acgatgtgtt ccccaaggga cccagcggag 1080

gtgggcacgg gagcatggtg ttcaacctgt cctttctgta ccaagagttt aagaagcacc 1140

acaataagtg gtcgcgcagc aacctggact ttgacttagt gagatacaag ggcacagtga 1200

taaagctgta cagacaccag gactttgact acatagtgtg gataagcagg acccctccct 1260

tccaggagag cctgctcaca gtaatgaccc accagcccag cgtcatgctg caggcaaaaa 1320

agtgcataat agtaaagagc tacaggaccc acccgggggg caaaccctat gtaactgcaa 1380

aagttaggcc ccccagactc ctaactgaca agtggtactt ccagtcagac ttctgcaacg 1440

ttccgctttt tagcctacag tttgcccttg cggaactgcg gtttccgatc tgctcaccac 1500

aaactgacac caattgcatt aacttcctgg tgttagatga catctactac aagtttctag 1560

ataataagcc taaacagagt tcagacccta atgacgaaaa cagaataaaa ttctggcacg 1620

gcctatggtc cactatgaga tatttaaaca ccacctacat aaacacactg tttccaggca 1680

cagacagtct agtggccgcc aaagatactg acaatagtgt aaataaatac cccagcacag 1740

ccactaaaca gccctacaaa gacagtcagt acatgcaaaa tatatggaat acatcaaaaa 1800

tacatgcctt atatacgtgg gtagcagaga caaactacaa aagactgcag gcctactaca 1860

cacagaccta cggaggctac cagagacaat ttttcacagg aaaacagtac tgggactaca 1920

gagtaggcat gtttagtcca gccttcctga gtcccagcag actaaatccc cagaacccag 1980

gggcatacac agaggtctcc tacaacccct ggacagacga gggcacgggc aacgtagtgt 2040

gcctgcagta tctgactaaa gagacctcag actacaaacc aggtggtggg agcaagttct 2100

gcatagaagg tgtgcctcta tgggcagcgc tggtgggata cgtagacatg tgtaaaaaag 2160

agggcaagga cccgggcatc agactaaact gtctcctgtt agtcaagtgt ccctatacaa 2220

agcctcagct gtatgacaaa aaaaaccccg agaaactgtt tgtaccttac tcctataact 2280

ttgggcacgg caagatgccg gggggagaca aatacatacc catagagttc aaagacaggt 2340

ggtacccctg cctgctccac caagaggagt ggatagagga cattgtcagg tcgggaccct 2400

tcgttccaaa agacatgccc agcagcgtca cctgcatgat gaggtacagc tctcttttta 2460

actggggcgg taatataatc caagaacagg ccgtggaaga cccctgtaag aaaggcacct 2520

tcgtcgttcc cggaaccagt ggcatcgctc gcatactaca agtcagcaac ccggccaagc 2580

agacccccac gacaacctgg cactcgtggg actggagacg atccctcttt acagagacgg 2640

gtcttaaaag aatgcgcgaa caacaaccat atgatgaact gtcttatacg ggccctaaaa 2700

agccaaaact gtcccttccc gcagggcccg ccgtccccgg tgccgccgtc gcctcctcct 2760

ggtgggaaac aaaacaggtc acctcgccag acgtcagcga gacggagacc gaagcagaag 2820

cccaccaaga ggaagagacg gagccggagg agggagtcca gctccagcag ctgtgggagc 2880

agcaactcct gcaaaagcga cagctgggag tcgtgttcca gcaactcctc cgactcagac 2940

agggggcgga gatccacccg ggcctcgtat aattcctggg ccccagaacc cgtacctgct 3000

tttcccggag caggcccctc caaaagtgcc tatttttgac ccctttggtc agaaaacaga 3060

gctagagctg tgcggctgct tcgacaggcc gcccaggaac aacccctacg accacccctt 3120

ctacccctgg ctgcccaaag agcctccctc ctactaccag ggctacaaag tgtctttcaa 3180

actagggttc cacccagaca agcatgtgtg aaccccgcca ataaaccact gctgctacac 3240

tgattcttag gccgtgggag tctcactggt cggtgtctac ctcttaaggt cactaagcac 3300

tccgagcgtt agcgaggagt gcgaccctac cccctgggcc cacttcttcg gagccgcgcg 3360

ctacgccttc ggctgcgcgc ggcacctcag acccccgctc gtgctgacac gcttgcgcgt 3420

gtcagaccac ttcgggctcg cgggggtcgg g 3451

<210> 10
<211> 166
<212> PRT
<213> Torque teno virus

<400> 10
Met Ala Ser Gly Arg Arg Gly Trp Met Pro Pro Asp Leu Thr Val Gln
1 5 10 15

Glu Arg Glu Asp Ala Trp Trp Thr Ser Phe Cys Ala Ser His Arg Ser
20 25 30

Phe Cys Ser Cys Asp Asp Pro Val Gly His Ile Asn Thr Leu Ala Arg
35 40 45

Asp Asn Ser Pro Leu Ala Gln Thr Pro Thr Thr Thr Ser Gly Gln Gly
50 55 60

Pro Pro Pro Pro Pro Thr Pro Pro Arg Thr Pro Gly Pro Arg Pro Gly
65 70 75 80

Ser Ala Pro Asp Gln Gly Gly Arg Ile Arg Ala Ser Trp Thr Tyr Pro
85 90 95

Leu Ala Pro Gly Gly Pro Gly Ser Thr Pro Trp Pro Thr Gly Gly Ala
100 105 110

Gly Asp Ala Gly Gly Ala Ala Gly Gly Gly Ala Gly Val Leu Ser Ala
115 120 125

Ala Ala Gly Gly Gly Gly Glu Gly Asp Ala Gly Pro Glu Gly Ala Gly
130 135 140

Gly Gly Glu Gly Asp Asp Val Arg Asp Leu Leu Ala Ala Ile Glu Gly
145 150 155 160

Asp Val Gly Ala Asp Gly
165

<210> 11
<211> 348
<212> PRT
<213> Torque teno virus

<400> 11
Met Ala Ser Gly Arg Arg Gly Trp Met Pro Pro Asp Leu Thr Val Gln
1   5   10   15

Glu Arg Glu Asp Ala Trp Trp Thr Ser Phe Cys Ala Ser His Arg Ser
            20   25   30

Phe Cys Ser Cys Asp Asp Pro Val Gly His Ile Asn Thr Leu Ala Arg
        35   40   45

Asp Asn Ser Pro Leu Ala Gln Thr Pro Thr Thr Thr Ser Gly Gln Gly
    50   55   60

Pro Pro Pro Pro Pro Thr Pro Pro Arg Thr Pro Gly Pro Arg Pro Gly
65   70   75   80

Ser Ala Pro Asp Gln Gly Gly Arg Ile Arg Ala Ser Trp Thr Tyr Pro
                85   90   95

Leu Ala Pro Gly Gly Pro Gly Ser Thr Pro Trp Pro Thr Gly Gly Ala
            100   105   110

Gly Asp Ala Gly Gly Ala Ala Gly Gly Gly Ala Gly Val Leu Ser Ala
        115   120   125

Ala Ala Gly Gly Gly Gly Glu Gly Asp Ala Gly Pro Glu Gly Ala Gly
    130   135   140

Gly Gly Glu Gly Asp Asp Val Arg Asp Leu Leu Ala Ala Ile Glu Gly
145   150   155   160

Asp Val Gly Ala Asp Gly Pro Trp Lys Thr Pro Val Arg Lys Ala Pro
                165   170   175

Ser Ser Phe Pro Glu Pro Val Ala Ser Leu Ala Tyr Tyr Lys Ser Ala
            180   185   190

```
Thr Arg Pro Ser Arg Pro Pro Arg Gln Pro Gly Thr Arg Gly Thr Gly
        195                 200                 205

Asp Asp Pro Ser Leu Gln Arg Arg Val Leu Lys Glu Cys Ala Asn Asn
    210                 215                 220

Asn His Met Met Asn Cys Leu Ile Arg Ala Leu Lys Ser Gln Asn Cys
225                 230                 235                 240

Pro Phe Pro Gln Gly Pro Pro Ser Pro Val Pro Pro Ser Pro Pro Pro
                245                 250                 255

Gly Gly Lys Gln Asn Arg Ser Pro Arg Gln Thr Ser Ala Arg Arg Arg
            260                 265                 270

Pro Lys Gln Lys Pro Thr Lys Arg Lys Arg Arg Ser Arg Arg Arg Glu
        275                 280                 285

Ser Ser Ser Ser Ser Cys Gly Ser Ser Asn Ser Cys Lys Ser Asp Ser
    290                 295                 300

Trp Glu Ser Cys Ser Ser Asn Ser Ser Asp Ser Asp Arg Gly Arg Arg
305                 310                 315                 320

Ser Thr Arg Ala Ser Tyr Asn Ser Trp Ala Pro Glu Pro Val Pro Ala
                325                 330                 335

Phe Pro Gly Ala Gly Pro Ser Lys Ser Ala Tyr Phe
            340                 345


<210> 12
<211> 327
<212> PRT
<213> Torque teno virus

<400> 12
Met Ala Ser Gly Arg Arg Gly Trp Met Pro Pro Asp Leu Thr Val Gln
1               5                   10                  15

Glu Arg Glu Asp Ala Trp Trp Thr Ser Phe Cys Ala Ser His Arg Ser
            20                  25                  30
```

```
Phe Cys Ser Cys Asp Asp Pro Val Gly His Ile Asn Thr Leu Ala Arg
        35                  40                  45

Asp Asn Ser Pro Leu Ala Gln Thr Pro Thr Thr Thr Ser Gly Gln Gly
    50                  55                  60

Pro Pro Pro Pro Pro Thr Pro Pro Arg Thr Pro Gly Pro Arg Pro Gly
65                  70                  75                  80

Ser Ala Pro Asp Gln Gly Gly Arg Ile Arg Ala Ser Trp Thr Tyr Pro
                85                  90                  95

Leu Ala Pro Gly Gly Pro Gly Ser Thr Pro Trp Pro Thr Gly Gly Ala
            100                 105                 110

Gly Asp Ala Gly Gly Ala Ala Gly Gly Gly Ala Gly Val Leu Ser Ala
        115                 120                 125

Ala Ala Gly Gly Gly Gly Glu Gly Asp Ala Gly Pro Glu Gly Ala Gly
    130                 135                 140

Gly Gly Glu Gly Asp Asp Val Arg Asp Leu Leu Ala Ala Ile Glu Gly
145                 150                 155                 160

Asp Val Gly Ala Asp Gly Ala Arg Arg Pro Arg Cys Arg Arg Arg Leu
                165                 170                 175

Leu Leu Val Gly Asn Lys Thr Gly His Leu Ala Arg Arg Gln Arg Asp
            180                 185                 190

Gly Asp Arg Ser Arg Ser Pro Pro Arg Gly Arg Asp Gly Ala Gly Gly
        195                 200                 205

Gly Ser Pro Ala Pro Ala Ala Val Gly Ala Ala Thr Pro Ala Lys Ala
    210                 215                 220

Thr Ala Gly Ser Arg Val Pro Ala Thr Pro Pro Thr Gln Thr Gly Gly
225                 230                 235                 240

Gly Asp Pro Pro Gly Pro Arg Ile Ile Pro Gly Pro Gln Asn Pro Tyr
                245                 250                 255
```

```
Leu Leu Phe Pro Glu Gln Ala Pro Pro Lys Val Pro Ile Phe Asp Pro
            260                 265                 270

Phe Gly Gln Lys Thr Glu Leu Glu Leu Cys Gly Cys Phe Asp Arg Pro
        275                 280                 285

Pro Arg Asn Asn Pro Tyr Asp His Pro Phe Tyr Pro Trp Leu Pro Lys
    290                 295                 300

Glu Pro Pro Ser Tyr Tyr Gln Gly Tyr Lys Val Ser Phe Lys Leu Gly
305                 310                 315                 320

Phe His Pro Asp Lys His Val
                325
```

<210> 13
<211> 747
<212> PRT
<213> Torque teno virus

<400> 13

```
Met Ala Tyr Trp Trp Gly Arg Arg Arg Arg Trp Arg Arg Trp Arg Arg
1               5                   10                  15

Arg Arg Arg Pro Leu Arg Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg
            20                  25                  30

Trp Pro Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg Arg Ala Arg Pro
        35                  40                  45

Ala Arg Arg Tyr Arg Arg Arg Arg Gly Arg Arg Arg Val Arg Arg Arg
    50                  55                  60

Arg Arg Pro Gln Lys Leu Val Leu Thr Gln Trp Asn Pro Gln Thr Val
65                  70                  75                  80

Arg Lys Cys Val Ile Arg Gly Phe Leu Pro Leu Phe Phe Cys Gly Gln
                85                  90                  95

Gly Ala Tyr His Arg Asn Phe Thr Asp His Tyr Asp Asp Val Phe Pro
            100                 105                 110
```

```
Lys Gly Pro Ser Gly Gly Gly His Gly Ser Met Val Phe Asn Leu Ser
        115                 120                 125

Phe Leu Tyr Gln Glu Phe Lys Lys His His Asn Lys Trp Ser Arg Ser
    130                 135                 140

Asn Leu Asp Phe Asp Leu Val Arg Tyr Lys Gly Thr Val Ile Lys Leu
145                 150                 155                 160

Tyr Arg His Gln Asp Phe Asp Tyr Ile Val Trp Ile Ser Arg Thr Pro
                165                 170                 175

Pro Phe Gln Glu Ser Leu Leu Thr Val Met Thr His Gln Pro Ser Val
            180                 185                 190

Met Leu Gln Ala Lys Lys Cys Ile Ile Val Lys Ser Tyr Arg Thr His
        195                 200                 205

Pro Gly Gly Lys Pro Tyr Val Thr Ala Lys Val Arg Pro Pro Arg Leu
    210                 215                 220

Leu Thr Asp Lys Trp Tyr Phe Gln Ser Asp Phe Cys Asn Val Pro Leu
225                 230                 235                 240

Phe Ser Leu Gln Phe Ala Leu Ala Glu Leu Arg Phe Pro Ile Cys Ser
                245                 250                 255

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

Tyr Tyr Lys Phe Leu Asp Asn Lys Pro Lys Gln Ser Ser Asp Pro Asn
        275                 280                 285

Asp Glu Asn Arg Ile Lys Phe Trp His Gly Leu Trp Ser Thr Met Arg
    290                 295                 300

Tyr Leu Asn Thr Thr Tyr Ile Asn Thr Leu Phe Pro Gly Thr Asp Ser
305                 310                 315                 320
```

```
Leu Val Ala Ala Lys Asp Thr Asp Asn Ser Val Asn Lys Tyr Pro Ser
                325                 330                 335

Thr Ala Thr Lys Gln Pro Tyr Lys Asp Ser Gln Tyr Met Gln Asn Ile
            340                 345                 350

Trp Asn Thr Ser Lys Ile His Ala Leu Tyr Thr Trp Val Ala Glu Thr
        355                 360                 365

Asn Tyr Lys Arg Leu Gln Ala Tyr Tyr Thr Gln Thr Tyr Gly Gly Tyr
    370                 375                 380

Gln Arg Gln Phe Phe Thr Gly Lys Gln Tyr Trp Asp Tyr Arg Val Gly
385                 390                 395                 400

Met Phe Ser Pro Ala Phe Leu Ser Pro Ser Arg Leu Asn Pro Gln Asn
                405                 410                 415

Pro Gly Ala Tyr Thr Glu Val Ser Tyr Asn Pro Trp Thr Asp Glu Gly
            420                 425                 430

Thr Gly Asn Val Val Cys Leu Gln Tyr Leu Thr Lys Glu Thr Ser Asp
        435                 440                 445

Tyr Lys Pro Gly Gly Gly Ser Lys Phe Cys Ile Glu Gly Val Pro Leu
    450                 455                 460

Trp Ala Ala Leu Val Gly Tyr Val Asp Met Cys Lys Lys Glu Gly Lys
465                 470                 475                 480

Asp Pro Gly Ile Arg Leu Asn Cys Leu Leu Leu Val Lys Cys Pro Tyr
                485                 490                 495

Thr Lys Pro Gln Leu Tyr Asp Lys Lys Asn Pro Glu Lys Leu Phe Val
            500                 505                 510

Pro Tyr Ser Tyr Asn Phe Gly His Gly Lys Met Pro Gly Gly Asp Lys
        515                 520                 525

Tyr Ile Pro Ile Glu Phe Lys Asp Arg Trp Tyr Pro Cys Leu Leu His
    530                 535                 540
```

```
Gln Glu Glu Trp Ile Glu Asp Ile Val Arg Ser Gly Pro Phe Val Pro
545                 550                 555                 560

Lys Asp Met Pro Ser Ser Val Thr Cys Met Met Arg Tyr Ser Ser Leu
                565                 570                 575

Phe Asn Trp Gly Gly Asn Ile Ile Gln Glu Gln Ala Val Glu Asp Pro
            580                 585                 590

Cys Lys Lys Gly Thr Phe Val Val Pro Gly Thr Ser Gly Ile Ala Arg
        595                 600                 605

Ile Leu Gln Val Ser Asn Pro Ala Lys Gln Thr Pro Thr Thr Thr Trp
    610                 615                 620

His Ser Trp Asp Trp Arg Arg Ser Leu Phe Thr Glu Thr Gly Leu Lys
625                 630                 635                 640

Arg Met Arg Glu Gln Gln Pro Tyr Asp Glu Leu Ser Tyr Thr Gly Pro
                645                 650                 655

Lys Lys Pro Lys Leu Ser Leu Pro Ala Gly Pro Ala Val Pro Gly Ala
            660                 665                 670

Ala Val Ala Ser Ser Trp Trp Glu Thr Lys Gln Val Thr Ser Pro Asp
        675                 680                 685

Val Ser Glu Thr Glu Thr Glu Ala Glu Ala His Gln Glu Glu Glu Thr
    690                 695                 700

Glu Pro Glu Glu Gly Val Gln Leu Gln Gln Leu Trp Glu Gln Gln Leu
705                 710                 715                 720

Leu Gln Lys Arg Gln Leu Gly Val Val Phe Gln Gln Leu Leu Arg Leu
                725                 730                 735

Arg Gln Gly Ala Glu Ile His Pro Gly Leu Val
            740                 745
```

<210> 14
<211> 220
<212> PRT
<213> Torque teno virus

<400> 14
Met Ala Tyr Trp Trp Gly Arg Arg Arg Arg Trp Arg Arg Trp Arg Arg
1   5   10   15

Arg Arg Arg Pro Leu Arg Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg
            20   25   30

Trp Pro Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg Arg Ala Arg Pro
        35   40   45

Ala Arg Arg Tyr Arg Arg Arg Arg Gly Arg Arg Arg Ala Val Glu Asp
    50   55   60

Pro Cys Lys Lys Gly Thr Phe Val Val Pro Gly Thr Ser Gly Ile Ala
65   70   75   80

Arg Ile Leu Gln Val Ser Asn Pro Ala Lys Gln Thr Pro Thr Thr Thr
                85   90   95

Trp His Ser Trp Asp Trp Arg Arg Ser Leu Phe Thr Glu Thr Gly Leu
            100   105   110

Lys Arg Met Arg Glu Gln Gln Pro Tyr Asp Glu Leu Ser Tyr Thr Gly
        115   120   125

Pro Lys Lys Pro Lys Leu Ser Leu Pro Ala Gly Pro Ala Val Pro Gly
    130   135   140

Ala Ala Val Ala Ser Ser Trp Trp Glu Thr Lys Gln Val Thr Ser Pro
145   150   155   160

Asp Val Ser Glu Thr Glu Thr Glu Ala Glu Ala His Gln Glu Glu Glu
                165   170   175

Thr Glu Pro Glu Glu Gly Val Gln Leu Gln Gln Leu Trp Glu Gln Gln
            180   185   190

Leu Leu Gln Lys Arg Gln Leu Gly Val Val Phe Gln Gln Leu Leu Arg
195 200 205

Leu Arg Gln Gly Ala Glu Ile His Pro Gly Leu Val
210 215 220

<210> 15
<211> 164
<212> PRT
<213> Torque teno virus

<400> 15
Met Ala Tyr Trp Trp Gly Arg Arg Arg Arg Trp Arg Arg Trp Arg Arg
1 5 10 15

Arg Arg Arg Pro Leu Arg Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg
20 25 30

Trp Pro Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg Arg Ala Arg Pro
35 40 45

Ala Arg Arg Tyr Arg Arg Arg Arg Gly Arg Arg Arg Gly Pro Pro Ser
50 55 60

Pro Val Pro Pro Ser Pro Pro Pro Gly Gly Lys Gln Asn Arg Ser Pro
65 70 75 80

Arg Gln Thr Ser Ala Arg Arg Arg Pro Lys Gln Lys Pro Thr Lys Arg
85 90 95

Lys Arg Arg Ser Arg Arg Arg Glu Ser Ser Ser Ser Ser Cys Gly Ser
100 105 110

Ser Asn Ser Cys Lys Ser Asp Ser Trp Glu Ser Cys Ser Ser Asn Ser
115 120 125

Ser Asp Ser Asp Arg Gly Arg Arg Ser Thr Arg Ala Ser Tyr Asn Ser
130 135 140

Trp Ala Pro Glu Pro Val Pro Ala Phe Pro Gly Ala Gly Pro Ser Lys
145 150 155 160

Ser Ala Tyr Phe

<210> 16
<211> 3753
<212> DNA
<213> Torque teno virus

<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

tacctttaat 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 tacttttgg 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> Torque teno virus

<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> Torque teno virus

<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
            260                 265


<210> 19
<211> 276
<212> PRT
<213> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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
<211> 3878
<212> DNA
<213> Torque teno virus

<400> 24
aaatacgtca ctaaccacgt gactcccaca ggccaaccac agtctatgtc gtgcacttcc 60

tgggcatggt ctacgtgata atataaagcg gtgcacttcc gaatggctga gttttccacg 120

cccgtccgca gcgagatcgc gacgtaggag cgatcgagcg tcccgagggc gggtgccgga 180

ggtgagttta cacaccgcag tcaaggggca attcgggctc gggaggccgg gccatgggca 240

aggctcttaa aaagctatgt ttctcggtaa aatctacagg aagaaaagga aactgcttct 300

gcaggctgtg cgtgctccgc agacgccatc ttccatgagc cgctgctggt gtccccctcg 360

gggtgatgtc tcctcccgcg agtctcgatg gtacgaggcg gttcgaggaa gccacgatgc 420

tttttgtggc tgtagtgatc ctattcttca tctttctcgt ctggctgcac gttttaacca 480

tcagggacct ccgacgcccc ccacggacga ccgtgcgccg cagaataccc cagtgagacg 540

cctgctgcct ctccccagct accccggcga gggtccccag gctagatggc ctggtgggga 600

tggaggcgcc gctggtggcg accgaagaga aggtggagat ggcggcgcgc gcgccgccga 660

agacgagtac cagcccgaag acctagacga gcttttcggc gctatcgaac aagaacagta 720

aggaggaggc gaagggggag gcggaggggc taccggcgcc gttacagact gagacgctat 780

gccagacgca ggttccgacg caaaaagata gtactgactc agtggaaccc ccagactacc 840

agaaaatgta taataagggg catgatgcca gtactgtggg ccggcatggg tacgggggggc 900

agaaactatg cagtgaggtc agatgactat gtggtgaaca aagggttcgg gggctccttc 960

gccacggaga ccttctccct gaaggttctc tatgaccagt ttcaaagggg cttcaacagg 1020

tggtcccaca ctaacgagga cctagacctg gcccgctaca ggggctgcag gtggactttt 1080

tacagacata aagacacaga ctttatagtg tactttacaa acaatcctcc catgaagacc 1140

aaccagttct ccgcgcccct gacgaccccc ggcatgctca tgcgcagtaa atacaaagtc 1200

ctcattccca gcttccagac cagacccaag ggtcgcaaaa cagtaaccgt taaaataaga 1260

ccccccaaac tatttcaaga caagtggtac acccagcagg acctgtgttc agttcctctt 1320

gtccaactga acgtgaccgc agctgatttc acacatccgt tcggctcacc actaactgaa 1380

actccttgcg tagagttcca ggtgctgggt gacttgtaca atacatgtct caatatcgac 1440

cttccgcaat ttagtgaatt aggagaaata actagtgcct actcaaaacc aaactcaaat 1500

aacctaaaag aattatacaa agaattgttc acaaaagcca catcaggaca ctactggcag 1560

acattcataa ccaacagcat ggtcagagca cacatagatg cagacaaagc taaagaagca 1620

caaagagcat ccaccacacc ctcatacaac aatgacccct tccccacaat acctgttaaa 1680

tcagagtttg cacagtggaa aaagaaattc acagacacta gagacagccc ctttcttttt 1740

gccacttacc atcccgaagc tataaaagac acaattatga aaatgagaga gaacaacttt 1800

aagctagaga caggacccaa tgacaagtat ggagactaca cagcacagta ccaaggaaac 1860

acacacatgc tagactacta ccttggcttt tacagcccca tattcctctc agatggaagg 1920

tctaacgtag aattcttcac tgcctacaga gacatagtat acaatccctt cttagacaag 1980

gcccagggca acatggtgtg gtttcagtac cacacaaaga cagacaacaa gtttaaaaaa 2040

ccagagtgcc actgggaaat caaagacatg cccctgtggg ccctcctaaa cggatatgta 2100

gactacttag agactcaaat acagtatggt gacctcagta aagaagggaa agtcctcatc 2160

aggtgtccct acaccaagcc agcactagta gaccccagag acgacactgc aggatatgta 2220

gtctacaaca gaaactttgg cagaggcaag tggatagacg gaggggggcta catccccctg 2280

cacgagagga caaaatggta cgtgatgctc agataccaga cggacgtctt ccatgacata 2340

gtgacctgtg ggccctggca gtacagagac gacaacaaaa acagccagct agtggccaaa 2400

taccgcttca gctttatatg gggaggtaac actgtccact ctcaggtcat cagaaacccg 2460

tgcaaagaca accaagtatc cggtccccgt cgacagccta gggatataca agtcgttgac 2520

ccgcaacgca tcacgccgcc gtgggtcctc cacagcttcg accagcgaag aggcctcttt 2580

actgaaacag ctctcaggcg cctgctccag gaaccactac ctggcgagta tgctgttagc 2640

accctcagga cacccctcct ctttctaccc tcagaatacc agcgagaaga cggcgctgca 2700

gaaagcgcct caggttcacc ggccaaaaga ccccgtatct ggtcagaaga gagtcagacg 2760

gagacgatct cctcggagga gaacccggcg gagacgacga gggagctcct ccagcgaaag 2820

ctccgagagc agcgagcact ccagttccaa ctccagcact tcgcggtcca actcgccaag 2880

acccaggcga atctccacgt aaaccccctg ttatctttcc cgcaatgaat aaggtctttc 2940

tgtttccccc agagggtccc aagcccatcc tgggcaaaga ggcctggcag gacgagtacg 3000

agacctgcag ggtctggaac agacctgcca gaacccacca cacagacacc cccttctatc 3060

cctgggcccc ccacaagttc catgtaagct tcaaacttgg cttccaataa aattactagg 3120

ccgtggaact ctcactggtc ggtgtctacc tcttaaggtc actaagcact ccgagcgtca 3180

gcgaggagtg cgaccctcta ccctggtgca acgccctcgg cggccgcgcg ctacgccttc 3240

ggctgcgcgc ggcacctcgg acccccgctc gtgctgacgc gctcgcgcgc gtcagaccac 3300

ttcgggctcg cgggggtcgg gaattttgct aaacagactc cgagttgcca ttggacactg 3360

tagctgtgaa tcagtaacga aagtgagtgg ggccagactt cgccataggg cctttatctt 3420

cttgccattg gtccgtgtag ggggtcgcca taggcttcga cctccctttt aggccttccg 3480

gactacaaaa atggcggatt cagtgacgtc acggccgcca ttttaagtag gtgccgtcca 3540

ggactgcagt tccgggtcag agtgcatcct cggcggaacc tgcacaaaat ggcggtcaat 3600

atcttccggg tcaaaggtca cacctacgtc ataagtcacg tgactgggtc ctgctacgtc 3660

atatgcggaa gtaggccccg ccacgtgact cgtcacgtgg gcgctgcgtc acggcggcca 3720

ttttgtatca caaaatggcg gacttccttc ctctttttta aaaataacgg cccagcggcg 3780

gcgcgcgcgc ttcgcgcgcg cgccgggggg ctccgccccc ccccgcgcat gcgcggggcc 3840

ccccccgcg gggggctccg ccccccggtc cccccccg 3878

<210> 25
<211> 128
<212> PRT
<213> Torque teno virus

<400> 25

```
Met Ser Arg Cys Trp Cys Pro Pro Arg Gly Asp Val Ser Ser Arg Glu
1               5                   10                  15


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


Cys Ser Asp Pro Ile Leu His Leu Ser Arg Leu Ala Ala Arg Phe Asn
        35                  40                  45


His Gln Gly Pro Pro Thr Pro Pro Thr Asp Asp Arg Ala Pro Gln Asn
    50                  55                  60


Thr Pro Val Arg Arg Leu Leu Pro Leu Pro Ser Tyr Pro Gly Glu Gly
65                  70                  75                  80


Pro Gln Ala Arg Trp Pro Gly Gly Asp Gly Gly Ala Ala Gly Gly Asp
                85                  90                  95


Arg Arg Glu Gly Gly Asp Gly Gly Ala Arg Ala Ala Glu Asp Glu Tyr
            100                 105                 110


Gln Pro Glu Asp Leu Asp Glu Leu Phe Gly Ala Ile Glu Gln Glu Gln
        115                 120                 125
```

<210> 26
<211> 279
<212> PRT
<213> Torque teno virus

<400> 26

```
Met Ser Arg Cys Trp Cys Pro Pro Arg Gly Asp Val Ser Ser Arg Glu
1               5                   10                  15


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


Cys Ser Asp Pro Ile Leu His Leu Ser Arg Leu Ala Ala Arg Phe Asn
        35                  40                  45


His Gln Gly Pro Pro Thr Pro Pro Thr Asp Asp Arg Ala Pro Gln Asn
    50                  55                  60
```

```
Thr Pro Val Arg Arg Leu Leu Pro Leu Pro Ser Tyr Pro Gly Glu Gly
65                  70                  75                  80

Pro Gln Ala Arg Trp Pro Gly Gly Asp Gly Gly Ala Ala Gly Gly Asp
                85                  90                  95

Arg Arg Glu Gly Gly Asp Gly Gly Ala Arg Ala Ala Glu Asp Glu Tyr
            100                 105                 110

Gln Pro Glu Asp Leu Asp Glu Leu Phe Gly Ala Ile Glu Gln Glu Gln
        115                 120                 125

Ser Ser Glu Thr Arg Ala Lys Thr Thr Lys Tyr Pro Val Pro Val Asp
    130                 135                 140

Ser Leu Gly Ile Tyr Lys Ser Leu Thr Arg Asn Ala Ser Arg Arg Arg
145                 150                 155                 160

Gly Ser Ser Thr Ala Ser Thr Ser Glu Glu Ala Ser Leu Leu Lys Gln
                165                 170                 175

Leu Ser Gly Ala Cys Ser Arg Asn His Tyr Leu Ala Ser Met Leu Leu
            180                 185                 190

Ala Pro Ser Gly His Pro Ser Ser Phe Tyr Pro Gln Asn Thr Ser Glu
        195                 200                 205

Lys Thr Ala Leu Gln Lys Ala Pro Gln Val His Arg Pro Lys Asp Pro
    210                 215                 220

Val Ser Gly Gln Lys Arg Val Arg Arg Arg Arg Ser Pro Arg Arg Arg
225                 230                 235                 240

Thr Arg Arg Arg Arg Arg Gly Ser Ser Ser Ser Glu Ser Ser Glu Ser
                245                 250                 255

Ser Glu His Ser Ser Ser Asn Ser Ser Thr Ser Arg Ser Asn Ser Pro
            260                 265                 270

Arg Pro Arg Arg Ile Ser Thr
        275
```

<210> 27
<211> 272
<212> PRT
<213> Torque teno virus

<400> 27
Met Ser Arg Cys Trp Cys Pro Pro Arg Gly Asp Val Ser Ser Arg Glu
1   5   10   15

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

Cys Ser Asp Pro Ile Leu His Leu Ser Arg Leu Ala Ala Arg Phe Asn
        35   40   45

His Gln Gly Pro Pro Thr Pro Pro Thr Asp Asp Arg Ala Pro Gln Asn
    50   55   60

Thr Pro Val Arg Arg Leu Leu Pro Leu Pro Ser Tyr Pro Gly Glu Gly
65   70   75   80

Pro Gln Ala Arg Trp Pro Gly Gly Asp Gly Gly Ala Ala Gly Gly Asp
                85   90   95

Arg Arg Glu Gly Gly Asp Gly Gly Ala Arg Ala Ala Glu Asp Glu Tyr
            100   105   110

Gln Pro Glu Asp Leu Asp Glu Leu Phe Gly Ala Ile Glu Gln Glu Gln
        115   120   125

Ile Pro Ala Arg Arg Arg Arg Cys Arg Lys Arg Leu Arg Phe Thr Gly
    130   135   140

Gln Lys Thr Pro Tyr Leu Val Arg Arg Glu Ser Asp Gly Asp Asp Leu
145   150   155   160

Leu Gly Gly Glu Pro Gly Gly Asp Asp Glu Gly Ala Pro Pro Ala Lys
                165   170   175

Ala Pro Arg Ala Ala Ser Thr Pro Val Pro Thr Pro Ala Leu Arg Gly
            180   185   190

```
Pro Thr Arg Gln Asp Pro Gly Glu Ser Pro Arg Lys Pro Pro Val Ile
        195                 200                 205

Phe Pro Ala Met Asn Lys Val Phe Leu Phe Pro Pro Glu Gly Pro Lys
    210                 215                 220

Pro Ile Leu Gly Lys Glu Ala Trp Gln Asp Glu Tyr Glu Thr Cys Arg
225                 230                 235                 240

Val Trp Asn Arg Pro Ala Arg Thr His His Thr Asp Thr Pro Phe Tyr
                245                 250                 255

Pro Trp Ala Pro His Lys Phe His Val Ser Phe Lys Leu Gly Phe Gln
            260                 265                 270
```

<210> 28
<211> 780
<212> PRT
<213> Torque teno virus

<400> 28

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

Trp Arg Trp Arg Arg Ala Arg Arg Arg Arg Arg Val Pro Ala Arg Arg
            20                  25                  30

Pro Arg Arg Ala Phe Arg Arg Tyr Arg Thr Arg Thr Val Arg Arg Arg
        35                  40                  45

Arg Arg Gly Arg Arg Arg Gly Tyr Arg Arg Arg Tyr Arg Leu Arg Arg
    50                  55                  60

Tyr Ala Arg Arg Arg Phe Arg Arg Lys Lys Ile Val Leu Thr Gln Trp
65                  70                  75                  80

Asn Pro Gln Thr Thr Arg Lys Cys Ile Ile Arg Gly Met Met Pro Val
                85                  90                  95

Leu Trp Ala Gly Met Gly Thr Gly Gly Arg Asn Tyr Ala Val Arg Ser
            100                 105                 110
```

```
Asp Asp Tyr Val Val Asn Lys Gly Phe Gly Gly Ser Phe Ala Thr Glu
        115                 120                 125

Thr Phe Ser Leu Lys Val Leu Tyr Asp Gln Phe Gln Arg Gly Phe Asn
    130                 135                 140

Arg Trp Ser His Thr Asn Glu Asp Leu Asp Leu Ala Arg Tyr Arg Gly
145                 150                 155                 160

Cys Arg Trp Thr Phe Tyr Arg His Lys Asp Thr Asp Phe Ile Val Tyr
                165                 170                 175

Phe Thr Asn Asn Pro Pro Met Lys Thr Asn Gln Phe Ser Ala Pro Leu
            180                 185                 190

Thr Thr Pro Gly Met Leu Met Arg Ser Lys Tyr Lys Val Leu Ile Pro
        195                 200                 205

Ser Phe Gln Thr Arg Pro Lys Gly Arg Lys Thr Val Thr Val Lys Ile
    210                 215                 220

Arg Pro Pro Lys Leu Phe Gln Asp Lys Trp Tyr Thr Gln Gln Asp Leu
225                 230                 235                 240

Cys Ser Val Pro Leu Val Gln Leu Asn Val Thr Ala Ala Asp Phe Thr
                245                 250                 255

His Pro Phe Gly Ser Pro Leu Thr Glu Thr Pro Cys Val Glu Phe Gln
            260                 265                 270

Val Leu Gly Asp Leu Tyr Asn Thr Cys Leu Asn Ile Asp Leu Pro Gln
        275                 280                 285

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

Asn Asn Leu Lys Glu Leu Tyr Lys Glu Leu Phe Thr Lys Ala Thr Ser
305                 310                 315                 320
```

```
Gly His Tyr Trp Gln Thr Phe Ile Thr Asn Ser Met Val Arg Ala His
               325                 330                 335

Ile Asp Ala Asp Lys Ala Lys Glu Ala Gln Arg Ala Ser Thr Thr Pro
            340                 345                 350

Ser Tyr Asn Asn Asp Pro Phe Pro Thr Ile Pro Val Lys Ser Glu Phe
        355                 360                 365

Ala Gln Trp Lys Lys Lys Phe Thr Asp Thr Arg Asp Ser Pro Phe Leu
    370                 375                 380

Phe Ala Thr Tyr His Pro Glu Ala Ile Lys Asp Thr Ile Met Lys Met
385                 390                 395                 400

Arg Glu Asn Asn Phe Lys Leu Glu Thr Gly Pro Asn Asp Lys Tyr Gly
               405                 410                 415

Asp Tyr Thr Ala Gln Tyr Gln Gly Asn Thr His Met Leu Asp Tyr Tyr
            420                 425                 430

Leu Gly Phe Tyr Ser Pro Ile Phe Leu Ser Asp Gly Arg Ser Asn Val
        435                 440                 445

Glu Phe Phe Thr Ala Tyr Arg Asp Ile Val Tyr Asn Pro Phe Leu Asp
    450                 455                 460

Lys Ala Gln Gly Asn Met Val Trp Phe Gln Tyr His Thr Lys Thr Asp
465                 470                 475                 480

Asn Lys Phe Lys Lys Pro Glu Cys His Trp Glu Ile Lys Asp Met Pro
               485                 490                 495

Leu Trp Ala Leu Leu Asn Gly Tyr Val Asp Tyr Leu Glu Thr Gln Ile
            500                 505                 510

Gln Tyr Gly Asp Leu Ser Lys Glu Gly Lys Val Leu Ile Arg Cys Pro
        515                 520                 525

Tyr Thr Lys Pro Ala Leu Val Asp Pro Arg Asp Asp Thr Ala Gly Tyr
    530                 535                 540
```

```
Val Val Tyr Asn Arg Asn Phe Gly Arg Gly Lys Trp Ile Asp Gly Gly
545                 550                 555                 560

Gly Tyr Ile Pro Leu His Glu Arg Thr Lys Trp Tyr Val Met Leu Arg
                565                 570                 575

Tyr Gln Thr Asp Val Phe His Asp Ile Val Thr Cys Gly Pro Trp Gln
            580                 585                 590

Tyr Arg Asp Asp Asn Lys Asn Ser Gln Leu Val Ala Lys Tyr Arg Phe
        595                 600                 605

Ser Phe Ile Trp Gly Gly Asn Thr Val His Ser Gln Val Ile Arg Asn
    610                 615                 620

Pro Cys Lys Asp Asn Gln Val Ser Gly Pro Arg Arg Gln Pro Arg Asp
625                 630                 635                 640

Ile Gln Val Val Asp Pro Gln Arg Ile Thr Pro Pro Trp Val Leu His
                645                 650                 655

Ser Phe Asp Gln Arg Arg Gly Leu Phe Thr Glu Thr Ala Leu Arg Arg
            660                 665                 670

Leu Leu Gln Glu Pro Leu Pro Gly Glu Tyr Ala Val Ser Thr Leu Arg
        675                 680                 685

Thr Pro Leu Leu Phe Leu Pro Ser Glu Tyr Gln Arg Glu Asp Gly Ala
    690                 695                 700

Ala Glu Ser Ala Ser Gly Ser Pro Ala Lys Arg Pro Arg Ile Trp Ser
705                 710                 715                 720

Glu Glu Ser Gln Thr Glu Thr Ile Ser Ser Glu Glu Asn Pro Ala Glu
                725                 730                 735

Thr Thr Arg Glu Leu Leu Gln Arg Lys Leu Arg Glu Gln Arg Ala Leu
            740                 745                 750
```

```
Gln Phe Gln Leu Gln His Phe Ala Val Gln Leu Ala Lys Thr Gln Ala
        755                 760                 765


Asn Leu His Val Asn Pro Leu Leu Ser Phe Pro Gln
    770                 775                 780


<210> 29
<211> 204
<212> PRT
<213> Torque teno virus

<400> 29
Met Ala Trp Trp Gly Trp Arg Arg Arg Trp Trp Arg Pro Lys Arg Arg
1               5                   10                  15


Trp Arg Trp Arg Arg Ala Arg Arg Arg Arg Arg Val Pro Ala Arg Arg
            20                  25                  30


Pro Arg Arg Ala Phe Arg Arg Tyr Arg Thr Arg Thr Val Ile Arg Asn
        35                  40                  45


Pro Cys Lys Asp Asn Gln Val Ser Gly Pro Arg Arg Gln Pro Arg Asp
    50                  55                  60


Ile Gln Val Val Asp Pro Gln Arg Ile Thr Pro Pro Trp Val Leu His
65                  70                  75                  80


Ser Phe Asp Gln Arg Arg Gly Leu Phe Thr Glu Thr Ala Leu Arg Arg
                85                  90                  95


Leu Leu Gln Glu Pro Leu Pro Gly Glu Tyr Ala Val Ser Thr Leu Arg
            100                 105                 110


Thr Pro Leu Leu Phe Leu Pro Ser Glu Tyr Gln Arg Glu Asp Gly Ala
        115                 120                 125


Ala Glu Ser Ala Ser Gly Ser Pro Ala Lys Arg Pro Arg Ile Trp Ser
    130                 135                 140


Glu Glu Ser Gln Thr Glu Thr Ile Ser Ser Glu Glu Asn Pro Ala Glu
145                 150                 155                 160
```

```
Thr Thr Arg Glu Leu Leu Gln Arg Lys Leu Arg Glu Gln Arg Ala Leu
                165                 170                 175

Gln Phe Gln Leu Gln His Phe Ala Val Gln Leu Ala Lys Thr Gln Ala
            180                 185                 190

Asn Leu His Val Asn Pro Leu Leu Ser Phe Pro Gln
        195                 200
```

<210> 30
<211> 119
<212> PRT
<213> Torque teno virus

<400> 30

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

Trp Arg Trp Arg Arg Ala Arg Arg Arg Arg Arg Val Pro Ala Arg Arg
            20                  25                  30

Pro Arg Arg Ala Phe Arg Arg Tyr Arg Thr Arg Thr Asn Thr Ser Glu
        35                  40                  45

Lys Thr Ala Leu Gln Lys Ala Pro Gln Val His Arg Pro Lys Asp Pro
    50                  55                  60

Val Ser Gly Gln Lys Arg Val Arg Arg Arg Arg Ser Pro Arg Arg Arg
65                  70                  75                  80

Thr Arg Arg Arg Arg Arg Gly Ser Ser Ser Ser Glu Ser Ser Glu Ser
                85                  90                  95

Ser Glu His Ser Ser Ser Asn Ser Ser Thr Ser Arg Ser Asn Ser Pro
            100                 105                 110

Arg Pro Arg Arg Ile Ser Thr
        115
```

<210> 31
<211> 3818
<212> DNA
<213> Torque teno virus

<400> 31
aagtccgcca ctaaccacgt gactcccgca ggccaaccca gtactatgtc gtccacttcc 60

tgggacgagt ctacgtcctg atataagtaa gtgcacttcc gaatggctga gttttccacg 120

cccgtccgca gcgagaacgc cacggagggg agtccgcgcg tcccgagggc gggtgccgga 180

ggtgagttta cacaccgcag tcaaggggca attcgggctc gggactggcc gggccccggg 240

caaggctctt aaaaaatgca ctttcgcaga gtgcgagcga aaaggaaact gctactgcaa 300

gctgtgcgag ctccaccgaa ggcacctgcc atgagcttca ccacacctac tattaatgcc 360

gggatccgag agcagcaatg gttcgagtcc acccttagat cccaccactc gttctgtggc 420

tgtggtgatc ccgtgcttca ttttactaac cttgctactc gctttaacta tctgcctgct 480

acctcttcgc ctctggaccc tcccggccca gcgccgcgag gccgcccggc gctccgccgc 540

ctcccggcac tcccttcagc ccccgcgacc ccttctagag aactagcatg gcctactggt 600

tcagaaggtg gggctggagg ccgaggcgcc ggtggagaag gtggcgccgc cgtcgaagga 660

gactaccgag aagaagaact agacgagctg ttcgcggcct tggaagaaga cgcaaaccaa 720

gggtaaggag gcgccgcaga actcgcagac gtacctacag acgggggtgg agacgcagga 780

ggtacataag acgggggcga cgcaaaaaga aactcatact gactcagtgg aacccggcaa 840

tagttaagag gtgcaacatt aagggcggac ttccaataat tatatgcgga gagcccaggg 900

cagcctttaa ctatggctac cacatggagg actacactcc tcaacctttc cccttcggag 960

ggggaatgag cacagtgact ttctctctga aagccttgta tgaccagtac ctaaaacacc 1020

aaaacaggtg gactttctca aacgaccagc tagacctcgc cagatacagg ggctgtaaac 1080

taaggttcta cagaagcccc gtctgtgact ttatagtaca ctacaaccta atacctccac 1140

taaaaatgaa ccagttcaca agtcccaaca cgcacccggg actactcatg ctcagcaaac 1200

acaagataat aattcccagc tttcaaacaa gacctggggg cagacgcttt gttaaaataa 1260

gacttaatcc ccccaaacta tttgaagaca agtggtacac tcagcaagac ctgtgcaagg 1320

ttccgctcgt tagtattaca gcaactgcgg ctgacttgcg gtatccgttc tgctcaccac 1380

aaacgaacaa cccttgcacc accttccagg tactgcgcaa gaactacaat acagttatag 1440

gaacttccgt aaaagaccaa gagtccacac aagactttga aaattggctt tataaaacag 1500

actcacacta tcaaacattt gccacagagg ctcaactagg cagaattcct gcatttaatc 1560

ctgatggcac taaaaacact aaacagcagt cgtggcaaga taactggagc aaaaaaaatt 1620

caccatggac aggtaactca ggtacatacc cacaaacaac cagtgaaatg tacaaaattc 1680

catatgacag taacttcggc tttcccacat acagagccca aaaagactac attttagaaa 1740

gaagacagtg caactttaac tatgaagtta ataatccagt tagcaaaaaa gtatggccac 1800

aacctagtac aacaacaccc acagtagact actatgaata ccactgtgga tggttcagca 1860

acatattcat aggccccaac agatacaacc tacagtttca aacagcatat gtagacacca 1920

catacaaccc actaatggac aagggcaaag gcaacaaaat atggtttcaa tatctgtcta 1980

aaaagggcac agactacaat gaaaaacaat gctactgcac cctagaagac atgcccctat 2040

gggcaatatg ctttggatac actgactatg tagagactca actaggaccc aatgtggacc 2100

atgaaacagc aggcttaata attatgatct gtccatacac tcaaccacct atgtatgaca 2160

aaaacagacc taactgggga tacgtagtct atgacacaaa ctttggcaat ggaaaaatgc 2220

cctcaggaag tggccaagtc ccagtatact ggcaatgccg atggaggccc atgctgtggt 2280

tccaacaaca agtactcaat gacatctcaa agactggacc gtacgcctac agagacgaat 2340

ataaaaatgt acaactgact ctctactaca actttatttt taactggggg ggcgacatgt 2400

attacccaca ggtcgttaaa aacccctgtg gagactccgg aatcgttccc ggttccggta 2460

gattcactcg agaagtacaa gtcgttagcc cgctttccat gggaccggcc tacatcttcc 2520

actacttcga ctccagacgc gggttcttta gtgaaaaagc tcttaaaaga atgcaacaac 2580

aacaagaatt tgatgaatct tttacattca aacctaagag acccaaactt tctacagcag 2640

ccgcagaaat cctccagctc gaagaagact cgacttcagg ggaaggaaaa tcgccactac 2700

agcaagaaga gaaagaagtc gaagtcctcc aaacgccgac agtacagctc cagctccagc 2760

gaaacatcca ggagcagctc gcaatcaagc agcagctcca attcctcttg ctccaactcc 2820

tcaaaaccca atccaatttg catttaaacc cacaattttt aagcccttca taaaatatga 2880

catgtttggg gacccccttc ctcacccccc aacagccgaa gagtgggaaa cagagtacca 2940

gtgctgtaag gcctttaaca gaccacctag aaccaaccta aaagacaccc ccttctaccc 3000

ctgggtacct aaacctaaac ctcaattccg tgtatctttt aaacttggtt ttcaataaac 3060

aaggccgtgg gagtttcact tgtcggtgtc aacctcttaa ggtcactaag cactccgagc 3120

gtaagcgagg agtgcgaccc tccccctgg ggcaactccc tcgaagtccg gcgctacgcg 3180

cttcgcgctg cgccggacat ctcggacccc ccctccaccc gaaacgcttg cgcgtttcgg 3240

accttcggcg tcgggggggt cgggggcttt actaaacaga ctccgaggtg ccattggaca 3300

ctgaggggat gaacagcaac gaaagtgagt ggggccagac ttcgccataa ggcctttatc 3360

ttcttgccat ttgtcagtat agagggtcgc cataggcttc ggcctccatt ttaacctcta 3420

aaaactacca aaatggccgt tccagtgacg tcacagccgc cattttaagt agctgacgtc 3480

aaggattgac gtgaaggtta aaggtcatcc tcggcggaag ctacacaaaa tggtggacaa 3540

catcttccgg gtcaaaggtc gtgcacacgt cataagtcac gtggtgggga cccgctgtaa 3600

cccggaagta ggccccgtca cgtgatttgt cacgtgtgta cacgtcacaa ccgccatttt 3660

gttttacaaa atggctgact tccttcctct tttttaaaaa aaacggccgt gcggcggcgc 3720

gcgcgcttcg cgcgcgcgcc gggggctgcc gccccccccc gcgcatgcgc gcggggcccc 3780

cccccgcggg gggctccgcc ccccggcccc cccccccg 3818

<210> 32
<211> 131
<212> PRT
<213> Torque teno virus

<400> 32
Met Ser Phe Thr Thr Pro Thr Ile Asn Ala Gly Ile Arg Glu Gln Gln
1 5 10 15

Trp Phe Glu Ser Thr Leu Arg Ser His His Ser Phe Cys Gly Cys Gly
20 25 30

Asp Pro Val Leu His Phe Thr Asn Leu Ala Thr Arg Phe Asn Tyr Leu
35 40 45

Pro Ala Thr Ser Ser Pro Leu Asp Pro Pro Gly Pro Ala Pro Arg Gly
50 55 60

Arg Pro Ala Leu Arg Arg Leu Pro Ala Leu Pro Ser Ala Pro Ala Thr
65 70 75 80

Pro Ser Arg Glu Leu Ala Trp Pro Thr Gly Ser Glu Gly Gly Ala Gly
85 90 95

Gly Arg Gly Ala Gly Gly Glu Gly Gly Ala Ala Val Glu Gly Asp Tyr

```
            100                 105                 110

Arg Glu Glu Glu Leu Asp Glu Leu Phe Ala Ala Leu Glu Glu Asp Ala
        115                 120                 125

Asn Gln Gly
    130
```

<210> 33
<211> 275
<212> PRT
<213> Torque teno virus

<400> 33

```
Met Ser Phe Thr Thr Pro Thr Ile Asn Ala Gly Ile Arg Glu Gln Gln
1               5                   10                  15

Trp Phe Glu Ser Thr Leu Arg Ser His His Ser Phe Cys Gly Cys Gly
            20                  25                  30

Asp Pro Val Leu His Phe Thr Asn Leu Ala Thr Arg Phe Asn Tyr Leu
        35                  40                  45

Pro Ala Thr Ser Ser Pro Leu Asp Pro Pro Gly Pro Ala Pro Arg Gly
    50                  55                  60

Arg Pro Ala Leu Arg Arg Leu Pro Ala Leu Pro Ser Ala Pro Ala Thr
65                  70                  75                  80

Pro Ser Arg Glu Leu Ala Trp Pro Thr Gly Ser Glu Gly Gly Ala Gly
                85                  90                  95

Gly Arg Gly Ala Gly Gly Glu Gly Gly Ala Ala Val Glu Gly Asp Tyr
            100                 105                 110

Arg Glu Glu Glu Leu Asp Glu Leu Phe Ala Ala Leu Glu Glu Asp Ala
        115                 120                 125

Asn Gln Gly Ser Leu Lys Thr Pro Val Glu Thr Pro Glu Ser Phe Pro
    130                 135                 140

Val Pro Val Asp Ser Leu Glu Lys Tyr Lys Ser Leu Ala Arg Phe Pro
```

145 150 155 160

Trp Asp Arg Pro Thr Ser Ser Thr Thr Ser Thr Pro Asp Ala Gly Ser
165 170 175

Leu Val Lys Lys Leu Leu Lys Glu Cys Asn Asn Asn Lys Asn Leu Met
180 185 190

Asn Leu Leu His Ser Asn Leu Arg Asp Pro Asn Phe Leu Gln Gln Pro
195 200 205

Gln Lys Ser Ser Ser Ser Lys Lys Thr Arg Leu Gln Gly Lys Glu Asn
210 215 220

Arg His Tyr Ser Lys Lys Arg Lys Lys Ser Lys Ser Ser Lys Arg Arg
225 230 235 240

Gln Tyr Ser Ser Ser Ser Ser Glu Thr Ser Arg Ser Ser Ser Gln Ser
245 250 255

Ser Ser Ser Ser Asn Ser Ser Cys Ser Asn Ser Ser Lys Pro Asn Pro
260 265 270

Ile Cys Ile
275

<210> 34
<211> 270
<212> PRT
<213> Torque teno virus

<400> 34
Met Ser Phe Thr Thr Pro Thr Ile Asn Ala Gly Ile Arg Glu Gln Gln
1 5 10 15

Trp Phe Glu Ser Thr Leu Arg Ser His His Ser Phe Cys Gly Cys Gly
20 25 30

Asp Pro Val Leu His Phe Thr Asn Leu Ala Thr Arg Phe Asn Tyr Leu
35 40 45

Pro Ala Thr Ser Ser Pro Leu Asp Pro Pro Gly Pro Ala Pro Arg Gly

```
    50                  55                  60

Arg Pro Ala Leu Arg Arg Leu Pro Ala Leu Pro Ser Ala Pro Ala Thr
65                  70                  75                  80

Pro Ser Arg Glu Leu Ala Trp Pro Thr Gly Ser Glu Gly Gly Ala Gly
                85                  90                  95

Gly Arg Gly Ala Gly Gly Glu Gly Gly Ala Ala Val Glu Gly Asp Tyr
            100                 105                 110

Arg Glu Glu Glu Leu Asp Glu Leu Phe Ala Ala Leu Glu Glu Asp Ala
        115                 120                 125

Asn Gln Gly Ser Arg Arg Asn Pro Pro Ala Arg Arg Arg Leu Asp Phe
    130                 135                 140

Arg Gly Arg Lys Ile Ala Thr Thr Ala Arg Arg Glu Arg Ser Arg Ser
145                 150                 155                 160

Pro Pro Asn Ala Asp Ser Thr Ala Pro Ala Pro Ala Lys His Pro Gly
                165                 170                 175

Ala Ala Arg Asn Gln Ala Ala Ala Pro Ile Pro Leu Ala Pro Thr Pro
            180                 185                 190

Gln Asn Pro Ile Gln Phe Ala Phe Lys Pro Thr Ile Phe Lys Pro Phe
        195                 200                 205

Ile Lys Tyr Asp Met Phe Gly Asp Pro Leu Pro His Pro Pro Thr Ala
    210                 215                 220

Glu Glu Trp Glu Thr Glu Tyr Gln Cys Cys Lys Ala Phe Asn Arg Pro
225                 230                 235                 240

Pro Arg Thr Asn Leu Lys Asp Thr Pro Phe Tyr Pro Trp Val Pro Lys
                245                 250                 255

Pro Lys Pro Gln Phe Arg Val Ser Phe Lys Leu Gly Phe Gln
            260                 265                 270
```

<210> 35
<211> 156
<212> PRT
<213> Torque teno virus

<400> 35
Met Ser Phe Thr Thr Pro Thr Ile Asn Ala Gly Ile Arg Glu Gln Gln
1 5 10 15

Cys Ser Arg Arg Asn Pro Pro Ala Arg Arg Arg Leu Asp Phe Arg Gly
            20 25 30

Arg Lys Ile Ala Thr Thr Ala Arg Arg Glu Arg Ser Arg Ser Pro Pro
        35 40 45

Asn Ala Asp Ser Thr Ala Pro Ala Pro Ala Lys His Pro Gly Ala Ala
    50 55 60

Arg Asn Gln Ala Ala Ala Pro Ile Pro Leu Ala Pro Thr Pro Gln Asn
65 70 75 80

Pro Ile Gln Phe Ala Phe Lys Pro Thr Ile Phe Lys Pro Phe Ile Lys
                85 90 95

Tyr Asp Met Phe Gly Asp Pro Leu Pro His Pro Pro Thr Ala Glu Glu
            100 105 110

Trp Glu Thr Glu Tyr Gln Cys Cys Lys Ala Phe Asn Arg Pro Pro Arg
        115 120 125

Thr Asn Leu Lys Asp Thr Pro Phe Tyr Pro Trp Val Pro Lys Pro Lys
    130 135 140

Pro Gln Phe Arg Val Ser Phe Lys Leu Gly Phe Gln
145 150 155

<210> 36
<211> 761
<212> PRT
<213> Torque teno virus

<400> 36
Met Ala Tyr Trp Phe Arg Arg Trp Gly Trp Arg Pro Arg Arg Arg Trp

```
1               5                   10                  15

Arg Arg Trp Arg Arg Arg Arg Arg Leu Pro Arg Arg Arg Thr Arg
            20                  25                  30

Arg Ala Val Arg Gly Leu Gly Arg Arg Arg Lys Pro Arg Val Arg Arg
        35                  40                  45

Arg Arg Arg Thr Arg Arg Arg Thr Tyr Arg Arg Gly Trp Arg Arg Arg
    50                  55                  60

Arg Tyr Ile Arg Arg Gly Arg Arg Lys Lys Lys Leu Ile Leu Thr Gln
65                  70                  75                  80

Trp Asn Pro Ala Ile Val Lys Arg Cys Asn Ile Lys Gly Gly Leu Pro
                85                  90                  95

Ile Ile Ile Cys Gly Glu Pro Arg Ala Ala Phe Asn Tyr Gly Tyr His
            100                 105                 110

Met Glu Asp Tyr Thr Pro Gln Pro Phe Pro Phe Gly Gly Gly Met Ser
        115                 120                 125

Thr Val Thr Phe Ser Leu Lys Ala Leu Tyr Asp Gln Tyr Leu Lys His
    130                 135                 140

Gln Asn Arg Trp Thr Phe Ser Asn Asp Gln Leu Asp Leu Ala Arg Tyr
145                 150                 155                 160

Arg Gly Cys Lys Leu Arg Phe Tyr Arg Ser Pro Val Cys Asp Phe Ile
                165                 170                 175

Val His Tyr Asn Leu Ile Pro Pro Leu Lys Met Asn Gln Phe Thr Ser
            180                 185                 190

Pro Asn Thr His Pro Gly Leu Leu Met Leu Ser Lys His Lys Ile Ile
        195                 200                 205

Ile Pro Ser Phe Gln Thr Arg Pro Gly Gly Arg Arg Phe Val Lys Ile
    210                 215                 220
```

```
Arg Leu Asn Pro Pro Lys Leu Phe Glu Asp Lys Trp Tyr Thr Gln Gln
225                 230                 235                 240

Asp Leu Cys Lys Val Pro Leu Val Ser Ile Thr Ala Thr Ala Ala Asp
                245                 250                 255

Leu Arg Tyr Pro Phe Cys Ser Pro Gln Thr Asn Asn Pro Cys Thr Thr
            260                 265                 270

Phe Gln Val Leu Arg Lys Asn Tyr Asn Thr Val Ile Gly Thr Ser Val
        275                 280                 285

Lys Asp Gln Glu Ser Thr Gln Asp Phe Glu Asn Trp Leu Tyr Lys Thr
    290                 295                 300

Asp Ser His Tyr Gln Thr Phe Ala Thr Glu Ala Gln Leu Gly Arg Ile
305                 310                 315                 320

Pro Ala Phe Asn Pro Asp Gly Thr Lys Asn Thr Lys Gln Gln Ser Trp
                325                 330                 335

Gln Asp Asn Trp Ser Lys Lys Asn Ser Pro Trp Thr Gly Asn Ser Gly
            340                 345                 350

Thr Tyr Pro Gln Thr Thr Ser Glu Met Tyr Lys Ile Pro Tyr Asp Ser
        355                 360                 365

Asn Phe Gly Phe Pro Thr Tyr Arg Ala Gln Lys Asp Tyr Ile Leu Glu
    370                 375                 380

Arg Arg Gln Cys Asn Phe Asn Tyr Glu Val Asn Asn Pro Val Ser Lys
385                 390                 395                 400

Lys Val Trp Pro Gln Pro Ser Thr Thr Thr Pro Thr Val Asp Tyr Tyr
                405                 410                 415

Glu Tyr His Cys Gly Trp Phe Ser Asn Ile Phe Ile Gly Pro Asn Arg
            420                 425                 430

Tyr Asn Leu Gln Phe Gln Thr Ala Tyr Val Asp Thr Thr Tyr Asn Pro
```

```
        435                 440                 445

Leu Met Asp Lys Gly Lys Gly Asn Lys Ile Trp Phe Gln Tyr Leu Ser
    450                 455                 460

Lys Lys Gly Thr Asp Tyr Asn Glu Lys Gln Cys Tyr Cys Thr Leu Glu
465                 470                 475                 480

Asp Met Pro Leu Trp Ala Ile Cys Phe Gly Tyr Thr Asp Tyr Val Glu
                485                 490                 495

Thr Gln Leu Gly Pro Asn Val Asp His Glu Thr Ala Gly Leu Ile Ile
            500                 505                 510

Met Ile Cys Pro Tyr Thr Gln Pro Pro Met Tyr Asp Lys Asn Arg Pro
        515                 520                 525

Asn Trp Gly Tyr Val Val Tyr Asp Thr Asn Phe Gly Asn Gly Lys Met
    530                 535                 540

Pro Ser Gly Ser Gly Gln Val Pro Val Tyr Trp Gln Cys Arg Trp Arg
545                 550                 555                 560

Pro Met Leu Trp Phe Gln Gln Gln Val Leu Asn Asp Ile Ser Lys Thr
                565                 570                 575

Gly Pro Tyr Ala Tyr Arg Asp Glu Tyr Lys Asn Val Gln Leu Thr Leu
            580                 585                 590

Tyr Tyr Asn Phe Ile Phe Asn Trp Gly Gly Asp Met Tyr Tyr Pro Gln
        595                 600                 605

Val Val Lys Asn Pro Cys Gly Asp Ser Gly Ile Val Pro Gly Ser Gly
    610                 615                 620

Arg Phe Thr Arg Glu Val Gln Val Val Ser Pro Leu Ser Met Gly Pro
625                 630                 635                 640

Ala Tyr Ile Phe His Tyr Phe Asp Ser Arg Arg Gly Phe Phe Ser Glu
                645                 650                 655
```

```
Lys Ala Leu Lys Arg Met Gln Gln Gln Glu Phe Asp Glu Ser Phe
            660                 665                 670

Thr Phe Lys Pro Lys Arg Pro Lys Leu Ser Thr Ala Ala Ala Glu Ile
        675                 680                 685

Leu Gln Leu Glu Glu Asp Ser Thr Ser Gly Glu Gly Lys Ser Pro Leu
    690                 695                 700

Gln Gln Glu Glu Lys Glu Val Glu Val Leu Gln Thr Pro Thr Val Gln
705                 710                 715                 720

Leu Gln Leu Gln Arg Asn Ile Gln Glu Gln Leu Ala Ile Lys Gln Gln
                725                 730                 735

Leu Gln Phe Leu Leu Leu Gln Leu Leu Lys Thr Gln Ser Asn Leu His
            740                 745                 750

Leu Asn Pro Gln Phe Leu Ser Pro Ser
        755                 760
```

<210> 37
<211> 198
<212> PRT
<213> Torque teno virus

<400> 37

```
Met Ala Tyr Trp Phe Arg Arg Trp Gly Trp Arg Pro Arg Arg Arg Trp
1               5                   10                  15

Arg Arg Trp Arg Arg Arg Arg Arg Arg Leu Pro Arg Arg Arg Thr Arg
            20                  25                  30

Arg Ala Val Arg Gly Leu Gly Arg Arg Arg Lys Pro Arg Val Val Lys
        35                  40                  45

Asn Pro Cys Gly Asp Ser Gly Ile Val Pro Gly Ser Gly Arg Phe Thr
    50                  55                  60

Arg Glu Val Gln Val Val Ser Pro Leu Ser Met Gly Pro Ala Tyr Ile
65                  70                  75                  80
```

```
Phe His Tyr Phe Asp Ser Arg Arg Gly Phe Phe Ser Glu Lys Ala Leu
                85                  90                  95

Lys Arg Met Gln Gln Gln Gln Glu Phe Asp Glu Ser Phe Thr Phe Lys
            100                 105                 110

Pro Lys Arg Pro Lys Leu Ser Thr Ala Ala Ala Glu Ile Leu Gln Leu
        115                 120                 125

Glu Glu Asp Ser Thr Ser Gly Glu Gly Lys Ser Pro Leu Gln Gln Glu
    130                 135                 140

Glu Lys Glu Val Glu Val Leu Gln Thr Pro Thr Val Gln Leu Gln Leu
145                 150                 155                 160

Gln Arg Asn Ile Gln Glu Gln Leu Ala Ile Lys Gln Gln Leu Gln Phe
                165                 170                 175

Leu Leu Leu Gln Leu Leu Lys Thr Gln Ser Asn Leu His Leu Asn Pro
            180                 185                 190

Gln Phe Leu Ser Pro Ser
        195
```

<210> 38
<211> 114
<212> PRT
<213> Torque teno virus

<400> 38

```
Met Ala Tyr Trp Phe Arg Arg Trp Gly Trp Arg Pro Arg Arg Arg Trp
1               5                   10                  15

Arg Arg Trp Arg Arg Arg Arg Arg Arg Leu Pro Arg Arg Arg Thr Arg
            20                  25                  30

Arg Ala Val Arg Gly Leu Gly Arg Arg Arg Lys Pro Arg Gln Pro Gln
        35                  40                  45

Lys Ser Ser Ser Ser Lys Lys Thr Arg Leu Gln Gly Lys Glu Asn Arg
    50                  55                  60
```

```
His Tyr Ser Lys Lys Arg Lys Lys Ser Lys Ser Ser Lys Arg Arg Gln
65                  70                  75                  80

Tyr Ser Ser Ser Ser Ser Glu Thr Ser Arg Ser Ser Ser Gln Ser Ser
                85                  90                  95

Ser Ser Ser Asn Ser Ser Cys Ser Asn Ser Ser Lys Pro Asn Pro Ile
            100                 105                 110

Cys Ile
```

<210> 39
<211> 3794
<212> DNA
<213> Torque teno virus

<400> 39

```
cccgaagtcc gtcactaacc acgtgactcc tgtcgcccaa tcagagtgta tgtcgtgcat       60

ttcctgggca tggtctacat cctgatataa ctaagtgcac ttccgaatgg ctgagttttc      120

cacgcccgtc cgcagcgagg gagcgacgga ggagctcccg agcgtcccga gggcgggtgc      180

cggaggtgag tttacacacc gcagtcaagg ggcaattcgg gctcgggact ggccgggcta      240

tgggcaaggc tcttagggtc ttcattctta atatgtttct tggcagagtt taccgccaca      300

agaaaaggaa agtgctactg tccacactgc gagctccaca ggcgtctcgc agggctatga      360

gttggcgacc cccggtacac gatgcacccg gcatcgagcg caattggtac gaggcctgtt      420

tcagagccca cgctggagct tgtggctgtg gcaattttat tatgcacctt aatcttttgg      480

ctgggcgtta tggttttact ccggggtcag cgccgccagg tggtcctcct ccgggcaccc      540

cgcagataag gagagccagg cctagtcccg ccgcaccaga gcagcccgct gccctaccat      600

ggcatgggga tggtggagat ggcggcgccg ctggcccgcc agacgctgga ggagacgccg      660

tcgccggcgc cccgtacgga gaacaagagc tcgccgacct gctcgacgct atagaagacg      720

acgaacagta agaaccaggc gaaggcggtg ggggcgcaga cggtacagac ggggctggag      780

acgcaggact tatgtgagaa aggggcgaca cagaaaaaag aaaaagagac tgatactgag      840

acagtggcaa ccagccacaa gacgcagatg taccataact gggtacctgc ccatagtgtt      900

ctgcggccac actaggggca ataaaaacta tgcactacac tctgacgact acacccccca      960
```

aggacaacca tttggagggg ctctaagcac tacctcattc tctttaaaag tactatttga 1020

ccagcatcag agaggactaa acaagtggtc ttttccaaac gaccaactag acctcgccag 1080

atatagaggc tgcaaattta tattttatag aacaaaacaa actgactggg tgggccagta 1140

tgacatatca gaaccctaca agctagacaa atacagctgc cccaactatc accctggaaa 1200

catgattaag gcaaagcaca aatttttaat accaagctat gacactaatc ctagaggcag 1260

acaaaaaatt atagttaaaa ttccccccc agacctcttt gtagacaagt ggtacactca 1320

agaggatctg tgttccgtta atcttgtgtc acttgcggtt tctgcggctt cctttctcca 1380

cccattcggc tcaccacaaa ctgacaaccc ttgctacacc ttccaggtgt tgaaagagtt 1440

ctactatcag gcaataggct tctctgcaag cacacaagca atgacatcag tattagacac 1500

gctatacaca caaaacagtt attgggaatc taatctaact cagttttatg tacttaatgc 1560

aaaaaaaggc agtgatacaa cacagccttt aactagcaat atgccaactc gtgaagagtt 1620

tatggcaaaa aaaaatacca attacaactg gtatacatac aaggccgcgt cagtaaaaaa 1680

taaactacat caaatgagac aaacctattt tgaggagtta acctctaagg ggccacaaac 1740

aacaaaaagt gaggaaggct acagtcagca ctggaccacc ccctccacaa acgcctacga 1800

atatcactta ggaatgttta gtgcaatatt tctagcccca gacaggccag tacctagatt 1860

tccatgcgcc taccaagatg taacttacaa ccccttaatg gacaaagggg tgggaaacca 1920

catttggttt cagtacaaca caaaggcaga cactcagcta atagtcacag gagggtcctg 1980

caaagcacac atacaagaca taccactgtg ggcggccttc tatggataca gtgactttat 2040

agagtcagaa ctaggcccct ttgtagatgc agagacggta ggcttagtgt gtgtaatatg 2100

cccttataca aaacccccca tgtacaacaa gacaaacccc gccatgggct acgtgttcta 2160

tgacagaaac tttggtgacg gaaaatggac tgacggacgg ggcaaaatag agccctactg 2220

gcaagttagg tggaggcccg aaatgctttt ccaagaaact gtaatggcag acctagttca 2280

gactgggccc tttagctaca aagacgaact taaaaacagc accctagtgt gcaagtacaa 2340

attctatttc acctggggag gtaacatgat gttccaacag acgatcaaaa acccgtgcaa 2400

gacggacgga caacccaccg actccagtag acaccctaga ggaatacaag tggcggaccc 2460

ggaacaaatg ggaccccgct gggtgttcca ctcctttgac tggcgaaggg gctatcttag 2520

cgagaaagct ctcaaacgcc tgcaagaaaa acctcttgac tatgacgaat attttacaca 2580

accaaaaaga cctagaatct ttcctccaac agaatcagca gagggagagt tccgagagcc 2640

cgaaaaaggc tcgtattcag aggaagaaag gtcgcaagcc tctgccgaag agcagacgca 2700

ggaggcgaca gtactcctcc tcaagcgacg actcagagag caacagcagc tccagcagca 2760

gctccaattc ctcacccgag aaatgttcaa aacgcaagcg ggtctccacc taaaccctat 2820

gttattaaac cagcgataaa ccaagtgtac ctgtttccag agagggcccc aaaaccccct 2880

cctagcagcc aagactggca gcaggagtac gaggcctgcg cagcctggga caggcccccc 2940

agatacaatc tgtcctctcc tcctttctac cccagctgcc cttcaaaatt ctgtgtaaaa 3000

ttcagccttg gctttaaata aatggcaact ttactgtgca aggccgtggg agtttcactg 3060

gtcggtgtct acctctaaag gtcactaagc actccgagcg ttagcgagga gtgcgaccct 3120

tccccctgac tcaacttctt cggagccgcg cgctacgcct tcggctgcgc gcggcacctc 3180

agacccccgc tcgtgctgac acgctcgcgc gtgtcagacc acttcgggct cgcgggggtc 3240

gggaattttg ctaaacagac tccgagttgc tcttggacac tgaggggggca tatcagtaac 3300

gaaagtgagt ggggccagac ttcgccataa ggcctttatc ttcttgccat tggatagtat 3360

cgagggttgc cataggcttc gacctccatt ttaggccttc cggactacaa aaatggccgt 3420

tttagtgacg tcacggccgc cattttaagt aaggcggaag cagctcggcg tacacaaaat 3480

ggcggcggag cacttccggc ttgcccaaaa tggtgggcaa cttcttccgg gtcaaaggtc 3540

acagctacgt cacaagtcac gtggggaggg ttggcgttta acccggaagc caatcctctt 3600

acgtggcctg tcacgtgact tgtacgtcac gaccaccatt ttgttttaca aaatggccga 3660

cttccttcct ctttttttaaa aataacggtt cggcggcggc gcgcgcgcta cgcgcgcgcg 3720

ccgggggggct gccgcccccc ccccgcgcat gcgcggggcc ccccccgcgg gggggctccg 3780

ccccccggcc cccc 3794

<210> 40
<211> 124
<212> PRT
<213> Torque teno virus

<400> 40
Met Ser Trp Arg Pro Pro Val His Asp Ala Pro Gly Ile Glu Arg Asn
1 5 10 15

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

Asn Phe Ile Met His Leu Asn Leu Leu Ala Gly Arg Tyr Gly Phe Thr
        35                  40                  45

Pro Gly Ser Ala Pro Pro Gly Gly Pro Pro Pro Gly Thr Pro Gln Ile
    50                  55                  60

Arg Arg Ala Arg Pro Ser Pro Ala Ala Pro Glu Gln Pro Ala Ala Leu
65                  70                  75                  80

Pro Trp His Gly Asp Gly Gly Asp Gly Gly Ala Ala Gly Pro Pro Asp
                85                  90                  95

Ala Gly Gly Asp Ala Val Ala Gly Ala Pro Tyr Gly Glu Gln Glu Leu
            100                 105                 110

Ala Asp Leu Leu Asp Ala Ile Glu Asp Asp Glu Gln
        115                 120
```

<210> 41
<211> 267
<212> PRT
<213> Torque teno virus

<400> 41

```
Met Ser Trp Arg Pro Pro Val His Asp Ala Pro Gly Ile Glu Arg Asn
1               5                   10                  15

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

Asn Phe Ile Met His Leu Asn Leu Leu Ala Gly Arg Tyr Gly Phe Thr
        35                  40                  45

Pro Gly Ser Ala Pro Pro Gly Gly Pro Pro Pro Gly Thr Pro Gln Ile
    50                  55                  60

Arg Arg Ala Arg Pro Ser Pro Ala Ala Pro Glu Gln Pro Ala Ala Leu
65                  70                  75                  80
```

```
Pro Trp His Gly Asp Gly Gly Asp Gly Gly Ala Ala Gly Pro Pro Asp
                85                  90                  95

Ala Gly Gly Asp Ala Val Ala Gly Ala Pro Tyr Gly Glu Gln Glu Leu
            100                 105                 110

Ala Asp Leu Leu Asp Ala Ile Glu Asp Asp Glu Gln Arg Ser Lys Thr
        115                 120                 125

Arg Ala Arg Arg Thr Asp Asn Pro Pro Thr Pro Val Asp Thr Leu Glu
    130                 135                 140

Glu Tyr Lys Trp Arg Thr Arg Asn Lys Trp Asp Pro Ala Gly Cys Ser
145                 150                 155                 160

Thr Pro Leu Thr Gly Glu Gly Ala Ile Leu Ala Arg Lys Leu Ser Asn
                165                 170                 175

Ala Cys Lys Lys Asn Leu Leu Thr Met Thr Asn Ile Leu His Asn Gln
            180                 185                 190

Lys Asp Leu Glu Ser Phe Leu Gln Gln Asn Gln Gln Arg Glu Ser Ser
        195                 200                 205

Glu Ser Pro Lys Lys Ala Arg Ile Gln Arg Lys Lys Gly Arg Lys Pro
    210                 215                 220

Leu Pro Lys Ser Arg Arg Arg Arg Arg Gln Tyr Ser Ser Ser Ser Asp
225                 230                 235                 240

Asp Ser Glu Ser Asn Ser Ser Ser Ser Ser Ser Ser Asn Ser Ser Pro
                245                 250                 255

Glu Lys Cys Ser Lys Arg Lys Arg Val Ser Thr
            260                 265
```

<210> 42
<211> 257
<212> PRT
<213> Torque teno virus

<400> 42

```
Met Ser Trp Arg Pro Pro Val His Asp Ala Pro Gly Ile Glu Arg Asn
1               5                  10                  15

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

Asn Phe Ile Met His Leu Asn Leu Leu Ala Gly Arg Tyr Gly Phe Thr
        35                  40                  45

Pro Gly Ser Ala Pro Pro Gly Gly Pro Pro Pro Gly Thr Pro Gln Ile
    50                  55                  60

Arg Arg Ala Arg Pro Ser Pro Ala Ala Pro Glu Gln Pro Ala Ala Leu
65                  70                  75                  80

Pro Trp His Gly Asp Gly Gly Asp Gly Gly Ala Ala Gly Pro Pro Asp
                85                  90                  95

Ala Gly Gly Asp Ala Val Ala Gly Ala Pro Tyr Gly Glu Gln Glu Leu
            100                 105                 110

Ala Asp Leu Leu Asp Ala Ile Glu Asp Asp Glu His Arg Gly Arg Val
        115                 120                 125

Pro Arg Ala Arg Lys Arg Leu Val Phe Arg Gly Arg Lys Val Ala Ser
    130                 135                 140

Leu Cys Arg Arg Ala Asp Ala Gly Gly Asp Ser Thr Pro Pro Gln Ala
145                 150                 155                 160

Thr Thr Gln Arg Ala Thr Ala Ala Pro Ala Ala Ala Pro Ile Pro His
                165                 170                 175

Pro Arg Asn Val Gln Asn Ala Ser Gly Ser Pro Pro Lys Pro Tyr Val
            180                 185                 190

Ile Lys Pro Ala Ile Asn Gln Val Tyr Leu Phe Pro Glu Arg Ala Pro
        195                 200                 205

Lys Pro Pro Pro Ser Ser Gln Asp Trp Gln Gln Glu Tyr Glu Ala Cys
    210                 215                 220
```

```
Ala Ala Trp Asp Arg Pro Pro Arg Tyr Asn Leu Ser Ser Pro Pro Phe
225                 230                 235                 240

Tyr Pro Ser Cys Pro Ser Lys Phe Cys Val Lys Phe Ser Leu Gly Phe
                245                 250                 255

Lys
```

<210> 43
<211> 150
<212> PRT
<213> Torque teno virus

<400> 43

```
Met Ser Trp Arg Pro Pro Val His Asp Ala Pro Gly Ile Glu Arg Asn
1               5                   10                  15

Cys Arg Gly Arg Val Pro Arg Ala Arg Lys Arg Leu Val Phe Arg Gly
            20                  25                  30

Arg Lys Val Ala Ser Leu Cys Arg Arg Ala Asp Ala Gly Gly Asp Ser
        35                  40                  45

Thr Pro Pro Gln Ala Thr Thr Gln Arg Ala Thr Ala Ala Pro Ala Ala
    50                  55                  60

Ala Pro Ile Pro His Pro Arg Asn Val Gln Asn Ala Ser Gly Ser Pro
65                  70                  75                  80

Pro Lys Pro Tyr Val Ile Lys Pro Ala Ile Asn Gln Val Tyr Leu Phe
                85                  90                  95

Pro Glu Arg Ala Pro Lys Pro Pro Pro Ser Ser Gln Asp Trp Gln Gln
            100                 105                 110

Glu Tyr Glu Ala Cys Ala Ala Trp Asp Arg Pro Pro Arg Tyr Asn Leu
        115                 120                 125

Ser Ser Pro Pro Phe Tyr Pro Ser Cys Pro Ser Lys Phe Cys Val Lys
    130                 135                 140
```

```
Phe Ser Leu Gly Phe Lys
145                 150


<210> 44
<211> 746
<212> PRT
<213> Torque teno virus

<400> 44
Met Ala Trp Gly Trp Trp Arg Trp Arg Arg Arg Trp Pro Ala Arg Arg
1               5                   10                  15


Trp Arg Arg Arg Arg Arg Arg Arg Pro Val Arg Arg Thr Arg Ala Arg
            20                  25                  30


Arg Pro Ala Arg Arg Tyr Arg Arg Arg Arg Thr Val Arg Thr Arg Arg
        35                  40                  45


Arg Arg Trp Gly Arg Arg Arg Tyr Arg Arg Gly Trp Arg Arg Arg Thr
    50                  55                  60


Tyr Val Arg Lys Gly Arg His Arg Lys Lys Lys Lys Arg Leu Ile Leu
65                  70                  75                  80


Arg Gln Trp Gln Pro Ala Thr Arg Arg Arg Cys Thr Ile Thr Gly Tyr
                85                  90                  95


Leu Pro Ile Val Phe Cys Gly His Thr Arg Gly Asn Lys Asn Tyr Ala
            100                 105                 110


Leu His Ser Asp Asp Tyr Thr Pro Gln Gly Gln Pro Phe Gly Gly Ala
        115                 120                 125


Leu Ser Thr Thr Ser Phe Ser Leu Lys Val Leu Phe Asp Gln His Gln
    130                 135                 140


Arg Gly Leu Asn Lys Trp Ser Phe Pro Asn Asp Gln Leu Asp Leu Ala
145                 150                 155                 160


Arg Tyr Arg Gly Cys Lys Phe Ile Phe Tyr Arg Thr Lys Gln Thr Asp
                165                 170                 175
```

```
Trp Val Gly Gln Tyr Asp Ile Ser Glu Pro Tyr Lys Leu Asp Lys Tyr
            180                 185                 190

Ser Cys Pro Asn Tyr His Pro Gly Asn Met Ile Lys Ala Lys His Lys
        195                 200                 205

Phe Leu Ile Pro Ser Tyr Asp Thr Asn Pro Arg Gly Arg Gln Lys Ile
    210                 215                 220

Ile Val Lys Ile Pro Pro Pro Asp Leu Phe Val Asp Lys Trp Tyr Thr
225                 230                 235                 240

Gln Glu Asp Leu Cys Ser Val Asn Leu Val Ser Leu Ala Val Ser Ala
                245                 250                 255

Ala Ser Phe Leu His Pro Phe Gly Ser Pro Gln Thr Asp Asn Pro Cys
            260                 265                 270

Tyr Thr Phe Gln Val Leu Lys Glu Phe Tyr Tyr Gln Ala Ile Gly Phe
        275                 280                 285

Ser Ala Ser Thr Gln Ala Met Thr Ser Val Leu Asp Thr Leu Tyr Thr
    290                 295                 300

Gln Asn Ser Tyr Trp Glu Ser Asn Leu Thr Gln Phe Tyr Val Leu Asn
305                 310                 315                 320

Ala Lys Lys Gly Ser Asp Thr Thr Gln Pro Leu Thr Ser Asn Met Pro
                325                 330                 335

Thr Arg Glu Glu Phe Met Ala Lys Lys Asn Thr Asn Tyr Asn Trp Tyr
            340                 345                 350

Thr Tyr Lys Ala Ala Ser Val Lys Asn Lys Leu His Gln Met Arg Gln
        355                 360                 365

Thr Tyr Phe Glu Glu Leu Thr Ser Lys Gly Pro Gln Thr Thr Lys Ser
    370                 375                 380
```

```
Glu Glu Gly Tyr Ser Gln His Trp Thr Thr Pro Ser Thr Asn Ala Tyr
385                 390                 395                 400

Glu Tyr His Leu Gly Met Phe Ser Ala Ile Phe Leu Ala Pro Asp Arg
                405                 410                 415

Pro Val Pro Arg Phe Pro Cys Ala Tyr Gln Asp Val Thr Tyr Asn Pro
            420                 425                 430

Leu Met Asp Lys Gly Val Gly Asn His Ile Trp Phe Gln Tyr Asn Thr
        435                 440                 445

Lys Ala Asp Thr Gln Leu Ile Val Thr Gly Gly Ser Cys Lys Ala His
    450                 455                 460

Ile Gln Asp Ile Pro Leu Trp Ala Ala Phe Tyr Gly Tyr Ser Asp Phe
465                 470                 475                 480

Ile Glu Ser Glu Leu Gly Pro Phe Val Asp Ala Glu Thr Val Gly Leu
                485                 490                 495

Val Cys Val Ile Cys Pro Tyr Thr Lys Pro Pro Met Tyr Asn Lys Thr
            500                 505                 510

Asn Pro Ala Met Gly Tyr Val Phe Tyr Asp Arg Asn Phe Gly Asp Gly
        515                 520                 525

Lys Trp Thr Asp Gly Arg Gly Lys Ile Glu Pro Tyr Trp Gln Val Arg
    530                 535                 540

Trp Arg Pro Glu Met Leu Phe Gln Glu Thr Val Met Ala Asp Leu Val
545                 550                 555                 560

Gln Thr Gly Pro Phe Ser Tyr Lys Asp Glu Leu Lys Asn Ser Thr Leu
                565                 570                 575

Val Cys Lys Tyr Lys Phe Tyr Phe Thr Trp Gly Gly Asn Met Met Phe
            580                 585                 590

Gln Gln Thr Ile Lys Asn Pro Cys Lys Thr Asp Gly Gln Pro Thr Asp
        595                 600                 605
```

```
Ser Ser Arg His Pro Arg Gly Ile Gln Val Ala Asp Pro Glu Gln Met
    610                 615                 620

Gly Pro Arg Trp Val Phe His Ser Phe Asp Trp Arg Arg Gly Tyr Leu
625                 630                 635                 640

Ser Glu Lys Ala Leu Lys Arg Leu Gln Glu Lys Pro Leu Asp Tyr Asp
                645                 650                 655

Glu Tyr Phe Thr Gln Pro Lys Arg Pro Arg Ile Phe Pro Pro Thr Glu
            660                 665                 670

Ser Ala Glu Gly Glu Phe Arg Glu Pro Glu Lys Gly Ser Tyr Ser Glu
        675                 680                 685

Glu Glu Arg Ser Gln Ala Ser Ala Glu Glu Gln Thr Gln Glu Ala Thr
    690                 695                 700

Val Leu Leu Leu Lys Arg Arg Leu Arg Glu Gln Gln Gln Leu Gln Gln
705                 710                 715                 720

Gln Leu Gln Phe Leu Thr Arg Glu Met Phe Lys Thr Gln Ala Gly Leu
                725                 730                 735

His Leu Asn Pro Met Leu Leu Asn Gln Arg
            740                 745
```

<210> 45
<211> 195
<212> PRT
<213> Torque teno virus

<400> 45

```
Met Ala Trp Gly Trp Trp Arg Trp Arg Arg Arg Trp Pro Ala Arg Arg
1               5                   10                  15

Trp Arg Arg Arg Arg Arg Arg Arg Pro Val Arg Arg Thr Arg Ala Arg
            20                  25                  30

Arg Pro Ala Arg Arg Tyr Arg Arg Arg Arg Thr Thr Ile Lys Asn Pro
        35                  40                  45
```

```
Cys Lys Thr Asp Gly Gln Pro Thr Asp Ser Ser Arg His Pro Arg Gly
    50                  55                  60

Ile Gln Val Ala Asp Pro Glu Gln Met Gly Pro Arg Trp Val Phe His
65                  70                  75                  80

Ser Phe Asp Trp Arg Arg Gly Tyr Leu Ser Glu Lys Ala Leu Lys Arg
                85                  90                  95

Leu Gln Glu Lys Pro Leu Asp Tyr Asp Glu Tyr Phe Thr Gln Pro Lys
            100                 105                 110

Arg Pro Arg Ile Phe Pro Pro Thr Glu Ser Ala Glu Gly Glu Phe Arg
        115                 120                 125

Glu Pro Glu Lys Gly Ser Tyr Ser Glu Glu Glu Arg Ser Gln Ala Ser
    130                 135                 140

Ala Glu Glu Gln Thr Gln Glu Ala Thr Val Leu Leu Leu Lys Arg Arg
145                 150                 155                 160

Leu Arg Glu Gln Gln Gln Leu Gln Gln Gln Leu Gln Phe Leu Thr Arg
                165                 170                 175

Glu Met Phe Lys Thr Gln Ala Gly Leu His Leu Asn Pro Met Leu Leu
            180                 185                 190

Asn Gln Arg
        195
```

<210> 46
<211> 107
<212> PRT
<213> Torque teno virus

<400> 46

```
Met Ala Trp Gly Trp Trp Arg Trp Arg Arg Arg Trp Pro Ala Arg Arg
1               5                   10                  15

Trp Arg Arg Arg Arg Arg Arg Arg Pro Val Arg Arg Thr Arg Ala Arg
            20                  25                  30
```

```
Arg Pro Ala Arg Arg Tyr Arg Arg Arg Arg Thr Gln Arg Glu Ser Ser
        35                  40                  45

Glu Ser Pro Lys Lys Ala Arg Ile Gln Arg Lys Lys Gly Arg Lys Pro
    50                  55                  60

Leu Pro Lys Ser Arg Arg Arg Arg Arg Gln Tyr Ser Ser Ser Ser Asp
65                  70                  75                  80

Asp Ser Glu Ser Asn Ser Ser Ser Ser Ser Ser Ser Asn Ser Ser Pro
                85                  90                  95

Glu Lys Cys Ser Lys Arg Lys Arg Val Ser Thr
            100                 105
```

<210> 47
<211> 3866
<212> DNA
<213> Torque teno virus

<400> 47

```
aagtccgtca ctaaccacgt gactcccgca ggccaatcag agtctatgtc gtgcacttcc       60

tgggcatggt ctacgttctc atataactaa ctgcacttcc gaatggctga gttttccacg      120

cccgtccgca gcggcagcac cacggagggt gatccccgcg tcccgagggc gggtgccgaa      180

ggtgagttta cacaccgcag tcaaggggca attcgggctc gggactggcc gggctatggg      240

caaggctctt agggctttca ttgttaaaaa tgtttctcgg caggccttac aggagaaaga      300

aaagggcgct gtcactgcct ggcgtgcgag ctgcacaggc gaaacaacct ggtgatatga      360

gctggagccg tccagtacat aatgccgccg ggatcgaaag gcagtggttc gaatccacct      420

ttagatccca cgctagttgc tgtggctgcg gcaattttgt taatcatatt aatgtactgg      480

ctgctcgcta cggctttact ggggggccga cgccgccagg tggtcctggg ccgcgtccac      540

aactgaggcc cgcgcttccc gcgccggacc ccgaccccca ggcgcccaac cgtgagccat      600

ggcgtggagc tggtggtggc aacgatggag aaggcgccgc tggaaaccca ggaggcgccg      660

ctggagacgt ctacgatgga gaagacctag acgcgctgtt cgccgccgtc gtcgaggacg      720

tagagtaagg aggcggaggt gggcgcgtag acgggggcga cgcagacggt acgccaccag      780
```

acgaaagaga cgttataggg gtcgccgctt taaaaagaaa ctagtactga ctcagtggca 840

ccctaatacc atgagacgct gcttaatcaa gggcatagtc cccctggtaa tatgcggcca 900

caccaggtgg aactacaact acgccctcca tagcaaggac tacacagagg agggtcgcta 960

ccctcacggg ggggccctca gcaccactac gtggtccctt aaggtgctgt atgacgagca 1020

cctcaaacac cacgacttct ggggctatcc caacaaccag ctagacctgg ccaggtacaa 1080

gggggccaag ttcaccttct acagacacaa aaagactgac tttataatat tctttaacag 1140

aaagcctccc tttaagctaa acaagtacag ctgtgcctcc tatcacccag gcatgctgat 1200

gcagcagaga cacaagatcc tgctacccag ctacgaaact aaacccaagg gcaggccaaa 1260

gataacagtt agaataaagc cccccactct gttagaggac aagtggtaca cccagcagga 1320

cctgtgcgac gttaacctgt tgcaacttgt ggtcactgcg gctgactttc gacatccact 1380

ctgctcacca caaacgaaca ctccaaccac aaccttccag gtgttgaaag acatctatta 1440

tgacactatg agcatatctg aacccacaga ctcctacact agtgttaaca ataaaagtac 1500

aacacaaact tttactaact actcaaacac cttagaaaac attctgtaca cacgagcctc 1560

ctactggaac tcgttccacg ccactgaata cctaaacccc aacatcatat acaaaaacgg 1620

tgaaaaacta ttcaaagaac atgaagactt aataacctgg atgacccaaa ctaacaatac 1680

cgggtttcta actaaaaaca acacagcttt tggcaacaac agctacaggc ccaatgcaga 1740

caaaattaaa aaagccagaa agacatactg gaacgcccta ataggcacca acgacctggc 1800

cactaatata ggccaggcca gagcagaaag gttcgagtac cacctaggct ggtactcccc 1860

catatttctc agcagacaca ggagcaacat gaactttgcc agggcctacc aagacgtcac 1920

atacaacccc aactgtgaca ggggagttaa caacagggtg tgggttcagc ctctaactaa 1980

acccaccaca gagttcgacg agaaaaggtg taagtgcgta gtgcagcacc tgcctctgtg 2040

ggcggctctg tactgctacc aagactttgt agaggaggag ctggggtcct cctcagagat 2100

attaaattca tgcctactgg tattacagtg cccttacacc tttcccccaa tgtatgacaa 2160

aaagctacca gacaagggat tcgtgtttta tgactccctt tttggagacg gcaaaatgtc 2220

tgacggacgc ggacaggtgg acattttctg gcaacagcga tggtaccctc gcttagccac 2280

tcagatgcaa gtcatgcacg acatcaccat gacgggcccc ttctcctacc gagacgagct 2340

agttagcacc caactgactg ccaagtacac ctttgacttt atgtggggcg gaaatatgat 2400

ctccacacag atcatcaaga acccctgcaa agacagtgga ctggaacccg cctaccccgg 2460

tagacagcgt cgcgacttac aaattgttga cccatactcc atgggccccc aattctcgtt 2520

ccacaactgg gactacagac atggcctttt tggccaagac gctatcgaca gagtgtctaa 2580

acaaccaaaa gatgatgcag actatcctaa cccatacaaa aggcctagat attttccacc 2640

cacagaccaa gccgcccaag agcaagaaaa agacttcagt ttcctcaaaa cagcaccgtc 2700

gaactcagaa gagagcgatc aagaagtcct ccaagaaacg caagtactcc gattccagcc 2760

agagcagcac aagcaactcc acctgcagct cgcagagcgg cagcgaatcg gagagcaact 2820

ccgataccta ctccaacaga tgttcaaaac tcaggccaat ctccacctaa acccatatac 2880

atttacccag ctgtaaagca ggtgtttatg tttgaccccc cgggccctaa ggctatctcg 2940

ggcgccaagg cctgggagga cgagttcctc accgcaaaag tgtggaaccg cccggtacgc 3000

aagtactact cagacacccc ctactacccc tgggccccca aaccccagta ctctgtcagt 3060

ttcaaactcg gctggaaata aaaaaagcct gctccactgt actaggccgt gggagtttca 3120

ctcgtcggtg tctacctctt aaggtcacca agcactccga gcgtcagcga ggagtgcgac 3180

ccttgggggt gggtgcaacg ccctcggcgg ccgcgcgcta cgccttcggc tgcgcgcggc 3240

acctcggacc cccgctcgtg ctgacgcgct tgcgcgcgtc agaccacttc gggctcgcgg 3300

gggtcggaaa ttttgctaaa cagactccga gttgccattg gacactggag ccgtgaatca 3360

gtaacgaaag tgagtggggc cagacttcgc cataaggcct ttatcttttt gccatttgtc 3420

cgtggggaag ggtcgctgca agcgcggacc ccgttttcac cccttccgga ctacaaaaat 3480

agcgcattag tgacgtcacg gccgccattt taagtaaggc ggaagcaact ccactttctc 3540

acaaaatggc ggcggagcac ttccggcttg cccaaaatgg ccgccaaaaa catccgggtc 3600

aaagttcgcc gctacgtcat aagtcacgtg actggggagg tacttaaaca cggaagtatc 3660

ctcaaccacg taactggtca cgtggtgcgc acgtcacggc aaccattttg ttttacaaaa 3720

tggcgcattt ccttcctctt ttttaaaaat taaccgttgg cggcggcgcg cgcgctacgc 3780

gcgcgcgccg gggagctctg ccccccccgg cgcatgcgcg cgggtccccc ccccgcgggg 3840

ggctccgccc cccggtcccc cccccg 3866

<210> 48
<211> 123
<212> PRT

<213> Torque teno virus

<400> 48
Met Ser Trp Ser Arg Pro Val His Asn Ala Ala Gly Ile Glu Arg Gln
1   5   10   15

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

Asn Phe Val Asn His Ile Asn Val Leu Ala Ala Arg Tyr Gly Phe Thr
        35   40   45

Gly Gly Pro Thr Pro Pro Gly Gly Pro Gly Pro Arg Pro Gln Leu Arg
    50   55   60

Pro Ala Leu Pro Ala Pro Asp Pro Asp Pro Gln Ala Pro Asn Arg Glu
65   70   75   80

Pro Trp Arg Gly Ala Gly Gly Gly Asn Asp Gly Glu Gly Ala Ala Gly
                85   90   95

Asn Pro Gly Gly Ala Ala Gly Asp Val Tyr Asp Gly Glu Asp Leu Asp
            100   105   110

Ala Leu Phe Ala Ala Val Val Glu Asp Val Glu
        115   120

<210> 49
<211> 275
<212> PRT
<213> Torque teno virus

<400> 49
Met Ser Trp Ser Arg Pro Val His Asn Ala Ala Gly Ile Glu Arg Gln
1   5   10   15

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

Asn Phe Val Asn His Ile Asn Val Leu Ala Ala Arg Tyr Gly Phe Thr
        35   40   45

Gly Gly Pro Thr Pro Pro Gly Gly Pro Gly Pro Arg Pro Gln Leu Arg

```
    50                  55                  60

Pro Ala Leu Pro Ala Pro Asp Pro Asp Pro Gln Ala Pro Asn Arg Glu
65                  70                  75                  80

Pro Trp Arg Gly Ala Gly Gly Gly Asn Asp Gly Glu Gly Ala Ala Gly
                85                  90                  95

Asn Pro Gly Gly Ala Ala Gly Asp Val Tyr Asp Gly Glu Asp Leu Asp
            100                 105                 110

Ala Leu Phe Ala Ala Val Val Glu Asp Val Glu Ser Ser Arg Thr Pro
        115                 120                 125

Ala Lys Thr Val Asp Trp Asn Pro Pro Thr Pro Val Asp Ser Val Ala
    130                 135                 140

Thr Tyr Lys Leu Leu Thr His Thr Pro Trp Ala Pro Asn Ser Arg Ser
145                 150                 155                 160

Thr Thr Gly Thr Thr Asp Met Ala Phe Leu Ala Lys Thr Leu Ser Thr
                165                 170                 175

Glu Cys Leu Asn Asn Gln Lys Met Met Gln Thr Ile Leu Thr His Thr
            180                 185                 190

Lys Gly Leu Asp Ile Phe His Pro Gln Thr Lys Pro Pro Lys Ser Lys
        195                 200                 205

Lys Lys Thr Ser Val Ser Ser Lys Gln His Arg Arg Thr Gln Lys Arg
    210                 215                 220

Ala Ile Lys Lys Ser Ser Lys Lys Arg Lys Tyr Ser Asp Ser Ser Gln
225                 230                 235                 240

Ser Ser Thr Ser Asn Ser Thr Cys Ser Ser Gln Ser Gly Ser Glu Ser
                245                 250                 255

Glu Ser Asn Ser Asp Thr Tyr Ser Asn Arg Cys Ser Lys Leu Arg Pro
            260                 265                 270
```

Ile Ser Thr
        275

<210> 50
<211> 267
<212> PRT
<213> Torque teno virus

<400> 50
Met Ser Trp Ser Arg Pro Val His Asn Ala Ala Gly Ile Glu Arg Gln
1   5   10   15

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

Asn Phe Val Asn His Ile Asn Val Leu Ala Ala Arg Tyr Gly Phe Thr
        35   40   45

Gly Gly Pro Thr Pro Pro Gly Gly Pro Gly Pro Arg Pro Gln Leu Arg
    50   55   60

Pro Ala Leu Pro Ala Pro Asp Pro Asp Pro Gln Ala Pro Asn Arg Glu
65   70   75   80

Pro Trp Arg Gly Ala Gly Gly Gly Asn Asp Gly Glu Gly Ala Ala Gly
                85   90   95

Asn Pro Gly Gly Ala Ala Gly Asp Val Tyr Asp Gly Glu Asp Leu Asp
            100   105   110

Ala Leu Phe Ala Ala Val Val Glu Asp Val Glu Pro Ser Arg Pro Arg
        115   120   125

Ala Arg Lys Arg Leu Gln Phe Pro Gln Asn Ser Thr Val Glu Leu Arg
    130   135   140

Arg Glu Arg Ser Arg Ser Pro Pro Arg Asn Ala Ser Thr Pro Ile Pro
145   150   155   160

Ala Arg Ala Ala Gln Ala Thr Pro Pro Ala Ala Arg Arg Ala Ala Ala
                165   170   175

```
Asn Arg Arg Ala Thr Pro Ile Pro Thr Pro Thr Asp Val Gln Asn Ser
            180                 185                 190

Gly Gln Ser Pro Pro Lys Pro Ile Tyr Ile Tyr Pro Ala Val Lys Gln
        195                 200                 205

Val Phe Met Phe Asp Pro Pro Gly Pro Lys Ala Ile Ser Gly Ala Lys
    210                 215                 220

Ala Trp Glu Asp Glu Phe Leu Thr Ala Lys Val Trp Asn Arg Pro Val
225                 230                 235                 240

Arg Lys Tyr Tyr Ser Asp Thr Pro Tyr Tyr Pro Trp Ala Pro Lys Pro
                245                 250                 255

Gln Tyr Ser Val Ser Phe Lys Leu Gly Trp Lys
            260                 265
```

<210> 51
<211> 765
<212> PRT
<213> Torque teno virus

<400> 51

```
Met Ala Trp Ser Trp Trp Trp Gln Arg Trp Arg Arg Arg Arg Trp Lys
1               5                   10                  15

Pro Arg Arg Arg Arg Trp Arg Arg Leu Arg Trp Arg Arg Pro Arg Arg
            20                  25                  30

Ala Val Arg Arg Arg Arg Arg Gly Arg Arg Val Arg Arg Arg Arg Trp
        35                  40                  45

Ala Arg Arg Arg Gly Arg Arg Arg Arg Tyr Ala Thr Arg Arg Lys Arg
    50                  55                  60

Arg Tyr Arg Gly Arg Arg Phe Lys Lys Lys Leu Val Leu Thr Gln Trp
65                  70                  75                  80

His Pro Asn Thr Met Arg Arg Cys Leu Ile Lys Gly Ile Val Pro Leu
                85                  90                  95
```

```
Val Ile Cys Gly His Thr Arg Trp Asn Tyr Asn Tyr Ala Leu His Ser
            100                 105                 110

Lys Asp Tyr Thr Glu Glu Gly Arg Tyr Pro His Gly Gly Ala Leu Ser
        115                 120                 125

Thr Thr Thr Trp Ser Leu Lys Val Leu Tyr Asp Glu His Leu Lys His
    130                 135                 140

His Asp Phe Trp Gly Tyr Pro Asn Asn Gln Leu Asp Leu Ala Arg Tyr
145                 150                 155                 160

Lys Gly Ala Lys Phe Thr Phe Tyr Arg His Lys Lys Thr Asp Phe Ile
                165                 170                 175

Ile Phe Phe Asn Arg Lys Pro Pro Phe Lys Leu Asn Lys Tyr Ser Cys
            180                 185                 190

Ala Ser Tyr His Pro Gly Met Leu Met Gln Gln Arg His Lys Ile Leu
        195                 200                 205

Leu Pro Ser Tyr Glu Thr Lys Pro Lys Gly Arg Pro Lys Ile Thr Val
    210                 215                 220

Arg Ile Lys Pro Pro Thr Leu Leu Glu Asp Lys Trp Tyr Thr Gln Gln
225                 230                 235                 240

Asp Leu Cys Asp Val Asn Leu Leu Gln Leu Val Val Thr Ala Ala Asp
                245                 250                 255

Phe Arg His Pro Leu Cys Ser Pro Gln Thr Asn Thr Pro Thr Thr Thr
            260                 265                 270

Phe Gln Val Leu Lys Asp Ile Tyr Tyr Asp Thr Met Ser Ile Ser Glu
        275                 280                 285

Pro Thr Asp Ser Tyr Thr Ser Val Asn Asn Lys Ser Thr Thr Gln Thr
    290                 295                 300

Phe Thr Asn Tyr Ser Asn Thr Leu Glu Asn Ile Leu Tyr Thr Arg Ala
```

```
305                 310                 315                 320

Ser Tyr Trp Asn Ser Phe His Ala Thr Glu Tyr Leu Asn Pro Asn Ile
                325                 330                 335

Ile Tyr Lys Asn Gly Glu Lys Leu Phe Lys Glu His Glu Asp Leu Ile
            340                 345                 350

Thr Trp Met Thr Gln Thr Asn Asn Thr Gly Phe Leu Thr Lys Asn Asn
        355                 360                 365

Thr Ala Phe Gly Asn Asn Ser Tyr Arg Pro Asn Ala Asp Lys Ile Lys
    370                 375                 380

Lys Ala Arg Lys Thr Tyr Trp Asn Ala Leu Ile Gly Thr Asn Asp Leu
385                 390                 395                 400

Ala Thr Asn Ile Gly Gln Ala Arg Ala Glu Arg Phe Glu Tyr His Leu
                405                 410                 415

Gly Trp Tyr Ser Pro Ile Phe Leu Ser Arg His Arg Ser Asn Met Asn
            420                 425                 430

Phe Ala Arg Ala Tyr Gln Asp Val Thr Tyr Asn Pro Asn Cys Asp Arg
        435                 440                 445

Gly Val Asn Asn Arg Val Trp Val Gln Pro Leu Thr Lys Pro Thr Thr
    450                 455                 460

Glu Phe Asp Glu Lys Arg Cys Lys Cys Val Val Gln His Leu Pro Leu
465                 470                 475                 480

Trp Ala Ala Leu Tyr Cys Tyr Gln Asp Phe Val Glu Glu Glu Leu Gly
                485                 490                 495

Ser Ser Ser Glu Ile Leu Asn Ser Cys Leu Leu Val Leu Gln Cys Pro
            500                 505                 510

Tyr Thr Phe Pro Pro Met Tyr Asp Lys Lys Leu Pro Asp Lys Gly Phe
        515                 520                 525
```

```
Val Phe Tyr Asp Ser Leu Phe Gly Asp Gly Lys Met Ser Asp Gly Arg
    530                 535                 540

Gly Gln Val Asp Ile Phe Trp Gln Gln Arg Trp Tyr Pro Arg Leu Ala
545                 550                 555                 560

Thr Gln Met Gln Val Met His Asp Ile Thr Met Thr Gly Pro Phe Ser
                565                 570                 575

Tyr Arg Asp Glu Leu Val Ser Thr Gln Leu Thr Ala Lys Tyr Thr Phe
            580                 585                 590

Asp Phe Met Trp Gly Gly Asn Met Ile Ser Thr Gln Ile Ile Lys Asn
        595                 600                 605

Pro Cys Lys Asp Ser Gly Leu Glu Pro Ala Tyr Pro Gly Arg Gln Arg
    610                 615                 620

Arg Asp Leu Gln Ile Val Asp Pro Tyr Ser Met Gly Pro Gln Phe Ser
625                 630                 635                 640

Phe His Asn Trp Asp Tyr Arg His Gly Leu Phe Gly Gln Asp Ala Ile
                645                 650                 655

Asp Arg Val Ser Lys Gln Pro Lys Asp Asp Ala Asp Tyr Pro Asn Pro
            660                 665                 670

Tyr Lys Arg Pro Arg Tyr Phe Pro Pro Thr Asp Gln Ala Ala Gln Glu
        675                 680                 685

Gln Glu Lys Asp Phe Ser Phe Leu Lys Thr Ala Pro Ser Asn Ser Glu
    690                 695                 700

Glu Ser Asp Gln Glu Val Leu Gln Glu Thr Gln Val Leu Arg Phe Gln
705                 710                 715                 720

Pro Glu Gln His Lys Gln Leu His Leu Gln Leu Ala Glu Arg Gln Arg
                725                 730                 735

Ile Gly Glu Gln Leu Arg Tyr Leu Leu Gln Gln Met Phe Lys Thr Gln
```

```
            740                 745                 750

Ala Asn Leu His Leu Asn Pro Tyr Thr Phe Thr Gln Leu
        755                 760                 765


<210> 52
<211> 203
<212> PRT
<213> Torque teno virus

<400> 52
Met Ala Trp Ser Trp Trp Trp Gln Arg Trp Arg Arg Arg Arg Trp Lys
1               5                   10                  15

Pro Arg Arg Arg Arg Trp Arg Arg Leu Arg Trp Arg Arg Pro Arg Arg
            20                  25                  30

Ala Val Arg Arg Arg Arg Arg Gly Arg Arg Ile Ile Lys Asn Pro Cys
        35                  40                  45

Lys Asp Ser Gly Leu Glu Pro Ala Tyr Pro Gly Arg Gln Arg Arg Asp
    50                  55                  60

Leu Gln Ile Val Asp Pro Tyr Ser Met Gly Pro Gln Phe Ser Phe His
65                  70                  75                  80

Asn Trp Asp Tyr Arg His Gly Leu Phe Gly Gln Asp Ala Ile Asp Arg
                85                  90                  95

Val Ser Lys Gln Pro Lys Asp Asp Ala Asp Tyr Pro Asn Pro Tyr Lys
            100                 105                 110

Arg Pro Arg Tyr Phe Pro Pro Thr Asp Gln Ala Ala Gln Glu Gln Glu
        115                 120                 125

Lys Asp Phe Ser Phe Leu Lys Thr Ala Pro Ser Asn Ser Glu Glu Ser
    130                 135                 140

Asp Gln Glu Val Leu Gln Glu Thr Gln Val Leu Arg Phe Gln Pro Glu
145                 150                 155                 160

Gln His Lys Gln Leu His Leu Gln Leu Ala Glu Arg Gln Arg Ile Gly
```

```
                165                 170                 175

Glu Gln Leu Arg Tyr Leu Leu Gln Gln Met Phe Lys Thr Gln Ala Asn
            180                 185                 190

Leu His Leu Asn Pro Tyr Thr Phe Thr Gln Leu
        195                 200


<210> 53
<211> 116
<212> PRT
<213> Torque teno virus

<400> 53
Met Ala Trp Ser Trp Trp Trp Gln Arg Trp Arg Arg Arg Arg Trp Lys
1               5                   10                  15

Pro Arg Arg Arg Arg Trp Arg Arg Leu Arg Trp Arg Arg Pro Arg Arg
            20                  25                  30

Ala Val Arg Arg Arg Arg Arg Gly Arg Arg Thr Lys Pro Pro Lys Ser
        35                  40                  45

Lys Lys Lys Thr Ser Val Ser Ser Lys Gln His Arg Arg Thr Gln Lys
    50                  55                  60

Arg Ala Ile Lys Lys Ser Ser Lys Lys Arg Lys Tyr Ser Asp Ser Ser
65                  70                  75                  80

Gln Ser Ser Thr Ser Asn Ser Thr Cys Ser Ser Gln Ser Gly Ser Glu
                85                  90                  95

Ser Glu Ser Asn Ser Asp Thr Tyr Ser Asn Arg Cys Ser Lys Leu Arg
            100                 105                 110

Pro Ile Ser Thr
        115


<210> 54
<211> 2979
<212> DNA
<213> TTV-like mini virus
```

<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 gccccttta acccccaggg ggctccgccc 2940

cctgaaaccc ccaaggggggc tacgcccccc tacaccccc 2979

<210> 55
<211> 99
<212> PRT
<213> TTV-like mini virus

<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> TTV-like mini virus

<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> TTV-like mini virus

<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> TTV-like mini virus

<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> TTV-like mini virus

<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> TTV-like mini virus

<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
<211> 3242
<212> DNA
<213> Torque teno midi virus 1

<400> 61
aggtggagac tcttaagcta tataaccaag tggggtggcg aatggctgag tttaccccgc 60

tagacggtgc agggaccgga tcgagcgcag cgaggaggtc cccggctgcc cgtgggcggg 120

agcccgaggt gagtgaaacc accgaggtct aggggcaatt cgggctaggg cagtctagcg 180

gaacgggcaa gaaacttaaa aatatttctt ttacagatgc aaaacctatc agccaaagac 240

ttctacaaac catgcagata caactgtgaa actaaaaacc aaatgtggat gtctggcatt 300

gctgactccc atgacagttg gtgtgactgt gatactcctt ttgctcacct cctggctagt 360

atttttcctc ctggtcacac agatcgcaca cgaaccatcc aagaaatact taccagagat 420

tttaggaaaa catgcctttc tggtggggcc gacgcaacaa attctggtat ggccgaaact 480

atagaagaaa aaagagaaga tttccaaaaa gaagaaaaag aagattttac agaagaacaa 540

aatatagaag acctgctcgc cgccgtcgca gacgcagaag gaaggtaaga agaaaaaaaa 600

aaactcttat agtaagacaa tggcagccag actctattgt actctgtaaa attaaagggt 660

atgactctat aatatgggga gctgaaggca cacagtttca atgttctaca catgaaatgt 720

atgaatatac aagacaaaag taccctgggg gaggaggatt tggtgtacaa ctttacagct 780

tagagtattt gtatgaccaa tggaaactta gaaataatat atggactaaa acaaatcaac 840

tcaaagattt gtgtagatac ttaaaatgtg ttatgacctt ttacagacac caacacatag 900

attttgtaat tgtatatgaa agacaacccc catttgaaat agataaacta acatacatga 960

aatatcatcc atatatgtta ttacaaagaa agcataaaat aattttacct agtcaaacaa 1020

ctaatcctag aggtaaatta aaaaaaaaga aaactattaa acctcccaaa caaatgctca 1080

gcaaatggtt ttttcaacaa caatttgcta aatatgatct actacttatt gctgcagcag 1140

catgtagttt aagataccct agaataggct gctgcaatga aaatagaatg ataaccttat 1200

actgtttaaa tactaaattt tatcaagata cagaatgggg aactacaaaa caggcccccc 1260

actactttaa accatatgca acaattaata aatccatgat atttgtctct aactatggag 1320

gtaaaaaaac agaatataac ataggccaat ggatagaaac agatatacct ggagaaggta 1380

atctagcaag atactacaga tcaataagta aagaaggagg ttacttttca cctaaaatac 1440

tgcaagcata tcaaacaaaa gtaaagtctg tagactacaa acctttacca attgttttag 1500

gtagatataa cccagcaata gatgatggaa aaggcaacaa aatttactta caaactataa 1560

tgaatggcca ttggggccta cctcaaaaaa caccagatta tataatagaa gaggtccctc 1620

tttggctagg cttctgggga tactataact acttaaaaca aacaagaact gaagctatat 1680

ttccactaca catgtttgta gtgcaaagca aatacattca aacacaacaa acagaaacac 1740

ctaacaattt ttgggcattt atagacaaca gctttataca gggcaaaaac ccatgggact 1800

cagttattac ttactcagaa caaaagctat ggtttcctac agttgcatgg caactaaaaa 1860

ccataaatgc tatttgtgaa agtggaccat atgtacctaa actagacaat caaacatata 1920

gtacctggga actagcaact cattactcat ttcactttaa atggggtggt ccacagatat 1980

cagaccaacc agttgaagac ccaggaaaca aaaacaaata tgatgtgccc gatacaatca 2040

aagaagcatt acaaattgtt aacccagcaa aaaacattgc tgccacgatg ttccatgact 2100

gggactacag acggggttgc attacatcaa cagctattaa aagaatgcaa caaaacctcc 2160

caactgattc atctctcgaa tctgattcag actcagaacc agcacccaag aaaaaaagac 2220

tactaccagt cctccacgac ccacaaaaga aaacggaaaa gatcaaccaa tgtctcctct 2280

ctctctgcga agaaagtaca tgccaggagc aggaaacgga ggaaaacatc ctcaagctca 2340

tccagcagca gcagcagcag cagcagaaac tcaagcacaa cctcttagta ctaatcaagg 2400

acttaaaagt gaaacaaaga ttattacaac tacaaacggg ggtactagaa taacccttac 2460

cagatttaaa ccaggatttg agcaagaaac tgaaaaagag ttagcacaag catttaacag 2520

acccccctaga ctgttcaaag aagataaacc cttttacccc tggctaccca gatttacacc 2580

ccttgtaaac tttcacctta attttaaagg ctaggcctac actgctcact tagtggtgta 2640

tgtttattaa agtttgcacc ccagaaaaat tgtaaaataa aaaaaaaaaa aaaaaataaa 2700

aaattgcaaa aattcggcgc tcgcgcgcgc tgcgcgcgcg agcgccgtca cgcgccggcg 2760

ctcgcgcgcc gcgcgtatgt gctaacacac cacgcaccta gattggggtg cgcgcgtagc 2820

gcgcgcaccc caatgcgccc cgccctcgtt ccgacccgct tgcgcgggtc ggaccacttc 2880

gggctcgggg gggcgcgcct gcggcgctta tttactaaac agactccgag tcgccattgg 2940

gccccccta agctccgccc ccctcatgaa tattcataaa ggaaaccaca aaattagaat 3000

tgccgaccac aaactgccat atgctaatta gttccccttt tacacagtaa aaaggggaag 3060

tgggggggca gagccccccc acacccccg cgggggggc agagcccccc ccgcaccccc 3120

cctacgtcac aggccacgcc cccgccgcca tcttgggtgc ggcagggcgg ggactaaaat 3180

ggcgggaccc aatcatttta tactttcact ttccaattaa aacccgccac gtcacacaaa 3240

ag 3242

<210> 62
<211> 101
<212> PRT
<213> Torque teno midi virus 1

<400> 62
Met Trp Met Ser Gly Ile Ala Asp Ser His Asp Ser Trp Cys Asp Cys
1 5 10 15

Asp Thr Pro Phe Ala His Leu Leu Ala Ser Ile Phe Pro Pro Gly His
20 25 30

Thr Asp Arg Thr Arg Thr Ile Gln Glu Ile Leu Thr Arg Asp Phe Arg
35 40 45

Lys Thr Cys Leu Ser Gly Gly Ala Asp Ala Thr Asn Ser Gly Met Ala
50 55 60

Glu Thr Ile Glu Glu Lys Arg Glu Asp Phe Gln Lys Glu Glu Lys Glu
65 70 75 80

Asp Phe Thr Glu Glu Gln Asn Ile Glu Asp Leu Leu Ala Ala Val Ala
85 90 95

Asp Ala Glu Gly Arg
100

<210> 63
<211> 237
<212> PRT
<213> Torque teno midi virus 1

<400> 63
Met Trp Met Ser Gly Ile Ala Asp Ser His Asp Ser Trp Cys Asp Cys
1 5 10 15

Asp Thr Pro Phe Ala His Leu Leu Ala Ser Ile Phe Pro Pro Gly His
20 25 30

Thr Asp Arg Thr Arg Thr Ile Gln Glu Ile Leu Thr Arg Asp Phe Arg
35 40 45

Lys Thr Cys Leu Ser Gly Gly Ala Asp Ala Thr Asn Ser Gly Met Ala
50 55 60

Glu Thr Ile Glu Glu Lys Arg Glu Asp Phe Gln Lys Glu Glu Lys Glu
65 70 75 80

Asp Phe Thr Glu Glu Gln Asn Ile Glu Asp Leu Leu Ala Ala Val Ala
85 90 95

Asp Ala Glu Gly Arg Tyr Gln Thr Asn Gln Leu Lys Thr Gln Glu Thr
100 105 110

Lys Thr Asn Met Met Cys Pro Ile Gln Ser Lys Lys His Tyr Lys Leu
115 120 125

Leu Thr Gln Gln Lys Thr Leu Leu Pro Arg Cys Ser Met Thr Gly Thr
130 135 140

Thr Asp Gly Val Ala Leu His Gln Gln Leu Leu Lys Glu Cys Asn Lys
145 150 155 160

```
Thr Ser Gln Leu Ile His Leu Ser Asn Leu Ile Gln Thr Gln Asn Gln
                165                 170                 175


His Pro Arg Lys Lys Asp Tyr Tyr Gln Ser Ser Thr Thr His Lys Arg
            180                 185                 190


Lys Arg Lys Arg Ser Thr Asn Val Ser Ser Leu Ser Ala Lys Lys Val
        195                 200                 205


His Ala Arg Ser Arg Lys Arg Arg Lys Thr Ser Ser Ser Ser Ser Ser
    210                 215                 220


Ser Ser Ser Ser Ser Ser Arg Asn Ser Ser Thr Thr Ser
225                 230                 235


<210> 64
<211> 239
<212> PRT
<213> Torque teno midi virus 1

<400> 64
Met Trp Met Ser Gly Ile Ala Asp Ser His Asp Ser Trp Cys Asp Cys
1               5                   10                  15


Asp Thr Pro Phe Ala His Leu Leu Ala Ser Ile Phe Pro Pro Gly His
            20                  25                  30


Thr Asp Arg Thr Arg Thr Ile Gln Glu Ile Leu Thr Arg Asp Phe Arg
        35                  40                  45


Lys Thr Cys Leu Ser Gly Gly Ala Asp Ala Thr Asn Ser Gly Met Ala
    50                  55                  60


Glu Thr Ile Glu Glu Lys Arg Glu Asp Phe Gln Lys Glu Glu Lys Glu
65                  70                  75                  80


Asp Phe Thr Glu Glu Gln Asn Ile Glu Asp Leu Leu Ala Ala Val Ala
                85                  90                  95


Asp Ala Glu Gly Arg Thr Ser Thr Gln Glu Lys Lys Thr Thr Thr Ser
            100                 105                 110
```

```
Pro Pro Arg Pro Thr Lys Glu Asn Gly Lys Asp Gln Pro Met Ser Pro
        115                 120                 125


Leu Ser Leu Arg Arg Lys Tyr Met Pro Gly Ala Gly Asn Gly Gly Lys
    130                 135                 140


His Pro Gln Ala His Pro Ala Ala Ala Ala Ala Ala Ala Glu Thr Gln
145                 150                 155                 160


Ala Gln Pro Leu Ser Thr Asn Gln Gly Leu Lys Ser Glu Thr Lys Ile
                165                 170                 175


Ile Thr Thr Thr Asn Gly Gly Thr Arg Ile Thr Leu Thr Arg Phe Lys
            180                 185                 190


Pro Gly Phe Glu Gln Glu Thr Glu Lys Glu Leu Ala Gln Ala Phe Asn
        195                 200                 205


Arg Pro Pro Arg Leu Phe Lys Glu Asp Lys Pro Phe Tyr Pro Trp Leu
    210                 215                 220


Pro Arg Phe Thr Pro Leu Val Asn Phe His Leu Asn Phe Lys Gly
225                 230                 235


<210> 65
<211> 673
<212> PRT
<213> Torque teno midi virus 1

<400> 65
Met Pro Phe Trp Trp Gly Arg Arg Asn Lys Phe Trp Tyr Gly Arg Asn
1               5                   10                  15


Tyr Arg Arg Lys Lys Arg Arg Phe Pro Lys Arg Arg Lys Arg Arg Phe
            20                  25                  30


Tyr Arg Arg Thr Lys Tyr Arg Arg Pro Ala Arg Arg Arg Arg Arg Arg
        35                  40                  45


Arg Arg Lys Val Arg Arg Lys Lys Lys Thr Leu Ile Val Arg Gln Trp
    50                  55                  60
```

```
Gln Pro Asp Ser Ile Val Leu Cys Lys Ile Lys Gly Tyr Asp Ser Ile
65                  70                  75                  80

Ile Trp Gly Ala Glu Gly Thr Gln Phe Gln Cys Ser Thr His Glu Met
                85                  90                  95

Tyr Glu Tyr Thr Arg Gln Lys Tyr Pro Gly Gly Gly Gly Phe Gly Val
            100                 105                 110

Gln Leu Tyr Ser Leu Glu Tyr Leu Tyr Asp Gln Trp Lys Leu Arg Asn
        115                 120                 125

Asn Ile Trp Thr Lys Thr Asn Gln Leu Lys Asp Leu Cys Arg Tyr Leu
    130                 135                 140

Lys Cys Val Met Thr Phe Tyr Arg His Gln His Ile Asp Phe Val Ile
145                 150                 155                 160

Val Tyr Glu Arg Gln Pro Pro Phe Glu Ile Asp Lys Leu Thr Tyr Met
                165                 170                 175

Lys Tyr His Pro Tyr Met Leu Leu Gln Arg Lys His Lys Ile Ile Leu
            180                 185                 190

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

Ile Lys Pro Pro Lys Gln Met Leu Ser Lys Trp Phe Phe Gln Gln Gln
    210                 215                 220

Phe Ala Lys Tyr Asp Leu Leu Leu Ile Ala Ala Ala Ala Cys Ser Leu
225                 230                 235                 240

Arg Tyr Pro Arg Ile Gly Cys Cys Asn Glu Asn Arg Met Ile Thr Leu
                245                 250                 255

Tyr Cys Leu Asn Thr Lys Phe Tyr Gln Asp Thr Glu Trp Gly Thr Thr
            260                 265                 270

Lys Gln Ala Pro His Tyr Phe Lys Pro Tyr Ala Thr Ile Asn Lys Ser
```

```
        275                 280                 285

Met Ile Phe Val Ser Asn Tyr Gly Gly Lys Lys Thr Glu Tyr Asn Ile
    290                 295                 300

Gly Gln Trp Ile Glu Thr Asp Ile Pro Gly Glu Gly Asn Leu Ala Arg
305                 310                 315                 320

Tyr Tyr Arg Ser Ile Ser Lys Glu Gly Gly Tyr Phe Ser Pro Lys Ile
                325                 330                 335

Leu Gln Ala Tyr Gln Thr Lys Val Lys Ser Val Asp Tyr Lys Pro Leu
            340                 345                 350

Pro Ile Val Leu Gly Arg Tyr Asn Pro Ala Ile Asp Asp Gly Lys Gly
        355                 360                 365

Asn Lys Ile Tyr Leu Gln Thr Ile Met Asn Gly His Trp Gly Leu Pro
    370                 375                 380

Gln Lys Thr Pro Asp Tyr Ile Ile Glu Glu Val Pro Leu Trp Leu Gly
385                 390                 395                 400

Phe Trp Gly Tyr Tyr Asn Tyr Leu Lys Gln Thr Arg Thr Glu Ala Ile
                405                 410                 415

Phe Pro Leu His Met Phe Val Val Gln Ser Lys Tyr Ile Gln Thr Gln
            420                 425                 430

Gln Thr Glu Thr Pro Asn Asn Phe Trp Ala Phe Ile Asp Asn Ser Phe
        435                 440                 445

Ile Gln Gly Lys Asn Pro Trp Asp Ser Val Ile Thr Tyr Ser Glu Gln
    450                 455                 460

Lys Leu Trp Phe Pro Thr Val Ala Trp Gln Leu Lys Thr Ile Asn Ala
465                 470                 475                 480

Ile Cys Glu Ser Gly Pro Tyr Val Pro Lys Leu Asp Asn Gln Thr Tyr
                485                 490                 495
```

```
Ser Thr Trp Glu Leu Ala Thr His Tyr Ser Phe His Phe Lys Trp Gly
            500                 505                 510

Gly Pro Gln Ile Ser Asp Gln Pro Val Glu Asp Pro Gly Asn Lys Asn
        515                 520                 525

Lys Tyr Asp Val Pro Asp Thr Ile Lys Glu Ala Leu Gln Ile Val Asn
    530                 535                 540

Pro Ala Lys Asn Ile Ala Ala Thr Met Phe His Asp Trp Asp Tyr Arg
545                 550                 555                 560

Arg Gly Cys Ile Thr Ser Thr Ala Ile Lys Arg Met Gln Gln Asn Leu
                565                 570                 575

Pro Thr Asp Ser Ser Leu Glu Ser Asp Ser Asp Ser Glu Pro Ala Pro
            580                 585                 590

Lys Lys Lys Arg Leu Leu Pro Val Leu His Asp Pro Gln Lys Lys Thr
        595                 600                 605

Glu Lys Ile Asn Gln Cys Leu Leu Ser Leu Cys Glu Glu Ser Thr Cys
    610                 615                 620

Gln Glu Gln Glu Thr Glu Glu Asn Ile Leu Lys Leu Ile Gln Gln Gln
625                 630                 635                 640

Gln Gln Gln Gln Gln Lys Leu Lys His Asn Leu Leu Val Leu Ile Lys
                645                 650                 655

Asp Leu Lys Val Lys Gln Arg Leu Leu Gln Leu Gln Thr Gly Val Leu
            660                 665                 670

Glu
```

<210> 66
<211> 209
<212> PRT
<213> Torque teno midi virus 1

<400> 66
Met Pro Phe Trp Trp Gly Arg Arg Asn Lys Phe Trp Tyr Gly Arg Asn
1   5   10   15

Tyr Arg Arg Lys Lys Arg Arg Phe Pro Lys Arg Arg Lys Arg Arg Phe
            20   25   30

Tyr Arg Arg Thr Lys Tyr Arg Arg Pro Ala Arg Arg Arg Arg Arg Arg
        35   40   45

Arg Arg Lys Ile Ser Asp Gln Pro Val Glu Asp Pro Gly Asn Lys Asn
    50   55   60

Lys Tyr Asp Val Pro Asp Thr Ile Lys Glu Ala Leu Gln Ile Val Asn
65   70   75   80

Pro Ala Lys Asn Ile Ala Ala Thr Met Phe His Asp Trp Asp Tyr Arg
                85   90   95

Arg Gly Cys Ile Thr Ser Thr Ala Ile Lys Arg Met Gln Gln Asn Leu
            100   105   110

Pro Thr Asp Ser Ser Leu Glu Ser Asp Ser Asp Ser Glu Pro Ala Pro
        115   120   125

Lys Lys Lys Arg Leu Leu Pro Val Leu His Asp Pro Gln Lys Lys Thr
    130   135   140

Glu Lys Ile Asn Gln Cys Leu Leu Ser Leu Cys Glu Glu Ser Thr Cys
145   150   155   160

Gln Glu Gln Glu Thr Glu Glu Asn Ile Leu Lys Leu Ile Gln Gln Gln
                165   170   175

Gln Gln Gln Gln Gln Lys Leu Lys His Asn Leu Leu Val Leu Ile Lys
            180   185   190

Asp Leu Lys Val Lys Gln Arg Leu Leu Gln Leu Gln Thr Gly Val Leu
        195   200   205

Glu

<210> 67
<211> 209
<212> PRT
<213> Torque teno midi virus 1

<400> 67
Met Pro Phe Trp Trp Gly Arg Arg Asn Lys Phe Trp Tyr Gly Arg Asn
1   5   10   15

Tyr Arg Arg Lys Lys Arg Arg Phe Pro Lys Arg Arg Lys Arg Arg Phe
            20   25   30

Tyr Arg Arg Thr Lys Tyr Arg Arg Pro Ala Arg Arg Arg Arg Arg Arg
        35   40   45

Arg Arg Lys Ile Ser Asp Gln Pro Val Glu Asp Pro Gly Asn Lys Asn
    50   55   60

Lys Tyr Asp Val Pro Asp Thr Ile Lys Glu Ala Leu Gln Ile Val Asn
65   70   75   80

Pro Ala Lys Asn Ile Ala Ala Thr Met Phe His Asp Trp Asp Tyr Arg
                85   90   95

Arg Gly Cys Ile Thr Ser Thr Ala Ile Lys Arg Met Gln Gln Asn Leu
            100   105   110

Pro Thr Asp Ser Ser Leu Glu Ser Asp Ser Asp Ser Glu Pro Ala Pro
        115   120   125

Lys Lys Lys Arg Leu Leu Pro Val Leu His Asp Pro Gln Lys Lys Thr
    130   135   140

Glu Lys Ile Asn Gln Cys Leu Leu Ser Leu Cys Glu Glu Ser Thr Cys
145   150   155   160

Gln Glu Gln Glu Thr Glu Glu Asn Ile Leu Lys Leu Ile Gln Gln Gln
                165   170   175

Gln Gln Gln Gln Gln Lys Leu Lys His Asn Leu Leu Val Leu Ile Lys

180 185 190

Asp Leu Lys Val Lys Gln Arg Leu Leu Gln Leu Gln Thr Gly Val Leu
195 200 205

Glu

<210> 68
<211> 25
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<220>
<221> VARIANT
<222> (3)..(3)
<223> /replace="Ile"

<220>
<221> MOD_RES
<222> (9)..(10)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (15)..(15)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (17)..(17)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (20)..(20)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (24)..(24)
<223> Any amino acid

<220>
<221> SITE

<222> (1)..(25)
<223> /note="Variant residues given in the sequence have no
preference with respect to those in the annotations
for variant positions"

<400> 68
Leu Ile Leu Arg Gln Trp Gln Pro Xaa Xaa Ile Arg Arg Cys Xaa Ile
1 5 10 15

Xaa Gly Tyr Xaa Pro Leu Ile Xaa Cys
20 25

<210> 69
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (3)..(4)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (6)..(6)
<223> Any amino acid

<400> 69
Asn Tyr Xaa Xaa His Xaa Asp
1 5

<210> 70
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (4)..(5)

<223> Any amino acid

<400> 70
Phe Ser Leu Xaa Xaa Leu Tyr Asp Glx
1   5

<210> 71
<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (2)..(2)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (8)..(8)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (16)..(16)
<223> Any amino acid

<400> 71
Asn Xaa Trp Thr Xaa Ser Asn Xaa Asp Leu Asp Leu Cys Arg Tyr Xaa
1   5   10   15

Gly Cys

<210> 72
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<220>
<221> MOD_RES
<222> (2)..(2)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(9)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (11)..(11)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (13)..(13)
<223> Any amino acid

<400> 72
Thr Xaa Pro Ser Xaa His Pro Gly Xaa Met Xaa Leu Xaa Lys His Lys
1 5 10 15

<210> 73
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(9)
<223> Any amino acid

<400> 73
Ile Pro Ser Leu Xaa Thr Arg Pro Xaa Gly
1   5   10

<210> 74
<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (3)..(3)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (6)..(6)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(9)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (16)..(16)
<223> Any amino acid

<400> 74
Arg Ile Xaa Pro Pro Xaa Leu Phe Xaa Asp Lys Trp Tyr Phe Gln Xaa
1   5   10   15

Asp Leu

<210> 75
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic

peptide"

<220>
<221> MOD_RES
<222> (3)..(3)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<400> 75
Leu Leu Xaa Ile Xaa Ala Thr Ala
1 5

<210> 76
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (2)..(3)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (6)..(6)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(9)
<223> Any amino acid

<400> 76
Leu Xaa Xaa Pro Phe Xaa Ser Pro Xaa Thr Asp
1 5 10

<210> 77
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (4)..(5)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(9)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (12)..(12)
<223> Any amino acid

<400> 77
Tyr Asn Pro Xaa Xaa Asp Lys Gly Xaa Gly Asn Xaa Ile Trp
1 5 10

<210> 78
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (7)..(7)
<223> Any amino acid

<400> 78
Cys Pro Tyr Thr Glx Pro Xaa Leu
1 5

<210> 79
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (1)..(1)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (4)..(4)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (6)..(6)
<223> Any amino acid

<400> 79
Xaa Phe Gly Xaa Gly Xaa Met Pro
1 5

<210> 80
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (3)..(3)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (6)..(6)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (8)..(8)
<223> Any amino acid

<400> 80
His Gln Xaa Glu Val Xaa Glu Xaa
1 5

<210> 81
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (3)..(3)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (7)..(7)
<223> Any amino acid

<400> 81
Lys Tyr Xaa Phe Xaa Phe Xaa Trp Gly Gly Asn Pro
1 5 10

<210> 82
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<400> 82
His Ser Trp Asp Xaa Arg Arg Gly
1 5

<210> 83

<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<400> 83
Ala Ile Lys Arg Xaa Gln Gln
1 5

<210> 84
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (1)..(1)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (5)..(6)
<223> Any amino acid

<400> 84
Xaa Gln Glx Gln Xaa Xaa Leu Arg
1 5

<210> 85
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic

peptide"

<220>
<221> MOD_RES
<222> (3)..(3)
<223> Any amino acid

<220>
<221> VARIANT
<222> (4)..(4)
<223> /replace="Ile"

<220>
<221> MOD_RES
<222> (6)..(7)
<223> Any amino acid

<220>
<221> SITE
<222> (1)..(9)
<223> /note="Variant residues given in the sequence have no
preference with respect to those in the annotations
for variant positions"

<400> 85
Pro Arg Xaa Leu Gln Xaa Xaa Asp Pro
1 5

<210> 86
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<400> 86
His Ser Trp Asp Xaa Arg Arg Gly
1 5

<210> 87
<211> 7
<212> PRT

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<400> 87
Ala Ile Lys Arg Xaa Gln Gln
1 5

<210> 88
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (4)..(5)
<223> Any amino acid

<400> 88
Gln Glx Gln Xaa Xaa Leu Arg
1 5

<210> 89
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (2)..(2)
<223> Any amino acid

<400> 89
Lys Xaa Lys Arg Arg Arg Arg
1   5

<210> 90
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (3)..(3)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (6)..(7)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (10)..(11)
<223> Any amino acid

<400> 90
Pro Ile Xaa Ser Leu Xaa Xaa Tyr Lys Xaa Xaa Thr Arg
1   5   10

<210> 91
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (3)..(3)
<223> Any amino acid

<220>

<221> MOD_RES
<222> (10)..(10)
<223> Any amino acid

<400> 91
Leu Ala Xaa Gln Leu Leu Lys Glu Cys Xaa Lys Asn
1   5   10

<210> 92
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (4)..(4)
<223> Any amino acid

<400> 92
His Leu Asn Xaa Leu Ala
1   5

<210> 93
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<400> 93
Asp Arg Pro Pro Arg
1   5

<210> 94
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (2)..(2)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (8)..(8)
<223> Any amino acid

<400> 94
Asp Xaa Pro Phe Tyr Pro Trp Xaa Pro
1   5

<210> 95
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (2)..(2)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (6)..(6)
<223> Any amino acid

<400> 95
Val Xaa Phe Lys Leu Xaa Phe
1   5

<210> 96
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (2)..(2)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(9)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (14)..(14)
<223> Any amino acid

<400> 96
Trp Xaa Pro Pro Val His Asx Val Xaa Gly Ile Glu Arg Xaa Trp
1   5   10   15

<210> 97
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (6)..(6)
<223> Any amino acid

<400> 97
Ala Lys Arg Lys Leu Xaa
1   5

<210> 98
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES

<222> (4)..(4)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (7)..(8)
<223> Any amino acid

<400> 98
Pro Ser Ser Xaa Asp Trp Xaa Xaa Glu Tyr
1   5   10

<210> 99
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<400> 99
Asp Arg Pro Pro Arg
1   5

<210> 100
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<400> 100
Pro Phe Tyr Pro Trp
1   5

<210> 101
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> MOD_RES
<222> (3)..(3)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (7)..(7)
<223> Any amino acid

<400> 101
Asn Val Xaa Phe Lys Leu Xaa Phe
1   5

<210> 102
<211> 25
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
      peptide"

<220>
<221> VARIANT
<222> (1)..(1)
<223> /replace="Ile"

<220>
<221> MOD_RES
<222> (2)..(5)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(13)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (15)..(15)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (17)..(17)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (19)..(21)

<223> Any amino acid

<220>
<221> VARIANT
<222> (22)..(22)
<223> /replace="Ile"

<220>
<221> SITE
<222> (1)..(25)
<223> /note="Variant residues given in the sequence have no
preference with respect to those in the annotations
for variant positions"

<400> 102
Leu Xaa Xaa Xaa Xaa Trp Gln Pro Xaa Xaa Xaa Xaa Xaa Cys Xaa Ile
1 5 10 15

Xaa Gly Xaa Xaa Xaa Leu Trp Gln Pro
20 25

<210> 103
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (2)..(2)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (4)..(6)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (8)..(11)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (13)..(13)
<223> Any amino acid

<400> 103
Asn Xaa Trp Xaa Xaa Xaa Asn Xaa Xaa Xaa Xaa Leu Xaa Arg Tyr
1   5   10   15

<210> 104
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="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

<400> 104
Tyr Asn Pro Xaa Xaa Asp Xaa Gly
1   5

<210> 105
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="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>
<221> source

<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220>
<221> variation
<222> (10)..(10)
<223> /replace="c" or " "

<220>
<221> variation
<222> (12)..(12)
<223> /replace=" "

<220>
<221> variation
<222> (30)..(30)
<223> /replace=" "

<220>
<221> variation
<222> (31)..(31)
<223> /replace=" "

<220>
<221> variation
<222> (32)..(32)
<223> /replace=" "

<220>
<221> variation
<222> (34)..(34)
<223> /replace="a" or " "

<220>
<221> variation
<222> (43)..(43)
<223> /replace="g" or "t" or " "

<220>
<221> variation
<222> (44)..(44)
<223> /replace="c" or " "

<220>
<221> variation

<222> (45)..(45)
<223> /replace="c" or " "

<220>
<221> variation
<222> (46)..(46)
<223> /replace="a" or " "

<220>
<221> variation
<222> (52)..(52)
<223> /replace="t" or " "

<220>
<221> variation
<222> (53)..(53)
<223> /replace=" "

<220>
<221> variation
<222> (54)..(54)
<223> /replace="c" or " "

<220>
<221> variation
<222> (70)..(70)
<223> /replace=" "

<220>
<221> variation
<222> (71)..(71)
<223> /replace=" "

<220>
<221> variation
<222> (89)..(89)
<223> /replace="t" or " "

<220>
<221> variation
<222> (90)..(90)
<223> /replace="c" or "a" or " "

<220>
<221> variation
<222> (103)..(103)
<223> /replace=" "

<220>
<221> misc_feature
<222> (1)..(117)
<223> /note="Variant nucleotides given in the sequence have no
preference with respect to those in the annotations
for variant positions"

<400> 120
cggcggsggg gcsscgcgct dcgcgcgcsg cccgsyrggg graggcmwgc sgcgcccccc 60

cscgcgcatg cgcrcgggkc cccccccccg sgggggggctc cgcccccccg gcccccc 117

<210> 121
<211> 169
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 126
gcggcgg 7

<210> 127
<211> 7
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 127
gggggcg 7

<210> 128
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 128
gccgcg 6

<210> 129
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 129
ttcgcgcgcc gcccaccagg gggtg 25

<210> 130
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 130
ctgcg 5

<210> 131
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 131
cgccccccc cgcgcat 17

<210> 132
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 132
gcgcggggcc ccccccc 17

<210> 133
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 133
ggggggctc cgccccccg gccccccccc gtgctaaacc caccgcgcat gcgcgaccac 60

gcccccgccg cc 72

<210> 134
<211> 115
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="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>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<400> 135
cggcggcggc ggcg 14

<210> 136
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<400> 136
cgcgcgctgc gcgcgcg 17

<210> 137
<211> 19
<212> DNA
<213> Artificial Sequence

<220>

<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 137
cgccgggggg gcgccagcg 19

<210> 138
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 138
ccccccccc cgcgcat 17

<210> 139
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 139
gcacgggtcc ccccccccac ggggggctcc g 31

<210> 140
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 140
ccccccggcc ccccccc 17

<210> 141
<211> 121
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 141
ccgtcggcgg gggggccgcg cgctgcgcgc gcggcccccg ggggaggcac agcctccccc 60

ccccgcgcgc atgcgcgcgg gtccccccccc ctccgggggg ctccgccccc cggccccccc 120

c 121

<210> 142
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 142
ccgtcggcgg gggggccgcg cgctgcgcgc gcggccc 37

<210> 143
<211> 84
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 145
cggcggcggc g 11

<210> 146
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 146
cgcgcgctac gcgcgcg 17

<210> 147
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 147
cgccgggggg 10

<210> 148
<211> 7
<212> DNA
<213> Artificial Sequence

<220>

<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 148
ctgccgc 7

<210> 149
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 149
ccccccccg cgcat 15

<210> 150
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 150
gcgcggggcc ccccccc 17

<210> 151
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 151
gcgggggct ccg 13

<210> 152
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 152
ccccccggcc cccc 14

<210> 153
<211> 122
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 153
gccgccgcgg cggcgggggg cggcgcgctg cgcgcgccgc ccagtagggg gagccatgcg 60

ccccccccg cgcatgcgcg gggcccccc ccgcgggggg ctccgccccc cggcccccc 120

cg 122

<210> 154
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 154
gccgccgcgg cggcggggg 19

<210> 155
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 155
gcggcgcgct gcgcgcgccg cccagtaggg ggagccatgc g 41

<210> 156
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 156
ccccccccg cgcat 15

<210> 157
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 157
gcgcggggcc ccccccc 17

<210> 158
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 158
gcgggggct ccg 13

<210> 159
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 159
ccccccggcc ccccccg 17

<210> 160
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 160
cgcgctgcgc gcgccgccca gtaggggggag ccatgc 36

<210> 161
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 163
cgcgctgcgc gcgccgccca gtagggggag ccatgc 36

<210> 164
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 164
gcgctdcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 165
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 165
gcgcttcgcg cgccgcccac taggggcgt tgcgcg 36

<210> 166
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 166
gcgctgcgcg cgccgcccag taggggcgc aatgcg 36

<210> 167
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 167
gcgctgcgcg cgcggccccc gggggaggca ttgcct 36

<210> 168
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 168
gcgctgcgcg cgcgcgccgg gggggcgcca gcgccc 36

<210> 169
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 169
gcgcttcgcg cgcgcgccgg ggggctccgc cccccc 36

<210> 170
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 170
gcgcttcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 171
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 171
gcgctacgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 172
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 172
gcgctacgcg cgcgcgccgg ggggctctgc cccccc 36

<210> 173
<211> 680
<212> PRT
<213> Torque teno virus

<400> 173
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1 5 10 15

Pro Trp Arg Pro 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 Val Leu Arg Gln Trp Gln Pro Asp Val Ile Arg
65 70 75 80

```
His Cys Lys Ile Thr Gly Arg 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 Asp 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 Leu 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 His Arg Glu Asp Arg Leu
        275                 280                 285
```

```
Asn Ile Ile Asn Asn Thr Leu His Pro His Asp Ile Thr Gly Pro 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 Leu Leu Arg Glu Tyr Gly
                325                 330                 335

Leu Tyr Ser Pro Tyr Tyr Leu Asn Pro Thr Arg Ile Asn Leu Asp Trp
            340                 345                 350

Met Thr Pro Tyr Thr His Val Arg Tyr Asn Pro Leu Val Asp Lys Gly
        355                 360                 365

Phe Gly Asn Arg Ile Tyr Ile Gln Trp Cys Ser Glu Ala Asp Val Ser
    370                 375                 380

Tyr Asn Arg Thr Lys Ser Lys Cys Leu Leu Gln Asp Met Pro Leu Phe
385                 390                 395                 400

Phe Met Cys Tyr Gly Tyr Ile Asp Trp Ala Ile Lys Asn Thr Gly Val
                405                 410                 415

Ser Ser Leu Ala Arg Asp Ala Arg Ile Cys Ile Arg Cys Pro Tyr Thr
            420                 425                 430

Glu Pro Gln Leu Val Gly Ser Thr Glu Asp Ile Gly Phe Val Pro Ile
        435                 440                 445

Thr Glu Thr Phe Met Arg Gly Asp Met Pro Val Leu Ala Pro Tyr Ile
    450                 455                 460

Pro Leu Ser Trp Phe Cys Lys Trp Tyr Pro Asn Ile Ala His Gln Lys
465                 470                 475                 480

Glu Val Leu Glu Ala Ile Ile Ser Cys Ser Pro Phe Met Pro Arg Asp
                485                 490                 495

Gln Gly Met Asn Gly Trp Asp Ile Thr Ile Gly Tyr Lys Met Asp Phe
            500                 505                 510
```

```
Leu Trp Gly Gly Ser Pro Leu Pro Ser Gln Pro Ile Asp Asp Pro Cys
        515                 520                 525

Gln Gln Gly Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu
    530                 535                 540

Leu Gln Val Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His
545                 550                 555                 560

Lys Trp Asp Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg
                565                 570                 575

Val Ser Glu Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro
            580                 585                 590

Ser Lys Arg Asn Lys Leu Asp Ser Ala Phe Arg Gly Glu Asn Pro Glu
        595                 600                 605

Gln Lys Glu Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr
    610                 615                 620

Pro Glu Glu Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu
625                 630                 635                 640

Leu Leu His Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg
                645                 650                 655

Arg Gly Leu Lys Leu Val Phe Thr Asp Ile Leu Arg Leu Arg Gln Gly
            660                 665                 670

Val His Trp Asn Pro Glu Leu Thr
        675                 680
```

<210> 174
<211> 66
<212> PRT
<213> Torque teno virus

<400> 174

```
Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1               5                   10                  15
```

Pro Trp Arg Pro 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> 175
<211> 211
<212> PRT
<213> Torque teno virus

<400> 175
Pro Thr Leu Val Leu Arg Gln Trp Gln Pro Asp Val Ile Arg His Cys
1   5   10   15

Lys Ile Thr Gly Arg 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 Asp 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 Leu 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> 176
<211> 70
<212> PRT
<213> Torque teno virus

<400> 176

```
Ser Asn Leu Pro Gln His Arg Glu Asp Arg Leu Asn Ile Ile Asn Asn
1               5                   10                  15

Thr Leu His Pro His Asp Ile Thr Gly Pro 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 Leu Leu Arg Glu Tyr Gly Leu Tyr Ser Pro Tyr
    50                  55                  60

Tyr Leu Asn Pro Thr Arg
65                  70
```

<210> 177
<211> 166

<212> PRT
<213> Torque teno virus

<400> 177
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 Ala Arg 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 Thr 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 Ala Ile Ile Ser Cys Ser Pro
    130   135   140

Phe Met Pro Arg Asp Gln Gly Met Asn Gly Trp Asp Ile Thr Ile Gly
145   150   155   160

Tyr Lys Met Asp Phe Leu
                165

<210> 178
<211> 167
<212> PRT
<213> Torque teno virus

<400> 178

```
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 Pro 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> 179
<211> 747
<212> PRT
<213> Torque teno virus

<400> 179

```
Met Ala Tyr Trp Trp Gly Arg Arg Arg Arg Trp Arg Arg Trp Arg Arg
1               5                  10                  15
```

```
Arg Arg Arg Pro Leu Arg Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg
            20                  25                  30

Trp Pro Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg Arg Ala Arg Pro
        35                  40                  45

Ala Arg Arg Tyr Arg Arg Arg Arg Gly Arg Arg Arg Val Arg Arg Arg
    50                  55                  60

Arg Arg Pro Gln Lys Leu Val Leu Thr Gln Trp Asn Pro Gln Thr Val
65                  70                  75                  80

Arg Lys Cys Val Ile Arg Gly Phe Leu Pro Leu Phe Phe Cys Gly Gln
                85                  90                  95

Gly Ala Tyr His Arg Asn Phe Thr Asp His Tyr Asp Asp Val Phe Pro
            100                 105                 110

Lys Gly Pro Ser Gly Gly Gly His Gly Ser Met Val Phe Asn Leu Ser
        115                 120                 125

Phe Leu Tyr Gln Glu Phe Lys Lys His His Asn Lys Trp Ser Arg Ser
    130                 135                 140

Asn Leu Asp Phe Asp Leu Val Arg Tyr Lys Gly Thr Val Ile Lys Leu
145                 150                 155                 160

Tyr Arg His Gln Asp Phe Asp Tyr Ile Val Trp Ile Ser Arg Thr Pro
                165                 170                 175

Pro Phe Gln Glu Ser Leu Leu Thr Val Met Thr His Gln Pro Ser Val
            180                 185                 190

Met Leu Gln Ala Lys Lys Cys Ile Ile Val Lys Ser Tyr Arg Thr His
        195                 200                 205

Pro Gly Gly Lys Pro Tyr Val Thr Ala Lys Val Arg Pro Pro Arg Leu
    210                 215                 220

Leu Thr Asp Lys Trp Tyr Phe Gln Ser Asp Phe Cys Asn Val Pro Leu
225                 230                 235                 240
```

```
Phe Ser Leu Gln Phe Ala Leu Ala Glu Leu Arg Phe Pro Ile Cys Ser
                245                 250                 255

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

Tyr Tyr Lys Phe Leu Asp Asn Lys Pro Lys Gln Ser Ser Asp Pro Asn
        275                 280                 285

Asp Glu Asn Arg Ile Lys Phe Trp His Gly Leu Trp Ser Thr Met Arg
    290                 295                 300

Tyr Leu Asn Thr Thr Tyr Ile Asn Thr Leu Phe Pro Gly Thr Asp Ser
305                 310                 315                 320

Leu Val Ala Ala Lys Asp Thr Asp Asn Ser Val Asn Lys Tyr Pro Ser
                325                 330                 335

Thr Ala Thr Lys Gln Pro Tyr Lys Asp Ser Gln Tyr Met Gln Asn Ile
            340                 345                 350

Trp Asn Thr Ser Lys Ile His Ala Leu Tyr Thr Trp Val Ala Glu Thr
        355                 360                 365

Asn Tyr Lys Arg Leu Gln Ala Tyr Tyr Thr Gln Thr Tyr Gly Gly Tyr
    370                 375                 380

Gln Arg Gln Phe Phe Thr Gly Lys Gln Tyr Trp Asp Tyr Arg Val Gly
385                 390                 395                 400

Met Phe Ser Pro Ala Phe Leu Ser Pro Ser Arg Leu Asn Pro Gln Asn
                405                 410                 415

Pro Gly Ala Tyr Thr Glu Val Ser Tyr Asn Pro Trp Thr Asp Glu Gly
            420                 425                 430

Thr Gly Asn Val Val Cys Leu Gln Tyr Leu Thr Lys Glu Thr Ser Asp
        435                 440                 445
```

```
Tyr Lys Pro Gly Gly Gly Ser Lys Phe Cys Ile Glu Gly Val Pro Leu
    450                 455                 460

Trp Ala Ala Leu Val Gly Tyr Val Asp Met Cys Lys Lys Glu Gly Lys
465                 470                 475                 480

Asp Pro Gly Ile Arg Leu Asn Cys Leu Leu Leu Val Lys Cys Pro Tyr
                485                 490                 495

Thr Lys Pro Gln Leu Tyr Asp Lys Lys Asn Pro Glu Lys Leu Phe Val
            500                 505                 510

Pro Tyr Ser Tyr Asn Phe Gly His Gly Lys Met Pro Gly Gly Asp Lys
        515                 520                 525

Tyr Ile Pro Ile Glu Phe Lys Asp Arg Trp Tyr Pro Cys Leu Leu His
    530                 535                 540

Gln Glu Glu Trp Ile Glu Asp Ile Val Arg Ser Gly Pro Phe Val Pro
545                 550                 555                 560

Lys Asp Met Pro Ser Ser Val Thr Cys Met Met Arg Tyr Ser Ser Leu
                565                 570                 575

Phe Asn Trp Gly Gly Asn Ile Ile Gln Glu Gln Ala Val Glu Asp Pro
            580                 585                 590

Cys Lys Lys Gly Thr Phe Val Val Pro Gly Thr Ser Gly Ile Ala Arg
        595                 600                 605

Ile Leu Gln Val Ser Asn Pro Ala Lys Gln Thr Pro Thr Thr Thr Trp
    610                 615                 620

His Ser Trp Asp Trp Arg Arg Ser Leu Phe Thr Glu Thr Gly Leu Lys
625                 630                 635                 640

Arg Met Arg Glu Gln Gln Pro Tyr Asp Glu Leu Ser Tyr Thr Gly Pro
                645                 650                 655

Lys Lys Pro Lys Leu Ser Leu Pro Ala Gly Pro Ala Val Pro Gly Ala
            660                 665                 670
```

Ala Val Ala Ser Ser Trp Trp Glu Thr Lys Gln Val Thr Ser Pro Asp
675 680 685

Val Ser Glu Thr Glu Thr Glu Ala Glu Ala His Gln Glu Glu Glu Thr
690 695 700

Glu Pro Glu Glu Gly Val Gln Leu Gln Gln Leu Trp Glu Gln Gln Leu
705 710 715 720

Leu Gln Lys Arg Gln Leu Gly Val Val Phe Gln Gln Leu Leu Arg Leu
725 730 735

Arg Gln Gly Ala Glu Ile His Pro Gly Leu Val
740 745

<210> 180
<211> 69
<212> PRT
<213> Torque teno virus

<400> 180
Met Ala Tyr Trp Trp Gly Arg Arg Arg Arg Trp Arg Arg Trp Arg Arg
1 5 10 15

Arg Arg Arg Pro Leu Arg Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg
20 25 30

Trp Pro Arg Arg Arg Arg Trp Arg Arg Arg Arg Arg Arg Ala Arg Pro
35 40 45

Ala Arg Arg Tyr Arg Arg Arg Arg Gly Arg Arg Arg Val Arg Arg Arg
50 55 60

Arg Arg Pro Gln Lys
65

<210> 181
<211> 210
<212> PRT
<213> Torque teno virus

<400> 181

```
Leu Val Leu Thr Gln Trp Asn Pro Gln Thr Val Arg Lys Cys Val Ile
1               5                  10                  15

Arg Gly Phe Leu Pro Leu Phe Phe Cys Gly Gln Gly Ala Tyr His Arg
            20                  25                  30

Asn Phe Thr Asp His Tyr Asp Asp Val Phe Pro Lys Gly Pro Ser Gly
        35                  40                  45

Gly Gly His Gly Ser Met Val Phe Asn Leu Ser Phe Leu Tyr Gln Glu
    50                  55                  60

Phe Lys Lys His His Asn Lys Trp Ser Arg Ser Asn Leu Asp Phe Asp
65                  70                  75                  80

Leu Val Arg Tyr Lys Gly Thr Val Ile Lys Leu Tyr Arg His Gln Asp
                85                  90                  95

Phe Asp Tyr Ile Val Trp Ile Ser Arg Thr Pro Pro Phe Gln Glu Ser
            100                 105                 110

Leu Leu Thr Val Met Thr His Gln Pro Ser Val Met Leu Gln Ala Lys
        115                 120                 125

Lys Cys Ile Ile Val Lys Ser Tyr Arg Thr His Pro Gly Gly Lys Pro
    130                 135                 140

Tyr Val Thr Ala Lys Val Arg Pro Pro Arg Leu Leu Thr Asp Lys Trp
145                 150                 155                 160

Tyr Phe Gln Ser Asp Phe Cys Asn Val Pro Leu Phe Ser Leu Gln Phe
                165                 170                 175

Ala Leu Ala Glu Leu Arg Phe Pro Ile Cys Ser Pro Gln Thr Asp Thr
            180                 185                 190

Asn Cys Ile Asn Phe Leu Val Leu Asp Asp Ile Tyr Tyr Lys Phe Leu
        195                 200                 205

Asp Asn
    210
```

<210> 182
<211> 132
<212> PRT
<213> Torque teno virus

<400> 182
Lys Pro Lys Gln Ser Ser Asp Pro Asn Asp Glu Asn Arg Ile Lys Phe
1   5   10   15

Trp His Gly Leu Trp Ser Thr Met Arg Tyr Leu Asn Thr Thr Tyr Ile
            20   25   30

Asn Thr Leu Phe Pro Gly Thr Asp Ser Leu Val Ala Ala Lys Asp Thr
        35   40   45

Asp Asn Ser Val Asn Lys Tyr Pro Ser Thr Ala Thr Lys Gln Pro Tyr
    50   55   60

Lys Asp Ser Gln Tyr Met Gln Asn Ile Trp Asn Thr Ser Lys Ile His
65   70   75   80

Ala Leu Tyr Thr Trp Val Ala Glu Thr Asn Tyr Lys Arg Leu Gln Ala
                85   90   95

Tyr Tyr Thr Gln Thr Tyr Gly Gly Tyr Gln Arg Gln Phe Phe Thr Gly
            100   105   110

Lys Gln Tyr Trp Asp Tyr Arg Val Gly Met Phe Ser Pro Ala Phe Leu
        115   120   125

Ser Pro Ser Arg
    130

<210> 183
<211> 167
<212> PRT
<213> Torque teno virus

<400> 183
Leu Asn Pro Gln Asn Pro Gly Ala Tyr Thr Glu Val Ser Tyr Asn Pro
1   5   10   15

```
Trp Thr Asp Glu Gly Thr Gly Asn Val Val Cys Leu Gln Tyr Leu Thr
            20                  25                  30

Lys Glu Thr Ser Asp Tyr Lys Pro Gly Gly Gly Ser Lys Phe Cys Ile
        35                  40                  45

Glu Gly Val Pro Leu Trp Ala Ala Leu Val Gly Tyr Val Asp Met Cys
    50                  55                  60

Lys Lys Glu Gly Lys Asp Pro Gly Ile Arg Leu Asn Cys Leu Leu Leu
65                  70                  75                  80

Val Lys Cys Pro Tyr Thr Lys Pro Gln Leu Tyr Asp Lys Lys Asn Pro
                85                  90                  95

Glu Lys Leu Phe Val Pro Tyr Ser Tyr Asn Phe Gly His Gly Lys Met
            100                 105                 110

Pro Gly Gly Asp Lys Tyr Ile Pro Ile Glu Phe Lys Asp Arg Trp Tyr
        115                 120                 125

Pro Cys Leu Leu His Gln Glu Glu Trp Ile Glu Asp Ile Val Arg Ser
    130                 135                 140

Gly Pro Phe Val Pro Lys Asp Met Pro Ser Ser Val Thr Cys Met Met
145                 150                 155                 160

Arg Tyr Ser Ser Leu Phe Asn
                165
```

<210> 184
<211> 169
<212> PRT
<213> Torque teno virus

```
<400> 184
Trp Gly Gly Asn Ile Ile Gln Glu Gln Ala Val Glu Asp Pro Cys Lys
1               5                   10                  15

Lys Gly Thr Phe Val Val Pro Gly Thr Ser Gly Ile Ala Arg Ile Leu
            20                  25                  30
```

```
Gln Val Ser Asn Pro Ala Lys Gln Thr Pro Thr Thr Thr Trp His Ser
        35                  40                  45

Trp Asp Trp Arg Arg Ser Leu Phe Thr Glu Thr Gly Leu Lys Arg Met
    50                  55                  60

Arg Glu Gln Gln Pro Tyr Asp Glu Leu Ser Tyr Thr Gly Pro Lys Lys
65                  70                  75                  80

Pro Lys Leu Ser Leu Pro Ala Gly Pro Ala Val Pro Gly Ala Ala Val
                85                  90                  95

Ala Ser Ser Trp Trp Glu Thr Lys Gln Val Thr Ser Pro Asp Val Ser
            100                 105                 110

Glu Thr Glu Thr Glu Ala Glu Ala His Gln Glu Glu Glu Thr Glu Pro
        115                 120                 125

Glu Glu Gly Val Gln Leu Gln Gln Leu Trp Glu Gln Gln Leu Leu Gln
    130                 135                 140

Lys Arg Gln Leu Gly Val Val Phe Gln Gln Leu Leu Arg Leu Arg Gln
145                 150                 155                 160

Gly Ala Glu Ile His Pro Gly Leu Val
                165
```

<210> 185
<211> 743
<212> PRT
<213> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

```
<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
<211> 780
<212> PRT
<213> Torque teno virus

<400> 191
Met Ala Trp Trp Gly Trp Arg Arg Arg Trp Trp Arg Pro Lys Arg Arg
1 5 10 15

Trp Arg Trp Arg Arg Ala Arg Arg Arg Arg Arg Val Pro Ala Arg Arg
20 25 30

Pro Arg Arg Ala Phe Arg Arg Tyr Arg Thr Arg Thr Val Arg Arg Arg
35 40 45

Arg Arg Gly Arg Arg Arg Gly Tyr Arg Arg Arg Tyr Arg Leu Arg Arg
50 55 60

Tyr Ala Arg Arg Arg Phe Arg Arg Lys Lys Ile Val Leu Thr Gln Trp
65 70 75 80

```
Asn Pro Gln Thr Thr Arg Lys Cys Ile Ile Arg Gly Met Met Pro Val
                85                  90                  95

Leu Trp Ala Gly Met Gly Thr Gly Gly Arg Asn Tyr Ala Val Arg Ser
            100                 105                 110

Asp Asp Tyr Val Val Asn Lys Gly Phe Gly Gly Ser Phe Ala Thr Glu
        115                 120                 125

Thr Phe Ser Leu Lys Val Leu Tyr Asp Gln Phe Gln Arg Gly Phe Asn
    130                 135                 140

Arg Trp Ser His Thr Asn Glu Asp Leu Asp Leu Ala Arg Tyr Arg Gly
145                 150                 155                 160

Cys Arg Trp Thr Phe Tyr Arg His Lys Asp Thr Asp Phe Ile Val Tyr
                165                 170                 175

Phe Thr Asn Asn Pro Pro Met Lys Thr Asn Gln Phe Ser Ala Pro Leu
            180                 185                 190

Thr Thr Pro Gly Met Leu Met Arg Ser Lys Tyr Lys Val Leu Ile Pro
        195                 200                 205

Ser Phe Gln Thr Arg Pro Lys Gly Arg Lys Thr Val Thr Val Lys Ile
    210                 215                 220

Arg Pro Pro Lys Leu Phe Gln Asp Lys Trp Tyr Thr Gln Gln Asp Leu
225                 230                 235                 240

Cys Ser Val Pro Leu Val Gln Leu Asn Val Thr Ala Ala Asp Phe Thr
                245                 250                 255

His Pro Phe Gly Ser Pro Leu Thr Glu Thr Pro Cys Val Glu Phe Gln
            260                 265                 270

Val Leu Gly Asp Leu Tyr Asn Thr Cys Leu Asn Ile Asp Leu Pro Gln
        275                 280                 285

Phe Ser Glu Leu Gly Glu Ile Thr Ser Ala Tyr Ser Lys Pro Asn Ser
    290                 295                 300
```

```
Asn Asn Leu Lys Glu Leu Tyr Lys Glu Leu Phe Thr Lys Ala Thr Ser
305                 310                 315                 320

Gly His Tyr Trp Gln Thr Phe Ile Thr Asn Ser Met Val Arg Ala His
                325                 330                 335

Ile Asp Ala Asp Lys Ala Lys Glu Ala Gln Arg Ala Ser Thr Thr Pro
            340                 345                 350

Ser Tyr Asn Asn Asp Pro Phe Pro Thr Ile Pro Val Lys Ser Glu Phe
        355                 360                 365

Ala Gln Trp Lys Lys Lys Phe Thr Asp Thr Arg Asp Ser Pro Phe Leu
    370                 375                 380

Phe Ala Thr Tyr His Pro Glu Ala Ile Lys Asp Thr Ile Met Lys Met
385                 390                 395                 400

Arg Glu Asn Asn Phe Lys Leu Glu Thr Gly Pro Asn Asp Lys Tyr Gly
                405                 410                 415

Asp Tyr Thr Ala Gln Tyr Gln Gly Asn Thr His Met Leu Asp Tyr Tyr
            420                 425                 430

Leu Gly Phe Tyr Ser Pro Ile Phe Leu Ser Asp Gly Arg Ser Asn Val
        435                 440                 445

Glu Phe Phe Thr Ala Tyr Arg Asp Ile Val Tyr Asn Pro Phe Leu Asp
    450                 455                 460

Lys Ala Gln Gly Asn Met Val Trp Phe Gln Tyr His Thr Lys Thr Asp
465                 470                 475                 480

Asn Lys Phe Lys Lys Pro Glu Cys His Trp Glu Ile Lys Asp Met Pro
                485                 490                 495

Leu Trp Ala Leu Leu Asn Gly Tyr Val Asp Tyr Leu Glu Thr Gln Ile
            500                 505                 510
```

```
Gln Tyr Gly Asp Leu Ser Lys Glu Gly Lys Val Leu Ile Arg Cys Pro
        515                 520                 525

Tyr Thr Lys Pro Ala Leu Val Asp Pro Arg Asp Asp Thr Ala Gly Tyr
    530                 535                 540

Val Val Tyr Asn Arg Asn Phe Gly Arg Gly Lys Trp Ile Asp Gly Gly
545                 550                 555                 560

Gly Tyr Ile Pro Leu His Glu Arg Thr Lys Trp Tyr Val Met Leu Arg
                565                 570                 575

Tyr Gln Thr Asp Val Phe His Asp Ile Val Thr Cys Gly Pro Trp Gln
            580                 585                 590

Tyr Arg Asp Asp Asn Lys Asn Ser Gln Leu Val Ala Lys Tyr Arg Phe
        595                 600                 605

Ser Phe Ile Trp Gly Gly Asn Thr Val His Ser Gln Val Ile Arg Asn
    610                 615                 620

Pro Cys Lys Asp Asn Gln Val Ser Gly Pro Arg Arg Gln Pro Arg Asp
625                 630                 635                 640

Ile Gln Val Val Asp Pro Gln Arg Ile Thr Pro Pro Trp Val Leu His
                645                 650                 655

Ser Phe Asp Gln Arg Arg Gly Leu Phe Thr Glu Thr Ala Leu Arg Arg
            660                 665                 670

Leu Leu Gln Glu Pro Leu Pro Gly Glu Tyr Ala Val Ser Thr Leu Arg
        675                 680                 685

Thr Pro Leu Leu Phe Leu Pro Ser Glu Tyr Gln Arg Glu Asp Gly Ala
    690                 695                 700

Ala Glu Ser Ala Ser Gly Ser Pro Ala Lys Arg Pro Arg Ile Trp Ser
705                 710                 715                 720

Glu Glu Ser Gln Thr Glu Thr Ile Ser Ser Glu Glu Asn Pro Ala Glu
                725                 730                 735
```

```
Thr Thr Arg Glu Leu Leu Gln Arg Lys Leu Arg Glu Gln Arg Ala Leu
            740                 745                 750


Gln Phe Gln Leu Gln His Phe Ala Val Gln Leu Ala Lys Thr Gln Ala
        755                 760                 765


Asn Leu His Val Asn Pro Leu Leu Ser Phe Pro Gln
    770                 775                 780


<210> 192
<211> 74
<212> PRT
<213> Torque teno virus

<400> 192
Met Ala Trp Trp Gly Trp Arg Arg Arg Trp Trp Arg Pro Lys Arg Arg
1               5                   10                  15


Trp Arg Trp Arg Arg Ala Arg Arg Arg Arg Arg Val Pro Ala Arg Arg
            20                  25                  30


Pro Arg Arg Ala Phe Arg Arg Tyr Arg Thr Arg Thr Val Arg Arg Arg
        35                  40                  45


Arg Arg Gly Arg Arg Arg Gly Tyr Arg Arg Arg Tyr Arg Leu Arg Arg
    50                  55                  60


Tyr Ala Arg Arg Arg Phe Arg Arg Lys Lys
65                  70


<210> 193
<211> 210
<212> PRT
<213> Torque teno virus

<400> 193
Ile Val Leu Thr Gln Trp Asn Pro Gln Thr Thr Arg Lys Cys Ile Ile
1               5                   10                  15


Arg Gly Met Met Pro Val Leu Trp Ala Gly Met Gly Thr Gly Gly Arg
            20                  25                  30
```

```
Asn Tyr Ala Val Arg Ser Asp Asp Tyr Val Val Asn Lys Gly Phe Gly
        35                  40                  45

Gly Ser Phe Ala Thr Glu Thr Phe Ser Leu Lys Val Leu Tyr Asp Gln
    50                  55                  60

Phe Gln Arg Gly Phe Asn Arg Trp Ser His Thr Asn Glu Asp Leu Asp
65                  70                  75                  80

Leu Ala Arg Tyr Arg Gly Cys Arg Trp Thr Phe Tyr Arg His Lys Asp
                85                  90                  95

Thr Asp Phe Ile Val Tyr Phe Thr Asn Asn Pro Pro Met Lys Thr Asn
            100                 105                 110

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

Tyr Lys Val Leu Ile Pro Ser Phe Gln Thr Arg Pro Lys Gly Arg Lys
    130                 135                 140

Thr Val Thr Val Lys Ile Arg Pro Pro Lys Leu Phe Gln Asp Lys Trp
145                 150                 155                 160

Tyr Thr Gln Gln Asp Leu Cys Ser Val Pro Leu Val Gln Leu Asn Val
                165                 170                 175

Thr Ala Ala Asp Phe Thr His Pro Phe Gly Ser Pro Leu Thr Glu Thr
            180                 185                 190

Pro Cys Val Glu Phe Gln Val Leu Gly Asp Leu Tyr Asn Thr Cys Leu
        195                 200                 205

Asn Ile
    210
```

<210> 194
<211> 161
<212> PRT
<213> Torque teno virus

<400> 194

```
Asp Leu Pro Gln Phe Ser Glu Leu Gly Glu Ile Thr Ser Ala Tyr Ser
1               5                   10                  15

Lys Pro Asn Ser Asn Asn Leu Lys Glu Leu Tyr Lys Glu Leu Phe Thr
            20                  25                  30

Lys Ala Thr Ser Gly His Tyr Trp Gln Thr Phe Ile Thr Asn Ser Met
        35                  40                  45

Val Arg Ala His Ile Asp Ala Asp Lys Ala Lys Glu Ala Gln Arg Ala
    50                  55                  60

Ser Thr Thr Pro Ser Tyr Asn Asn Asp Pro Phe Pro Thr Ile Pro Val
65                  70                  75                  80

Lys Ser Glu Phe Ala Gln Trp Lys Lys Lys Phe Thr Asp Thr Arg Asp
                85                  90                  95

Ser Pro Phe Leu Phe Ala Thr Tyr His Pro Glu Ala Ile Lys Asp Thr
            100                 105                 110

Ile Met Lys Met Arg Glu Asn Asn Phe Lys Leu Glu Thr Gly Pro Asn
        115                 120                 125

Asp Lys Tyr Gly Asp Tyr Thr Ala Gln Tyr Gln Gly Asn Thr His Met
    130                 135                 140

Leu Asp Tyr Tyr Leu Gly Phe Tyr Ser Pro Ile Phe Leu Ser Asp Gly
145                 150                 155                 160

Arg
```

<210> 195
<211> 166
<212> PRT
<213> Torque teno virus

<400> 195

```
Ser Asn Val Glu Phe Phe Thr Ala Tyr Arg Asp Ile Val Tyr Asn Pro
1               5                   10                  15
```

```
Phe Leu Asp Lys Ala Gln Gly Asn Met Val Trp Phe Gln Tyr His Thr
            20                  25                  30

Lys Thr Asp Asn Lys Phe Lys Lys Pro Glu Cys His Trp Glu Ile Lys
        35                  40                  45

Asp Met Pro Leu Trp Ala Leu Leu Asn Gly Tyr Val Asp Tyr Leu Glu
    50                  55                  60

Thr Gln Ile Gln Tyr Gly Asp Leu Ser Lys Glu Gly Lys Val Leu Ile
65                  70                  75                  80

Arg Cys Pro Tyr Thr Lys Pro Ala Leu Val Asp Pro Arg Asp Asp Thr
                85                  90                  95

Ala Gly Tyr Val Val Tyr Asn Arg Asn Phe Gly Arg Gly Lys Trp Ile
            100                 105                 110

Asp Gly Gly Gly Tyr Ile Pro Leu His Glu Arg Thr Lys Trp Tyr Val
        115                 120                 125

Met Leu Arg Tyr Gln Thr Asp Val Phe His Asp Ile Val Thr Cys Gly
    130                 135                 140

Pro Trp Gln Tyr Arg Asp Asp Asn Lys Asn Ser Gln Leu Val Ala Lys
145                 150                 155                 160

Tyr Arg Phe Ser Phe Ile
                165
```

<210> 196
<211> 169
<212> PRT
<213> Torque teno virus

```
<400> 196
Trp Gly Gly Asn Thr Val His Ser Gln Val Ile Arg Asn Pro Cys Lys
1               5                   10                  15

Asp Asn Gln Val Ser Gly Pro Arg Arg Gln Pro Arg Asp Ile Gln Val
            20                  25                  30
```

```
Val Asp Pro Gln Arg Ile Thr Pro Pro Trp Val Leu His Ser Phe Asp
        35                  40                  45

Gln Arg Arg Gly Leu Phe Thr Glu Thr Ala Leu Arg Arg Leu Leu Gln
    50                  55                  60

Glu Pro Leu Pro Gly Glu Tyr Ala Val Ser Thr Leu Arg Thr Pro Leu
65                  70                  75                  80

Leu Phe Leu Pro Ser Glu Tyr Gln Arg Glu Asp Gly Ala Ala Glu Ser
                85                  90                  95

Ala Ser Gly Ser Pro Ala Lys Arg Pro Arg Ile Trp Ser Glu Glu Ser
            100                 105                 110

Gln Thr Glu Thr Ile Ser Ser Glu Glu Asn Pro Ala Glu Thr Thr Arg
        115                 120                 125

Glu Leu Leu Gln Arg Lys Leu Arg Glu Gln Arg Ala Leu Gln Phe Gln
    130                 135                 140

Leu Gln His Phe Ala Val Gln Leu Ala Lys Thr Gln Ala Asn Leu His
145                 150                 155                 160

Val Asn Pro Leu Leu Ser Phe Pro Gln
                165

<210> 197
<211> 761
<212> PRT
<213> Torque teno virus

<400> 197
Met Ala Tyr Trp Phe Arg Arg Trp Gly Trp Arg Pro Arg Arg Arg Trp
1               5                   10                  15

Arg Arg Trp Arg Arg Arg Arg Arg Arg Leu Pro Arg Arg Arg Thr Arg
            20                  25                  30

Arg Ala Val Arg Gly Leu Gly Arg Arg Arg Lys Pro Arg Val Arg Arg
        35                  40                  45
```

```
Arg Arg Arg Thr Arg Arg Arg Thr Tyr Arg Arg Gly Trp Arg Arg Arg
    50                  55                  60

Arg Tyr Ile Arg Arg Gly Arg Arg Lys Lys Lys Leu Ile Leu Thr Gln
65                  70                  75                  80

Trp Asn Pro Ala Ile Val Lys Arg Cys Asn Ile Lys Gly Gly Leu Pro
                85                  90                  95

Ile Ile Ile Cys Gly Glu Pro Arg Ala Ala Phe Asn Tyr Gly Tyr His
            100                 105                 110

Met Glu Asp Tyr Thr Pro Gln Pro Phe Pro Phe Gly Gly Gly Met Ser
        115                 120                 125

Thr Val Thr Phe Ser Leu Lys Ala Leu Tyr Asp Gln Tyr Leu Lys His
    130                 135                 140

Gln Asn Arg Trp Thr Phe Ser Asn Asp Gln Leu Asp Leu Ala Arg Tyr
145                 150                 155                 160

Arg Gly Cys Lys Leu Arg Phe Tyr Arg Ser Pro Val Cys Asp Phe Ile
                165                 170                 175

Val His Tyr Asn Leu Ile Pro Pro Leu Lys Met Asn Gln Phe Thr Ser
            180                 185                 190

Pro Asn Thr His Pro Gly Leu Leu Met Leu Ser Lys His Lys Ile Ile
        195                 200                 205

Ile Pro Ser Phe Gln Thr Arg Pro Gly Gly Arg Arg Phe Val Lys Ile
    210                 215                 220

Arg Leu Asn Pro Pro Lys Leu Phe Glu Asp Lys Trp Tyr Thr Gln Gln
225                 230                 235                 240

Asp Leu Cys Lys Val Pro Leu Val Ser Ile Thr Ala Thr Ala Ala Asp
                245                 250                 255

Leu Arg Tyr Pro Phe Cys Ser Pro Gln Thr Asn Asn Pro Cys Thr Thr
            260                 265                 270
```

```
Phe Gln Val Leu Arg Lys Asn Tyr Asn Thr Val Ile Gly Thr Ser Val
        275                 280                 285

Lys Asp Gln Glu Ser Thr Gln Asp Phe Glu Asn Trp Leu Tyr Lys Thr
    290                 295                 300

Asp Ser His Tyr Gln Thr Phe Ala Thr Glu Ala Gln Leu Gly Arg Ile
305                 310                 315                 320

Pro Ala Phe Asn Pro Asp Gly Thr Lys Asn Thr Lys Gln Gln Ser Trp
                325                 330                 335

Gln Asp Asn Trp Ser Lys Lys Asn Ser Pro Trp Thr Gly Asn Ser Gly
            340                 345                 350

Thr Tyr Pro Gln Thr Thr Ser Glu Met Tyr Lys Ile Pro Tyr Asp Ser
        355                 360                 365

Asn Phe Gly Phe Pro Thr Tyr Arg Ala Gln Lys Asp Tyr Ile Leu Glu
    370                 375                 380

Arg Arg Gln Cys Asn Phe Asn Tyr Glu Val Asn Asn Pro Val Ser Lys
385                 390                 395                 400

Lys Val Trp Pro Gln Pro Ser Thr Thr Thr Pro Thr Val Asp Tyr Tyr
                405                 410                 415

Glu Tyr His Cys Gly Trp Phe Ser Asn Ile Phe Ile Gly Pro Asn Arg
            420                 425                 430

Tyr Asn Leu Gln Phe Gln Thr Ala Tyr Val Asp Thr Thr Tyr Asn Pro
        435                 440                 445

Leu Met Asp Lys Gly Lys Gly Asn Lys Ile Trp Phe Gln Tyr Leu Ser
    450                 455                 460

Lys Lys Gly Thr Asp Tyr Asn Glu Lys Gln Cys Tyr Cys Thr Leu Glu
465                 470                 475                 480
```

```
Asp Met Pro Leu Trp Ala Ile Cys Phe Gly Tyr Thr Asp Tyr Val Glu
                485                 490                 495

Thr Gln Leu Gly Pro Asn Val Asp His Glu Thr Ala Gly Leu Ile Ile
            500                 505                 510

Met Ile Cys Pro Tyr Thr Gln Pro Pro Met Tyr Asp Lys Asn Arg Pro
        515                 520                 525

Asn Trp Gly Tyr Val Val Tyr Asp Thr Asn Phe Gly Asn Gly Lys Met
    530                 535                 540

Pro Ser Gly Ser Gly Gln Val Pro Val Tyr Trp Gln Cys Arg Trp Arg
545                 550                 555                 560

Pro Met Leu Trp Phe Gln Gln Gln Val Leu Asn Asp Ile Ser Lys Thr
                565                 570                 575

Gly Pro Tyr Ala Tyr Arg Asp Glu Tyr Lys Asn Val Gln Leu Thr Leu
            580                 585                 590

Tyr Tyr Asn Phe Ile Phe Asn Trp Gly Gly Asp Met Tyr Tyr Pro Gln
        595                 600                 605

Val Val Lys Asn Pro Cys Gly Asp Ser Gly Ile Val Pro Gly Ser Gly
    610                 615                 620

Arg Phe Thr Arg Glu Val Gln Val Val Ser Pro Leu Ser Met Gly Pro
625                 630                 635                 640

Ala Tyr Ile Phe His Tyr Phe Asp Ser Arg Arg Gly Phe Phe Ser Glu
                645                 650                 655

Lys Ala Leu Lys Arg Met Gln Gln Gln Gln Glu Phe Asp Glu Ser Phe
            660                 665                 670

Thr Phe Lys Pro Lys Arg Pro Lys Leu Ser Thr Ala Ala Ala Glu Ile
        675                 680                 685

Leu Gln Leu Glu Glu Asp Ser Thr Ser Gly Glu Gly Lys Ser Pro Leu
    690                 695                 700
```

```
Gln Gln Glu Glu Lys Glu Val Glu Val Leu Gln Thr Pro Thr Val Gln
705                 710                 715                 720


Leu Gln Leu Gln Arg Asn Ile Gln Glu Gln Leu Ala Ile Lys Gln Gln
                725                 730                 735


Leu Gln Phe Leu Leu Leu Gln Leu Leu Lys Thr Gln Ser Asn Leu His
            740                 745                 750


Leu Asn Pro Gln Phe Leu Ser Pro Ser
        755                 760


<210> 198
<211> 75
<212> PRT
<213> Torque teno virus

<400> 198
Met Ala Tyr Trp Phe Arg Arg Trp Gly Trp Arg Pro Arg Arg Arg Trp
1               5                   10                  15


Arg Arg Trp Arg Arg Arg Arg Arg Arg Leu Pro Arg Arg Arg Thr Arg
            20                  25                  30


Arg Ala Val Arg Gly Leu Gly Arg Arg Arg Lys Pro Arg Val Arg Arg
        35                  40                  45


Arg Arg Arg Thr Arg Arg Arg Thr Tyr Arg Arg Gly Trp Arg Arg Arg
    50                  55                  60


Arg Tyr Ile Arg Arg Gly Arg Arg Lys Lys Lys
65                  70                  75


<210> 199
<211> 209
<212> PRT
<213> Torque teno virus

<400> 199
Leu Ile Leu Thr Gln Trp Asn Pro Ala Ile Val Lys Arg Cys Asn Ile
1               5                   10                  15
```

```
Lys Gly Gly Leu Pro Ile Ile Ile Cys Gly Glu Pro Arg Ala Ala Phe
            20                  25                  30

Asn Tyr Gly Tyr His Met Glu Asp Tyr Thr Pro Gln Pro Phe Pro Phe
        35                  40                  45

Gly Gly Gly Met Ser Thr Val Thr Phe Ser Leu Lys Ala Leu Tyr Asp
    50                  55                  60

Gln Tyr Leu Lys His Gln Asn Arg Trp Thr Phe Ser Asn Asp Gln Leu
65                  70                  75                  80

Asp Leu Ala Arg Tyr Arg Gly Cys Lys Leu Arg Phe Tyr Arg Ser Pro
                85                  90                  95

Val Cys Asp Phe Ile Val His Tyr Asn Leu Ile Pro Pro Leu Lys Met
            100                 105                 110

Asn Gln Phe Thr Ser Pro Asn Thr His Pro Gly Leu Leu Met Leu Ser
        115                 120                 125

Lys His Lys Ile Ile Ile Pro Ser Phe Gln Thr Arg Pro Gly Gly Arg
    130                 135                 140

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

Trp Tyr Thr Gln Gln Asp Leu Cys Lys Val Pro Leu Val Ser Ile Thr
                165                 170                 175

Ala Thr Ala Ala Asp Leu Arg Tyr Pro Phe Cys Ser Pro Gln Thr Asn
            180                 185                 190

Asn Pro Cys Thr Thr Phe Gln Val Leu Arg Lys Asn Tyr Asn Thr Val
        195                 200                 205

Ile
```

<210> 200
<211> 148

<212> PRT
<213> Torque teno virus

<400> 200
Gly Thr Ser Val Lys Asp Gln Glu Ser Thr Gln Asp Phe Glu Asn Trp
1   5   10   15

Leu Tyr Lys Thr Asp Ser His Tyr Gln Thr Phe Ala Thr Glu Ala Gln
            20   25   30

Leu Gly Arg Ile Pro Ala Phe Asn Pro Asp Gly Thr Lys Asn Thr Lys
        35   40   45

Gln Gln Ser Trp Gln Asp Asn Trp Ser Lys Lys Asn Ser Pro Trp Thr
    50   55   60

Gly Asn Ser Gly Thr Tyr Pro Gln Thr Thr Ser Glu Met Tyr Lys Ile
65   70   75   80

Pro Tyr Asp Ser Asn Phe Gly Phe Pro Thr Tyr Arg Ala Gln Lys Asp
                85   90   95

Tyr Ile Leu Glu Arg Arg Gln Cys Asn Phe Asn Tyr Glu Val Asn Asn
            100   105   110

Pro Val Ser Lys Lys Val Trp Pro Gln Pro Ser Thr Thr Thr Pro Thr
        115   120   125

Val Asp Tyr Tyr Glu Tyr His Cys Gly Trp Phe Ser Asn Ile Phe Ile
    130   135   140

Gly Pro Asn Arg
145

<210> 201
<211> 167
<212> PRT
<213> Torque teno virus

<400> 201
Tyr Asn Leu Gln Phe Gln Thr Ala Tyr Val Asp Thr Thr Tyr Asn Pro
1   5   10   15

```
Leu Met Asp Lys Gly Lys Gly Asn Lys Ile Trp Phe Gln Tyr Leu Ser
            20                  25                  30

Lys Lys Gly Thr Asp Tyr Asn Glu Lys Gln Cys Tyr Cys Thr Leu Glu
        35                  40                  45

Asp Met Pro Leu Trp Ala Ile Cys Phe Gly Tyr Thr Asp Tyr Val Glu
    50                  55                  60

Thr Gln Leu Gly Pro Asn Val Asp His Glu Thr Ala Gly Leu Ile Ile
65                  70                  75                  80

Met Ile Cys Pro Tyr Thr Gln Pro Pro Met Tyr Asp Lys Asn Arg Pro
                85                  90                  95

Asn Trp Gly Tyr Val Val Tyr Asp Thr Asn Phe Gly Asn Gly Lys Met
            100                 105                 110

Pro Ser Gly Ser Gly Gln Val Pro Val Tyr Trp Gln Cys Arg Trp Arg
        115                 120                 125

Pro Met Leu Trp Phe Gln Gln Gln Val Leu Asn Asp Ile Ser Lys Thr
    130                 135                 140

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

Tyr Tyr Asn Phe Ile Phe Asn
                165
```

<210> 202
<211> 162
<212> PRT
<213> Torque teno virus

```
<400> 202
Trp Gly Gly Asp Met Tyr Tyr Pro Gln Val Val Lys Asn Pro Cys Gly
1               5                   10                  15

Asp Ser Gly Ile Val Pro Gly Ser Gly Arg Phe Thr Arg Glu Val Gln
            20                  25                  30
```

```
Val Val Ser Pro Leu Ser Met Gly Pro Ala Tyr Ile Phe His Tyr Phe
        35                  40                  45

Asp Ser Arg Arg Gly Phe Phe Ser Glu Lys Ala Leu Lys Arg Met Gln
    50                  55                  60

Gln Gln Gln Glu Phe Asp Glu Ser Phe Thr Phe Lys Pro Lys Arg Pro
65                  70                  75                  80

Lys Leu Ser Thr Ala Ala Ala Glu Ile Leu Gln Leu Glu Glu Asp Ser
                85                  90                  95

Thr Ser Gly Glu Gly Lys Ser Pro Leu Gln Gln Glu Glu Lys Glu Val
            100                 105                 110

Glu Val Leu Gln Thr Pro Thr Val Gln Leu Gln Leu Gln Arg Asn Ile
        115                 120                 125

Gln Glu Gln Leu Ala Ile Lys Gln Gln Leu Gln Phe Leu Leu Leu Gln
    130                 135                 140

Leu Leu Lys Thr Gln Ser Asn Leu His Leu Asn Pro Gln Phe Leu Ser
145                 150                 155                 160

Pro Ser
```

<210> 203
<211> 746
<212> PRT
<213> Torque teno virus

<400> 203

```
Met Ala Trp Gly Trp Trp Arg Trp Arg Arg Arg Trp Pro Ala Arg Arg
1               5                   10                  15

Trp Arg Arg Arg Arg Arg Arg Arg Pro Val Arg Arg Thr Arg Ala Arg
            20                  25                  30

Arg Pro Ala Arg Arg Tyr Arg Arg Arg Arg Thr Val Arg Thr Arg Arg
        35                  40                  45
```

```
Arg Arg Trp Gly Arg Arg Arg Tyr Arg Arg Gly Trp Arg Arg Arg Thr
    50                  55                  60

Tyr Val Arg Lys Gly Arg His Arg Lys Lys Lys Lys Arg Leu Ile Leu
65                  70                  75                  80

Arg Gln Trp Gln Pro Ala Thr Arg Arg Arg Cys Thr Ile Thr Gly Tyr
                85                  90                  95

Leu Pro Ile Val Phe Cys Gly His Thr Arg Gly Asn Lys Asn Tyr Ala
            100                 105                 110

Leu His Ser Asp Asp Tyr Thr Pro Gln Gly Gln Pro Phe Gly Gly Ala
        115                 120                 125

Leu Ser Thr Thr Ser Phe Ser Leu Lys Val Leu Phe Asp Gln His Gln
    130                 135                 140

Arg Gly Leu Asn Lys Trp Ser Phe Pro Asn Asp Gln Leu Asp Leu Ala
145                 150                 155                 160

Arg Tyr Arg Gly Cys Lys Phe Ile Phe Tyr Arg Thr Lys Gln Thr Asp
                165                 170                 175

Trp Val Gly Gln Tyr Asp Ile Ser Glu Pro Tyr Lys Leu Asp Lys Tyr
            180                 185                 190

Ser Cys Pro Asn Tyr His Pro Gly Asn Met Ile Lys Ala Lys His Lys
        195                 200                 205

Phe Leu Ile Pro Ser Tyr Asp Thr Asn Pro Arg Gly Arg Gln Lys Ile
    210                 215                 220

Ile Val Lys Ile Pro Pro Pro Asp Leu Phe Val Asp Lys Trp Tyr Thr
225                 230                 235                 240

Gln Glu Asp Leu Cys Ser Val Asn Leu Val Ser Leu Ala Val Ser Ala
                245                 250                 255

Ala Ser Phe Leu His Pro Phe Gly Ser Pro Gln Thr Asp Asn Pro Cys
            260                 265                 270
```

```
Tyr Thr Phe Gln Val Leu Lys Glu Phe Tyr Tyr Gln Ala Ile Gly Phe
        275                 280                 285

Ser Ala Ser Thr Gln Ala Met Thr Ser Val Leu Asp Thr Leu Tyr Thr
    290                 295                 300

Gln Asn Ser Tyr Trp Glu Ser Asn Leu Thr Gln Phe Tyr Val Leu Asn
305                 310                 315                 320

Ala Lys Lys Gly Ser Asp Thr Thr Gln Pro Leu Thr Ser Asn Met Pro
                325                 330                 335

Thr Arg Glu Glu Phe Met Ala Lys Lys Asn Thr Asn Tyr Asn Trp Tyr
            340                 345                 350

Thr Tyr Lys Ala Ala Ser Val Lys Asn Lys Leu His Gln Met Arg Gln
        355                 360                 365

Thr Tyr Phe Glu Glu Leu Thr Ser Lys Gly Pro Gln Thr Thr Lys Ser
    370                 375                 380

Glu Glu Gly Tyr Ser Gln His Trp Thr Thr Pro Ser Thr Asn Ala Tyr
385                 390                 395                 400

Glu Tyr His Leu Gly Met Phe Ser Ala Ile Phe Leu Ala Pro Asp Arg
                405                 410                 415

Pro Val Pro Arg Phe Pro Cys Ala Tyr Gln Asp Val Thr Tyr Asn Pro
            420                 425                 430

Leu Met Asp Lys Gly Val Gly Asn His Ile Trp Phe Gln Tyr Asn Thr
        435                 440                 445

Lys Ala Asp Thr Gln Leu Ile Val Thr Gly Gly Ser Cys Lys Ala His
    450                 455                 460

Ile Gln Asp Ile Pro Leu Trp Ala Ala Phe Tyr Gly Tyr Ser Asp Phe
465                 470                 475                 480
```

```
Ile Glu Ser Glu Leu Gly Pro Phe Val Asp Ala Glu Thr Val Gly Leu
                485                 490                 495

Val Cys Val Ile Cys Pro Tyr Thr Lys Pro Pro Met Tyr Asn Lys Thr
            500                 505                 510

Asn Pro Ala Met Gly Tyr Val Phe Tyr Asp Arg Asn Phe Gly Asp Gly
        515                 520                 525

Lys Trp Thr Asp Gly Arg Gly Lys Ile Glu Pro Tyr Trp Gln Val Arg
    530                 535                 540

Trp Arg Pro Glu Met Leu Phe Gln Glu Thr Val Met Ala Asp Leu Val
545                 550                 555                 560

Gln Thr Gly Pro Phe Ser Tyr Lys Asp Glu Leu Lys Asn Ser Thr Leu
                565                 570                 575

Val Cys Lys Tyr Lys Phe Tyr Phe Thr Trp Gly Gly Asn Met Met Phe
            580                 585                 590

Gln Gln Thr Ile Lys Asn Pro Cys Lys Thr Asp Gly Gln Pro Thr Asp
        595                 600                 605

Ser Ser Arg His Pro Arg Gly Ile Gln Val Ala Asp Pro Glu Gln Met
    610                 615                 620

Gly Pro Arg Trp Val Phe His Ser Phe Asp Trp Arg Arg Gly Tyr Leu
625                 630                 635                 640

Ser Glu Lys Ala Leu Lys Arg Leu Gln Glu Lys Pro Leu Asp Tyr Asp
                645                 650                 655

Glu Tyr Phe Thr Gln Pro Lys Arg Pro Arg Ile Phe Pro Pro Thr Glu
            660                 665                 670

Ser Ala Glu Gly Glu Phe Arg Glu Pro Glu Lys Gly Ser Tyr Ser Glu
        675                 680                 685

Glu Glu Arg Ser Gln Ala Ser Ala Glu Glu Gln Thr Gln Glu Ala Thr
    690                 695                 700
```

```
Val Leu Leu Leu Lys Arg Arg Leu Arg Glu Gln Gln Gln Leu Gln Gln
705                 710                 715                 720


Gln Leu Gln Phe Leu Thr Arg Glu Met Phe Lys Thr Gln Ala Gly Leu
                725                 730                 735


His Leu Asn Pro Met Leu Leu Asn Gln Arg
            740                 745


<210> 204
<211> 77
<212> PRT
<213> Torque teno virus

<400> 204
Met Ala Trp Gly Trp Trp Arg Trp Arg Arg Arg Trp Pro Ala Arg Arg
1               5                   10                  15


Trp Arg Arg Arg Arg Arg Arg Arg Pro Val Arg Arg Thr Arg Ala Arg
            20                  25                  30


Arg Pro Ala Arg Arg Tyr Arg Arg Arg Arg Thr Val Arg Thr Arg Arg
        35                  40                  45


Arg Arg Trp Gly Arg Arg Arg Tyr Arg Arg Gly Trp Arg Arg Arg Thr
    50                  55                  60


Tyr Val Arg Lys Gly Arg His Arg Lys Lys Lys Lys Arg
65                  70                  75


<210> 205
<211> 209
<212> PRT
<213> Torque teno virus

<400> 205
Leu Ile Leu Arg Gln Trp Gln Pro Ala Thr Arg Arg Arg Cys Thr Ile
1               5                   10                  15


Thr Gly Tyr Leu Pro Ile Val Phe Cys Gly His Thr Arg Gly Asn Lys
            20                  25                  30
```

```
Asn Tyr Ala Leu His Ser Asp Asp Tyr Thr Pro Gln Gly Gln Pro Phe
        35                  40                  45

Gly Gly Ala Leu Ser Thr Thr Ser Phe Ser Leu Lys Val Leu Phe Asp
    50                  55                  60

Gln His Gln Arg Gly Leu Asn Lys Trp Ser Phe Pro Asn Asp Gln Leu
65                  70                  75                  80

Asp Leu Ala Arg Tyr Arg Gly Cys Lys Phe Ile Phe Tyr Arg Thr Lys
                85                  90                  95

Gln Thr Asp Trp Val Gly Gln Tyr Asp Ile Ser Glu Pro Tyr Lys Leu
            100                 105                 110

Asp Lys Tyr Ser Cys Pro Asn Tyr His Pro Gly Asn Met Ile Lys Ala
        115                 120                 125

Lys His Lys Phe Leu Ile Pro Ser Tyr Asp Thr Asn Pro Arg Gly Arg
    130                 135                 140

Gln Lys Ile Ile Val Lys Ile Pro Pro Pro Asp Leu Phe Val Asp Lys
145                 150                 155                 160

Trp Tyr Thr Gln Glu Asp Leu Cys Ser Val Asn Leu Val Ser Leu Ala
                165                 170                 175

Val Ser Ala Ala Ser Phe Leu His Pro Phe Gly Ser Pro Gln Thr Asp
            180                 185                 190

Asn Pro Cys Tyr Thr Phe Gln Val Leu Lys Glu Phe Tyr Tyr Gln Ala
        195                 200                 205

Ile
```

<210> 206
<211> 130
<212> PRT
<213> Torque teno virus

<400> 206

```
Gly Phe Ser Ala Ser Thr Gln Ala Met Thr Ser Val Leu Asp Thr Leu
1               5                   10                  15


Tyr Thr Gln Asn Ser Tyr Trp Glu Ser Asn Leu Thr Gln Phe Tyr Val
            20                  25                  30


Leu Asn Ala Lys Lys Gly Ser Asp Thr Thr Gln Pro Leu Thr Ser Asn
        35                  40                  45


Met Pro Thr Arg Glu Glu Phe Met Ala Lys Lys Asn Thr Asn Tyr Asn
    50                  55                  60


Trp Tyr Thr Tyr Lys Ala Ala Ser Val Lys Asn Lys Leu His Gln Met
65                  70                  75                  80


Arg Gln Thr Tyr Phe Glu Glu Leu Thr Ser Lys Gly Pro Gln Thr Thr
                85                  90                  95


Lys Ser Glu Glu Gly Tyr Ser Gln His Trp Thr Thr Pro Ser Thr Asn
            100                 105                 110


Ala Tyr Glu Tyr His Leu Gly Met Phe Ser Ala Ile Phe Leu Ala Pro
        115                 120                 125


Asp Arg
    130


<210> 207
<211> 169
<212> PRT
<213> Torque teno virus

<400> 207
Pro Val Pro Arg Phe Pro Cys Ala Tyr Gln Asp Val Thr Tyr Asn Pro
1               5                   10                  15


Leu Met Asp Lys Gly Val Gly Asn His Ile Trp Phe Gln Tyr Asn Thr
            20                  25                  30


Lys Ala Asp Thr Gln Leu Ile Val Thr Gly Gly Ser Cys Lys Ala His
        35                  40                  45
```

```
Ile Gln Asp Ile Pro Leu Trp Ala Ala Phe Tyr Gly Tyr Ser Asp Phe
    50                  55                  60

Ile Glu Ser Glu Leu Gly Pro Phe Val Asp Ala Glu Thr Val Gly Leu
65                  70                  75                  80

Val Cys Val Ile Cys Pro Tyr Thr Lys Pro Pro Met Tyr Asn Lys Thr
                85                  90                  95

Asn Pro Ala Met Gly Tyr Val Phe Tyr Asp Arg Asn Phe Gly Asp Gly
            100                 105                 110

Lys Trp Thr Asp Gly Arg Gly Lys Ile Glu Pro Tyr Trp Gln Val Arg
        115                 120                 125

Trp Arg Pro Glu Met Leu Phe Gln Glu Thr Val Met Ala Asp Leu Val
    130                 135                 140

Gln Thr Gly Pro Phe Ser Tyr Lys Asp Glu Leu Lys Asn Ser Thr Leu
145                 150                 155                 160

Val Cys Lys Tyr Lys Phe Tyr Phe Thr
                165
```

<210> 208
<211> 161
<212> PRT
<213> Torque teno virus

<400> 208

```
Trp Gly Gly Asn Met Met Phe Gln Gln Thr Ile Lys Asn Pro Cys Lys
1               5                   10                  15

Thr Asp Gly Gln Pro Thr Asp Ser Ser Arg His Pro Arg Gly Ile Gln
            20                  25                  30

Val Ala Asp Pro Glu Gln Met Gly Pro Arg Trp Val Phe His Ser Phe
        35                  40                  45

Asp Trp Arg Arg Gly Tyr Leu Ser Glu Lys Ala Leu Lys Arg Leu Gln
    50                  55                  60
```

Glu Lys Pro Leu Asp Tyr Asp Glu Tyr Phe Thr Gln Pro Lys Arg Pro
65 70 75 80

Arg Ile Phe Pro Pro Thr Glu Ser Ala Glu Gly Glu Phe Arg Glu Pro
85 90 95

Glu Lys Gly Ser Tyr Ser Glu Glu Glu Arg Ser Gln Ala Ser Ala Glu
100 105 110

Glu Gln Thr Gln Glu Ala Thr Val Leu Leu Leu Lys Arg Arg Leu Arg
115 120 125

Glu Gln Gln Gln Leu Gln Gln Gln Leu Gln Phe Leu Thr Arg Glu Met
130 135 140

Phe Lys Thr Gln Ala Gly Leu His Leu Asn Pro Met Leu Leu Asn Gln
145 150 155 160

Arg

<210> 209
<211> 765
<212> PRT
<213> Torque teno virus

<400> 209
Met Ala Trp Ser Trp Trp Trp Gln Arg Trp Arg Arg Arg Arg Trp Lys
1 5 10 15

Pro Arg Arg Arg Arg Trp Arg Arg Leu Arg Trp Arg Arg Pro Arg Arg
20 25 30

Ala Val Arg Arg Arg Arg Arg Gly Arg Arg Val Arg Arg Arg Arg Trp
35 40 45

Ala Arg Arg Arg Gly Arg Arg Arg Arg Tyr Ala Thr Arg Arg Lys Arg
50 55 60

Arg Tyr Arg Gly Arg Arg Phe Lys Lys Lys Leu Val Leu Thr Gln Trp
65 70 75 80

```
His Pro Asn Thr Met Arg Arg Cys Leu Ile Lys Gly Ile Val Pro Leu
                85                  90                  95

Val Ile Cys Gly His Thr Arg Trp Asn Tyr Asn Tyr Ala Leu His Ser
            100                 105                 110

Lys Asp Tyr Thr Glu Glu Gly Arg Tyr Pro His Gly Gly Ala Leu Ser
        115                 120                 125

Thr Thr Thr Trp Ser Leu Lys Val Leu Tyr Asp Glu His Leu Lys His
    130                 135                 140

His Asp Phe Trp Gly Tyr Pro Asn Asn Gln Leu Asp Leu Ala Arg Tyr
145                 150                 155                 160

Lys Gly Ala Lys Phe Thr Phe Tyr Arg His Lys Lys Thr Asp Phe Ile
                165                 170                 175

Ile Phe Phe Asn Arg Lys Pro Pro Phe Lys Leu Asn Lys Tyr Ser Cys
            180                 185                 190

Ala Ser Tyr His Pro Gly Met Leu Met Gln Gln Arg His Lys Ile Leu
        195                 200                 205

Leu Pro Ser Tyr Glu Thr Lys Pro Lys Gly Arg Pro Lys Ile Thr Val
    210                 215                 220

Arg Ile Lys Pro Pro Thr Leu Leu Glu Asp Lys Trp Tyr Thr Gln Gln
225                 230                 235                 240

Asp Leu Cys Asp Val Asn Leu Leu Gln Leu Val Val Thr Ala Ala Asp
                245                 250                 255

Phe Arg His Pro Leu Cys Ser Pro Gln Thr Asn Thr Pro Thr Thr Thr
            260                 265                 270

Phe Gln Val Leu Lys Asp Ile Tyr Tyr Asp Thr Met Ser Ile Ser Glu
        275                 280                 285

Pro Thr Asp Ser Tyr Thr Ser Val Asn Asn Lys Ser Thr Thr Gln Thr
    290                 295                 300
```

```
Phe Thr Asn Tyr Ser Asn Thr Leu Glu Asn Ile Leu Tyr Thr Arg Ala
305                 310                 315                 320

Ser Tyr Trp Asn Ser Phe His Ala Thr Glu Tyr Leu Asn Pro Asn Ile
                325                 330                 335

Ile Tyr Lys Asn Gly Glu Lys Leu Phe Lys Glu His Glu Asp Leu Ile
            340                 345                 350

Thr Trp Met Thr Gln Thr Asn Asn Thr Gly Phe Leu Thr Lys Asn Asn
        355                 360                 365

Thr Ala Phe Gly Asn Asn Ser Tyr Arg Pro Asn Ala Asp Lys Ile Lys
    370                 375                 380

Lys Ala Arg Lys Thr Tyr Trp Asn Ala Leu Ile Gly Thr Asn Asp Leu
385                 390                 395                 400

Ala Thr Asn Ile Gly Gln Ala Arg Ala Glu Arg Phe Glu Tyr His Leu
                405                 410                 415

Gly Trp Tyr Ser Pro Ile Phe Leu Ser Arg His Arg Ser Asn Met Asn
            420                 425                 430

Phe Ala Arg Ala Tyr Gln Asp Val Thr Tyr Asn Pro Asn Cys Asp Arg
        435                 440                 445

Gly Val Asn Asn Arg Val Trp Val Gln Pro Leu Thr Lys Pro Thr Thr
    450                 455                 460

Glu Phe Asp Glu Lys Arg Cys Lys Cys Val Val Gln His Leu Pro Leu
465                 470                 475                 480

Trp Ala Ala Leu Tyr Cys Tyr Gln Asp Phe Val Glu Glu Glu Leu Gly
                485                 490                 495

Ser Ser Ser Glu Ile Leu Asn Ser Cys Leu Leu Val Leu Gln Cys Pro
            500                 505                 510
```

```
Tyr Thr Phe Pro Pro Met Tyr Asp Lys Lys Leu Pro Asp Lys Gly Phe
        515                 520                 525

Val Phe Tyr Asp Ser Leu Phe Gly Asp Gly Lys Met Ser Asp Gly Arg
    530                 535                 540

Gly Gln Val Asp Ile Phe Trp Gln Gln Arg Trp Tyr Pro Arg Leu Ala
545                 550                 555                 560

Thr Gln Met Gln Val Met His Asp Ile Thr Met Thr Gly Pro Phe Ser
                565                 570                 575

Tyr Arg Asp Glu Leu Val Ser Thr Gln Leu Thr Ala Lys Tyr Thr Phe
            580                 585                 590

Asp Phe Met Trp Gly Gly Asn Met Ile Ser Thr Gln Ile Ile Lys Asn
        595                 600                 605

Pro Cys Lys Asp Ser Gly Leu Glu Pro Ala Tyr Pro Gly Arg Gln Arg
    610                 615                 620

Arg Asp Leu Gln Ile Val Asp Pro Tyr Ser Met Gly Pro Gln Phe Ser
625                 630                 635                 640

Phe His Asn Trp Asp Tyr Arg His Gly Leu Phe Gly Gln Asp Ala Ile
                645                 650                 655

Asp Arg Val Ser Lys Gln Pro Lys Asp Asp Ala Asp Tyr Pro Asn Pro
            660                 665                 670

Tyr Lys Arg Pro Arg Tyr Phe Pro Pro Thr Asp Gln Ala Ala Gln Glu
        675                 680                 685

Gln Glu Lys Asp Phe Ser Phe Leu Lys Thr Ala Pro Ser Asn Ser Glu
    690                 695                 700

Glu Ser Asp Gln Glu Val Leu Gln Glu Thr Gln Val Leu Arg Phe Gln
705                 710                 715                 720

Pro Glu Gln His Lys Gln Leu His Leu Gln Leu Ala Glu Arg Gln Arg
                725                 730                 735
```

```
Ile Gly Glu Gln Leu Arg Tyr Leu Leu Gln Gln Met Phe Lys Thr Gln
            740                 745                 750


Ala Asn Leu His Leu Asn Pro Tyr Thr Phe Thr Gln Leu
        755                 760                 765


<210> 210
<211> 74
<212> PRT
<213> Torque teno virus

<400> 210
Met Ala Trp Ser Trp Trp Trp Gln Arg Trp Arg Arg Arg Arg Trp Lys
1               5                   10                  15


Pro Arg Arg Arg Arg Trp Arg Arg Leu Arg Trp Arg Arg Pro Arg Arg
            20                  25                  30


Ala Val Arg Arg Arg Arg Arg Gly Arg Arg Val Arg Arg Arg Arg Trp
        35                  40                  45


Ala Arg Arg Arg Gly Arg Arg Arg Arg Tyr Ala Thr Arg Arg Lys Arg
    50                  55                  60


Arg Tyr Arg Gly Arg Arg Phe Lys Lys Lys
65                  70


<210> 211
<211> 212
<212> PRT
<213> Torque teno virus

<400> 211
Leu Val Leu Thr Gln Trp His Pro Asn Thr Met Arg Arg Cys Leu Ile
1               5                   10                  15


Lys Gly Ile Val Pro Leu Val Ile Cys Gly His Thr Arg Trp Asn Tyr
            20                  25                  30


Asn Tyr Ala Leu His Ser Lys Asp Tyr Thr Glu Glu Gly Arg Tyr Pro
        35                  40                  45
```

```
His Gly Gly Ala Leu Ser Thr Thr Thr Trp Ser Leu Lys Val Leu Tyr
    50                  55                  60

Asp Glu His Leu Lys His His Asp Phe Trp Gly Tyr Pro Asn Asn Gln
65                  70                  75                  80

Leu Asp Leu Ala Arg Tyr Lys Gly Ala Lys Phe Thr Phe Tyr Arg His
                85                  90                  95

Lys Lys Thr Asp Phe Ile Ile Phe Phe Asn Arg Lys Pro Pro Phe Lys
            100                 105                 110

Leu Asn Lys Tyr Ser Cys Ala Ser Tyr His Pro Gly Met Leu Met Gln
        115                 120                 125

Gln Arg His Lys Ile Leu Leu Pro Ser Tyr Glu Thr Lys Pro Lys Gly
    130                 135                 140

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

Lys Trp Tyr Thr Gln Gln Asp Leu Cys Asp Val Asn Leu Leu Gln Leu
                165                 170                 175

Val Val Thr Ala Ala Asp Phe Arg His Pro Leu Cys Ser Pro Gln Thr
            180                 185                 190

Asn Thr Pro Thr Thr Thr Phe Gln Val Leu Lys Asp Ile Tyr Tyr Asp
        195                 200                 205

Thr Met Ser Ile
    210
```

<210> 212
<211> 142
<212> PRT
<213> Torque teno virus

<400> 212

```
Ser Glu Pro Thr Asp Ser Tyr Thr Ser Val Asn Asn Lys Ser Thr Thr
1               5                   10                  15
```

```
Gln Thr Phe Thr Asn Tyr Ser Asn Thr Leu Glu Asn Ile Leu Tyr Thr
            20                  25                  30

Arg Ala Ser Tyr Trp Asn Ser Phe His Ala Thr Glu Tyr Leu Asn Pro
        35                  40                  45

Asn Ile Ile Tyr Lys Asn Gly Glu Lys Leu Phe Lys Glu His Glu Asp
    50                  55                  60

Leu Ile Thr Trp Met Thr Gln Thr Asn Asn Thr Gly Phe Leu Thr Lys
65                  70                  75                  80

Asn Asn Thr Ala Phe Gly Asn Asn Ser Tyr Arg Pro Asn Ala Asp Lys
                85                  90                  95

Ile Lys Lys Ala Arg Lys Thr Tyr Trp Asn Ala Leu Ile Gly Thr Asn
            100                 105                 110

Asp Leu Ala Thr Asn Ile Gly Gln Ala Arg Ala Glu Arg Phe Glu Tyr
        115                 120                 125

His Leu Gly Trp Tyr Ser Pro Ile Phe Leu Ser Arg His Arg
    130                 135                 140
```

<210> 213
<211> 167
<212> PRT
<213> Torque teno virus

<400> 213

```
Ser Asn Met Asn Phe Ala Arg Ala Tyr Gln Asp Val Thr Tyr Asn Pro
1               5                   10                  15

Asn Cys Asp Arg Gly Val Asn Asn Arg Val Trp Val Gln Pro Leu Thr
            20                  25                  30

Lys Pro Thr Thr Glu Phe Asp Glu Lys Arg Cys Lys Cys Val Val Gln
        35                  40                  45

His Leu Pro Leu Trp Ala Ala Leu Tyr Cys Tyr Gln Asp Phe Val Glu
    50                  55                  60
```

```
Glu Glu Leu Gly Ser Ser Ser Glu Ile Leu Asn Ser Cys Leu Leu Val
65                  70                  75                  80


Leu Gln Cys Pro Tyr Thr Phe Pro Pro Met Tyr Asp Lys Lys Leu Pro
                85                  90                  95


Asp Lys Gly Phe Val Phe Tyr Asp Ser Leu Phe Gly Asp Gly Lys Met
            100                 105                 110


Ser Asp Gly Arg Gly Gln Val Asp Ile Phe Trp Gln Gln Arg Trp Tyr
        115                 120                 125


Pro Arg Leu Ala Thr Gln Met Gln Val Met His Asp Ile Thr Met Thr
    130                 135                 140


Gly Pro Phe Ser Tyr Arg Asp Glu Leu Val Ser Thr Gln Leu Thr Ala
145                 150                 155                 160


Lys Tyr Thr Phe Asp Phe Met
                165


<210> 214
<211> 170
<212> PRT
<213> Torque teno virus

<400> 214
Trp Gly Gly Asn Met Ile Ser Thr Gln Ile Ile Lys Asn Pro Cys Lys
1               5                   10                  15


Asp Ser Gly Leu Glu Pro Ala Tyr Pro Gly Arg Gln Arg Arg Asp Leu
            20                  25                  30


Gln Ile Val Asp Pro Tyr Ser Met Gly Pro Gln Phe Ser Phe His Asn
        35                  40                  45


Trp Asp Tyr Arg His Gly Leu Phe Gly Gln Asp Ala Ile Asp Arg Val
    50                  55                  60


Ser Lys Gln Pro Lys Asp Asp Ala Asp Tyr Pro Asn Pro Tyr Lys Arg
65                  70                  75                  80
```

```
Pro Arg Tyr Phe Pro Pro Thr Asp Gln Ala Ala Gln Glu Gln Glu Lys
                85                  90                  95

Asp Phe Ser Phe Leu Lys Thr Ala Pro Ser Asn Ser Glu Glu Ser Asp
            100                 105                 110

Gln Glu Val Leu Gln Glu Thr Gln Val Leu Arg Phe Gln Pro Glu Gln
        115                 120                 125

His Lys Gln Leu His Leu Gln Leu Ala Glu Arg Gln Arg Ile Gly Glu
    130                 135                 140

Gln Leu Arg Tyr Leu Leu Gln Gln Met Phe Lys Thr Gln Ala Asn Leu
145                 150                 155                 160

His Leu Asn Pro Tyr Thr Phe Thr Gln Leu
                165                 170
```

<210> 215
<211> 666
<212> PRT
<213> TTV-like mini virus

<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> TTV-like mini virus

<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> TTV-like mini virus

<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> TTV-like mini virus

<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> TTV-like mini virus

<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> TTV-like mini virus

<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
<211> 673
<212> PRT
<213> Torque teno midi virus 1

<400> 221

```
Met Pro Phe Trp Trp Gly Arg Arg Asn Lys Phe Trp Tyr Gly Arg Asn
1               5                   10                  15

Tyr Arg Arg Lys Lys Arg Arg Phe Pro Lys Arg Arg Lys Arg Arg Phe
            20                  25                  30

Tyr Arg Arg Thr Lys Tyr Arg Arg Pro Ala Arg Arg Arg Arg Arg Arg
        35                  40                  45

Arg Arg Lys Val Arg Arg Lys Lys Lys Thr Leu Ile Val Arg Gln Trp
    50                  55                  60

Gln Pro Asp Ser Ile Val Leu Cys Lys Ile Lys Gly Tyr Asp Ser Ile
65                  70                  75                  80

Ile Trp Gly Ala Glu Gly Thr Gln Phe Gln Cys Ser Thr His Glu Met
                85                  90                  95
```

```
Tyr Glu Tyr Thr Arg Gln Lys Tyr Pro Gly Gly Gly Gly Phe Gly Val
            100                 105                 110

Gln Leu Tyr Ser Leu Glu Tyr Leu Tyr Asp Gln Trp Lys Leu Arg Asn
        115                 120                 125

Asn Ile Trp Thr Lys Thr Asn Gln Leu Lys Asp Leu Cys Arg Tyr Leu
    130                 135                 140

Lys Cys Val Met Thr Phe Tyr Arg His Gln His Ile Asp Phe Val Ile
145                 150                 155                 160

Val Tyr Glu Arg Gln Pro Pro Phe Glu Ile Asp Lys Leu Thr Tyr Met
                165                 170                 175

Lys Tyr His Pro Tyr Met Leu Leu Gln Arg Lys His Lys Ile Ile Leu
            180                 185                 190

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

Ile Lys Pro Pro Lys Gln Met Leu Ser Lys Trp Phe Phe Gln Gln Gln
    210                 215                 220

Phe Ala Lys Tyr Asp Leu Leu Leu Ile Ala Ala Ala Ala Cys Ser Leu
225                 230                 235                 240

Arg Tyr Pro Arg Ile Gly Cys Cys Asn Glu Asn Arg Met Ile Thr Leu
                245                 250                 255

Tyr Cys Leu Asn Thr Lys Phe Tyr Gln Asp Thr Glu Trp Gly Thr Thr
            260                 265                 270

Lys Gln Ala Pro His Tyr Phe Lys Pro Tyr Ala Thr Ile Asn Lys Ser
        275                 280                 285

Met Ile Phe Val Ser Asn Tyr Gly Gly Lys Lys Thr Glu Tyr Asn Ile
    290                 295                 300
```

```
Gly Gln Trp Ile Glu Thr Asp Ile Pro Gly Glu Gly Asn Leu Ala Arg
305                 310                 315                 320

Tyr Tyr Arg Ser Ile Ser Lys Glu Gly Gly Tyr Phe Ser Pro Lys Ile
                325                 330                 335

Leu Gln Ala Tyr Gln Thr Lys Val Lys Ser Val Asp Tyr Lys Pro Leu
            340                 345                 350

Pro Ile Val Leu Gly Arg Tyr Asn Pro Ala Ile Asp Asp Gly Lys Gly
        355                 360                 365

Asn Lys Ile Tyr Leu Gln Thr Ile Met Asn Gly His Trp Gly Leu Pro
    370                 375                 380

Gln Lys Thr Pro Asp Tyr Ile Ile Glu Glu Val Pro Leu Trp Leu Gly
385                 390                 395                 400

Phe Trp Gly Tyr Tyr Asn Tyr Leu Lys Gln Thr Arg Thr Glu Ala Ile
                405                 410                 415

Phe Pro Leu His Met Phe Val Val Gln Ser Lys Tyr Ile Gln Thr Gln
            420                 425                 430

Gln Thr Glu Thr Pro Asn Asn Phe Trp Ala Phe Ile Asp Asn Ser Phe
        435                 440                 445

Ile Gln Gly Lys Asn Pro Trp Asp Ser Val Ile Thr Tyr Ser Glu Gln
    450                 455                 460

Lys Leu Trp Phe Pro Thr Val Ala Trp Gln Leu Lys Thr Ile Asn Ala
465                 470                 475                 480

Ile Cys Glu Ser Gly Pro Tyr Val Pro Lys Leu Asp Asn Gln Thr Tyr
                485                 490                 495

Ser Thr Trp Glu Leu Ala Thr His Tyr Ser Phe His Phe Lys Trp Gly
            500                 505                 510

Gly Pro Gln Ile Ser Asp Gln Pro Val Glu Asp Pro Gly Asn Lys Asn
        515                 520                 525
```

```
Lys Tyr Asp Val Pro Asp Thr Ile Lys Glu Ala Leu Gln Ile Val Asn
    530                 535                 540

Pro Ala Lys Asn Ile Ala Ala Thr Met Phe His Asp Trp Asp Tyr Arg
545                 550                 555                 560

Arg Gly Cys Ile Thr Ser Thr Ala Ile Lys Arg Met Gln Gln Asn Leu
                565                 570                 575

Pro Thr Asp Ser Ser Leu Glu Ser Asp Ser Asp Ser Glu Pro Ala Pro
            580                 585                 590

Lys Lys Lys Arg Leu Leu Pro Val Leu His Asp Pro Gln Lys Lys Thr
        595                 600                 605

Glu Lys Ile Asn Gln Cys Leu Leu Ser Leu Cys Glu Glu Ser Thr Cys
    610                 615                 620

Gln Glu Gln Glu Thr Glu Glu Asn Ile Leu Lys Leu Ile Gln Gln Gln
625                 630                 635                 640

Gln Gln Gln Gln Gln Lys Leu Lys His Asn Leu Leu Val Leu Ile Lys
                645                 650                 655

Asp Leu Lys Val Lys Gln Arg Leu Leu Gln Leu Gln Thr Gly Val Leu
            660                 665                 670

Glu
```

<210> 222
<211> 57
<212> PRT
<213> Torque teno midi virus 1

<400> 222

```
Met Pro Phe Trp Trp Gly Arg Arg Asn Lys Phe Trp Tyr Gly Arg Asn
1               5                   10                  15

Tyr Arg Arg Lys Lys Arg Arg Phe Pro Lys Arg Arg Lys Arg Arg Phe
            20                  25                  30
```

Tyr Arg Arg Thr Lys Tyr Arg Arg Pro Ala Arg Arg Arg Arg Arg Arg
        35   40   45

Arg Arg Lys Val Arg Arg Lys Lys Lys
    50   55

<210> 223
<211> 202
<212> PRT
<213> Torque teno midi virus 1

<400> 223
Thr Leu Ile Val Arg Gln Trp Gln Pro Asp Ser Ile Val Leu Cys Lys
1   5   10   15

Ile Lys Gly Tyr Asp Ser Ile Ile Trp Gly Ala Glu Gly Thr Gln Phe
            20   25   30

Gln Cys Ser Thr His Glu Met Tyr Glu Tyr Thr Arg Gln Lys Tyr Pro
        35   40   45

Gly Gly Gly Gly Phe Gly Val Gln Leu Tyr Ser Leu Glu Tyr Leu Tyr
    50   55   60

Asp Gln Trp Lys Leu Arg Asn Asn Ile Trp Thr Lys Thr Asn Gln Leu
65   70   75   80

Lys Asp Leu Cys Arg Tyr Leu Lys Cys Val Met Thr Phe Tyr Arg His
                85   90   95

Gln His Ile Asp Phe Val Ile Val Tyr Glu Arg Gln Pro Pro Phe Glu
            100   105   110

Ile Asp Lys Leu Thr Tyr Met Lys Tyr His Pro Tyr Met Leu Leu Gln
        115   120   125

Arg Lys His Lys Ile Ile Leu Pro Ser Gln Thr Thr Asn Pro Arg Gly
    130   135   140

Lys Leu Lys Lys Lys Lys Thr Ile Lys Pro Pro Lys Gln Met Leu Ser
145   150   155   160

Lys Trp Phe Phe Gln Gln Gln Phe Ala Lys Tyr Asp Leu Leu Leu Ile
165 170 175

Ala Ala Ala Ala Cys Ser Leu Arg Tyr Pro Arg Ile Gly Cys Cys Asn
180 185 190

Glu Asn Arg Met Ile Thr Leu Tyr Cys Leu
195 200

<210> 224
<211> 92
<212> PRT
<213> Torque teno midi virus 1

<400> 224
Asn Thr Lys Phe Tyr Gln Asp Thr Glu Trp Gly Thr Thr Lys Gln Ala
1 5 10 15

Pro His Tyr Phe Lys Pro Tyr Ala Thr Ile Asn Lys Ser Met Ile Phe
20 25 30

Val Ser Asn Tyr Gly Gly Lys Lys Thr Glu Tyr Asn Ile Gly Gln Trp
35 40 45

Ile Glu Thr Asp Ile Pro Gly Glu Gly Asn Leu Ala Arg Tyr Tyr Arg
50 55 60

Ser Ile Ser Lys Glu Gly Gly Tyr Phe Ser Pro Lys Ile Leu Gln Ala
65 70 75 80

Tyr Gln Thr Lys Val Lys Ser Val Asp Tyr Lys Pro
85 90

<210> 225
<211> 159
<212> PRT
<213> Torque teno midi virus 1

<400> 225
Leu Pro Ile Val Leu Gly Arg Tyr Asn Pro Ala Ile Asp Asp Gly Lys
1 5 10 15

Gly Asn Lys Ile Tyr Leu Gln Thr Ile Met Asn Gly His Trp Gly Leu
            20   25   30

Pro Gln Lys Thr Pro Asp Tyr Ile Ile Glu Glu Val Pro Leu Trp Leu
        35   40   45

Gly Phe Trp Gly Tyr Tyr Asn Tyr Leu Lys Gln Thr Arg Thr Glu Ala
    50   55   60

Ile Phe Pro Leu His Met Phe Val Val Gln Ser Lys Tyr Ile Gln Thr
65   70   75   80

Gln Gln Thr Glu Thr Pro Asn Asn Phe Trp Ala Phe Ile Asp Asn Ser
                85   90   95

Phe Ile Gln Gly Lys Asn Pro Trp Asp Ser Val Ile Thr Tyr Ser Glu
            100   105   110

Gln Lys Leu Trp Phe Pro Thr Val Ala Trp Gln Leu Lys Thr Ile Asn
        115   120   125

Ala Ile Cys Glu Ser Gly Pro Tyr Val Pro Lys Leu Asp Asn Gln Thr
    130   135   140

Tyr Ser Thr Trp Glu Leu Ala Thr His Tyr Ser Phe His Phe Lys
145   150   155

<210> 226
<211> 163
<212> PRT
<213> Torque teno midi virus 1

<400> 226
Trp Gly Gly Pro Gln Ile Ser Asp Gln Pro Val Glu Asp Pro Gly Asn
1   5   10   15

Lys Asn Lys Tyr Asp Val Pro Asp Thr Ile Lys Glu Ala Leu Gln Ile
            20   25   30

Val Asn Pro Ala Lys Asn Ile Ala Ala Thr Met Phe His Asp Trp Asp
        35   40   45

```
Tyr Arg Arg Gly Cys Ile Thr Ser Thr Ala Ile Lys Arg Met Gln Gln
    50                  55                  60

Asn Leu Pro Thr Asp Ser Ser Leu Glu Ser Asp Ser Asp Ser Glu Pro
65                  70                  75                  80

Ala Pro Lys Lys Lys Arg Leu Leu Pro Val Leu His Asp Pro Gln Lys
                85                  90                  95

Lys Thr Glu Lys Ile Asn Gln Cys Leu Leu Ser Leu Cys Glu Glu Ser
            100                 105                 110

Thr Cys Gln Glu Gln Glu Thr Glu Glu Asn Ile Leu Lys Leu Ile Gln
        115                 120                 125

Gln Gln Gln Gln Gln Gln Gln Lys Leu Lys His Asn Leu Leu Val Leu
    130                 135                 140

Ile Lys Asp Leu Lys Val Lys Gln Arg Leu Leu Gln Leu Gln Thr Gly
145                 150                 155                 160

Val Leu Glu
```

<210> 227
<211> 220
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<220>
<221> MOD_RES
<222> (29)..(31)
<223> Any amino acid

<220>
<221> SITE
<222> (29)..(31)
<223> /note="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> /note="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>
<221> source
<223> /note="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> /note="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> /note="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>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic polypeptide"

<220>
<221> MOD_RES
<222> (20)..(22)
<223> Any amino acid

<220>
<221> SITE
<222> (20)..(22)
<223> /note="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> /note="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> /note="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> /note="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> /note="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> /note="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>
<221> source
<223> /note="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> /note="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>
<221> source
```

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<220>
<221> MOD_RES
<222> (1)..(10)
<223> Any amino acid

<220>
<221> SITE
<222> (1)..(10)
<223> /note="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> /note="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> /note="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> /note="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> /note="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>
<221> source
<223> /note="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> /note="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> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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> VARIANT
<222> (58)..(58)
<223> /replace="Ile"

<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> VARIANT
<222> (111)..(111)
<223> /replace="Ile"

<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

<220>
<221> SITE
<222> (1)..(162)
<223> /note="Variant residues given in the sequence have no
preference with respect to those in the annotations
for variant positions"

<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 Leu 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 Leu 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>
<221> source
<223> /note="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> /note="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> /note="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> /note="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> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 300
gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 301

<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 301
gcguacguca caagucacgu ggaggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 302
<211> 77
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 302
gccauuuuaa guagcugacg ucaaggauug acgugaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggug 77

<210> 303
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 303
gcacacguca uaagucacgu ggugggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugauuuguca cgugugua 78

<210> 304
<211> 66
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 304
cuuccggguc auaggucaca ccuacgucac aagucacgug gggaggguug gcguauagcc 60

cggaag 66

<210> 305
<211> 68
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 305
gccgggggggc ugccgccccc cccggggaaa gggggggggcc ccccccgggg gggggguuugc 60

cccccggc 68

<210> 306
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 306
auacgucauc agucacgugg gggaaggcgu gccuaaaccc ggaagcaucc ucguccacgu 60

gacugugacg uguguggc 78

<210> 307
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 307
cauuuuaagu aaggcggaag cagcucggcg uacacaaaau ggcggcggag cacuuccggc 60

uugcccaaaa ugg 73

<210> 308
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 308
gucacaaguc acguggggag gguuggcguu uaacccggaa gccaauccuc uuacguggcc 60

ugucacguga c 71

<210> 309
<211> 70
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 309
cgaccgcguc ccgaaggcgg guacccgagg ugaguuuaca caccgagguu aagggccaau 60

ucgggcuugg 70

<210> 310
<211> 59
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 310
cgcgguaucg uagccgacgc ggaccccguu uucggggccc ccgcggggcu cucggcgcg 59

<210> 311
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 311
cgccauuuug ugauacgcgc gucccccuccc ggcuuccgua caacgucagg cggggcgugg 60

ccguaucaga aaauggcg 78

<210> 312
<211> 77
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 312
gcuacgucau aagucacgug acugggcagg uacuaaaccc ggaaguaucc ucggucacgu 60

ggccugucac guaguug 77

<210> 313
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 313
ggcusugacg ucaaagucac gugggraggg uggcguuaaa cccggaaguc auccucguca 60

cgugaccuga cgucacagcc 80

<210> 314
<211> 66
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 314

gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

gaaggu 66

<210> 315
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 315
gguugugacg ucaaagucac gugggggaggg cggcguuaaa cccggaaguc auccucguca 60

cgugaccuga cgucacggcc 80

<210> 316
<211> 67
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 316
gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

guaggug 67

<210> 317
<211> 67
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 317
gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

guaggug 67

<210> 318

<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 318
ggcugugacg ucaaagucac gugggggaggg cggcguuaaa cccggaaguc auccucguca 60

cgugaccuga cgucacggcc 80

<210> 319
<211> 79
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 319
agaccacgug guaagucacg ugggggcagc ugcuguaaac ccggaaguag cugacccgcg 60

ugacugguca cgugaccug 79

<210> 320
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 320
cgccauuuua uaauacgcgc gucccccuccc ggcuuccgua cuacgucagg cggggcgugg 60

ccguauuaga aaauggug 78

<210> 321
<211> 72
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 321
uaaguaaggc ggaaccaggc ugucacccug ugucaaaggu caagggacag ccuuccggcu 60

ugcacaaaau gg 72

<210> 322
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 322
ugccuacguc auaagucacg uggggacggc ugcuguaaac acggaaguag cugacccgcg 60

ugacuuguca cgugagca 78

<210> 323
<211> 72
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 323
uuguguaagg cggaacaggc ugacaccccg ugucaaaggu caggggucag ccuccgcuuu 60

gcaccaaaug gu 72

<210> 324
<211> 79
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 324
uaccuacguc auaagucacg ugggaagagc ugcugugaac cuggaaguag cugacccgcg 60

uggcuuguca cgugagugc 79

<210> 325
<211> 75
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 325
uuuuccuggc ccguccgcgg cgagagcgcg agcgaagcga gcgaucgggc gucccgaggg 60

cgggugccgg aggug 75

<210> 326
<211> 68
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 326
aaagugagug gggccagacu ucgccauagg gccuuuaacu uccgggugcg ucugggggcc 60

gccauuuu 68

<210> 327
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 327
gugacguuac ucucacguga ugggggcgug cucuaacccg gaagcauccu cgaccacgug 60

acugugacgu cac 73

<210> 328
<211> 75
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 328
agcgucuacu acguacacuu ccuggggugu guccugccac uguauauaaa ccagaggggu 60

gacgaauggu agagu 75

<210> 329
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 329
gugacgucaa agucacgugg ugacggccau uuuaacccgg aaguggcugu ugucacguga 60

cuugacguca cgg 73

<210> 330
<211> 62
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 330
gcuuuagacg ccauuuuagg cccucgcggg cacccguagg cgcguuuuaa ugacgucacg 60

gc 62

<210> 331
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 331
cacccguagg cgcguuuuaa ugacgucacg gcagccauuu ugucgugacg uuugagacac 60

gugaugggggg cgu 73

<210> 332
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 332
gucgugacgu uugagacacg ugauggggggc gugccuaaac ccggaagcau cccuggucac 60

gugacucuga cgucacggcg 80

<210> 333
<211> 77
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 333
cgaaagugag uggggccaga cuucgccaua aggccuuuaa cuuccgggug cgugugggggg 60

ccgccauuuu agcuucg 77

<210> 334
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 334
cugugacguc aaagucacgu ggggagggcg gcguguaacc cggaagucau ccucgucacg 60

ugaccugacg ucacgg 76

<210> 335
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 335
cuguccgcca ucuugugacu uccuuccgcu uuuucaaaaa aaaagaggaa guaugacgua 60

gcggcggggg ggc 73

<210> 336
<211> 67
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 336
gguagaguuu uuuccgcccg uccgcagcga ggacgcgagc gcagcgagcg gccgagcgac 60

ccguggg 67

<210> 337
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 337
gcugugacgu uucagucacg uggggaggga acgccuaaac ccggaagcgu cccuggucac 60

gugauuguga cgucacggcc 80

<210> 338
<211> 63
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 338
ccgccauuuu gugacuuccu uccgcuuuuu caaaaaaaaa gaggaagugu gacguagcgg 60

cgg 63

<210> 339
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 339
gacugugacg ucaaagucac gugggggaggg cggcguguaa cccggaaguc auccucguca 60

cgugaccuga cgucacgg 78

<210> 340
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 340
cuguccgcca ucuugugacu uccuuccgcu uuuucaaaaa aaaagaggaa guaugacgug 60

gcggcggggg ggc 73

<210> 341
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 341

gguugugacg ucaaagucac gugggggaggg cggcguguaa cccggaaguc auccucguca 60

cgugaccuga cgucacggcc 80

<210> 342
<211> 65
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 342
cccgccaucu ugugacuucc uuccgcuuuu ucaaaaaaaa agaggaagug ugacguagcg 60

gcggg 65

<210> 343
<211> 67
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 343
gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

guaggug 67

<210> 344
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 344
gacugugacg ucaaagucac gugggggagga gggcguguaa cccggaaguc auccucguca 60

cgugaccuga cgucacgg 78

<210> 345

<211> 62
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 345
ucgcgucuua gugacgucac ggcagccauc uugguccuga cgucacuguc acguggggag 60

gg 62

<210> 346
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 346
ugacgucacu gucacguggg gagggaacac gugaacccgg aaguguccu ggucacguga 60

caugacguca cggccg 76

<210> 347
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 347
cgccauuuua aguaagcaug gcgggcggug augucaaaug uuaaagguca cagccgguca 60

ugcuugcaca aaauggcg 78

<210> 348
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 348
cgccauuuua aguaagcaug gcgggcggug acgugcaaug ucaaagguca cagccuguca 60

ugcuugcaca aaauggcg 78

<210> 349
<211> 72
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 349
ccaucuuaag uaguugaggc ggacgguggc gucgguucaa aggucaccau cagccacacc 60

uacucaaaau gg 72

<210> 350
<211> 67
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 350
gccugucaug cuugcacaaa auggcggacu uccgcuuccg ggucgccgcc auauuugguc 60

acgugac 67

<210> 351
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 351
gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 352
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 352
gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 353
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 353
gcauacguca caagucacgu gggggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 354
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 354
gccauuuuaa guagcugacg ucaaggauug acgugaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 355
<211> 78
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 355
gcacacguca uaagucacgu gguggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugauuuguca cgugugua 78

<210> 356
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 356
gccauuuuaa gucagcucug gggaggcgug acuuccaguu caaaggucau ccucaccaua 60

acuggcacaa aauggc 76

<210> 357
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 357
gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 358
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 358
gcauacguca caagucacgu ggaggggaca cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 359
<211> 79
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 359
gcgccauguu aaguggcugu cgccgaggau ugacgucaca guucaaaggu cauccucgac 60

gguaaccgca aacauggcg 79

<210> 360
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 360
caugcgucau aagucacaug acaggggucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 361
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 361
uggcagcacu uccgaauggc ugaguuuucc acgcccgucc gcggagaggg agccacggag 60

gugaucccga acg 73

<210> 362
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 362
gccauuuuaa gucagcgcug gggaggcaug acuguaaguu caaaggucau ccucaccgga 60

acugacacaa aauggccg 78

<210> 363
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 363
gccaucuuaa guggcugucg ccgaggauug acgucacagu ucaaagguca uccucggcgg 60

uaaccgcaaa gauggcgguc 80

<210> 364
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 364
auacgucaua agucacaugu cuaggggucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 365
<211> 77
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 365
ccauuuuaag uaaggcggaa gcagcugucc cuguaacaaa auggcggcga cagccuuccg 60

cuuugcacaa aauggag 77

<210> 366
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 366
gccaucuuaa guggcugucg cugaggauug acgucacagu ucaaagguca uccucggcgg 60

uaaccgcaaa gauggcgguc 80

<210> 367
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 367
cauacgucau aagucacaug acaggagucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 368
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 368

cgccaucuua aguggcuguc gccgaggauu ggcgucacag uucaaagguc auccucggcg 60

guaaccgcaa agauggcggu 80

<210> 369
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 369
cauacgucau aagucacaug acaggggucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 370
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 370
gcauacguca caagucacgu gggggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cguggugu 78

<210> 371
<211> 77
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 371
ccgccauuuu aggcuguugc cgggcguuug acuuccgugu uaaaggucaa acacccagcg 60

acaccaaaaa auggccg 77

<210> 372

<211> 77
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 372
cuacgucaua agucacguga cagggagggg cgacaaaccc ggaagucauc cucgcccacg 60

ugacuuacca cguggug 77

<210> 373
<211> 77
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 373
gccauuuuaa guaggugacg uccaggacug acguaaaguu caaaggucau ccucggcgga 60

accuauacaa aauggcg 77

<210> 374
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 374
cuacgucaua agucacgugg ggacggcugu acuuaaacac ggaaguaggc cccgucacgu 60

gauuuaccac guggug 76

<210> 375
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 375
gccauuuuaa guaaggcgga agagcucuag cuauacaaaa uggcggcgga gcacuuccgc 60

uuugcccaaa aug 73

<210> 376
<211> 77
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 376
gccauuuuaa guagcugacg ucaaggauug acguagaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggug 77

<210> 377
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 377
gcauacguca caagucacgu ggggggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 378
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 378
ggcgccauuu uaaguaagca uggcgggcgg cgacgucaca ugucaaaggu caccgcacuu 60

ccgugcuugc acaaaauggc 80

<210> 379
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 379
ugcuacguca ucgagacacg uggugccagc agcuguaaac ccggaagucg cugacacacg 60

ugucuuguca cgu 73

<210> 380
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 380
gccauuuuaa guaagcaccg ccuagggaug acguauaagu ucaaagguca uccucagccg 60

gaacuuacac aaaauggu 78

<210> 381
<211> 72
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 381
acgucauaug ucacgugggg aggcccugcu gcgcaaacgc ggaaguaggc cccgucacgu 60

gucauaccac gu 72

<210> 382
<211> 77
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 382
ccauuuuaag uaaggcggaa gcagcuccac uuucucacaa aauggcggcg gggcacuucc 60

ggcuugccca aaauggc 77

<210> 383
<211> 72
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 383
ccauuuuaag uaaggcggaa guuucuccac uauacaaaau ggcggcggag cacuuccggc 60

uugcccaaaa ug 72

<210> 384
<211> 72
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 384
ccaucuuaag uaguugaggc ggacgguggc gugaguucaa aggucaccau cagccacacc 60

uacucaaaau gg 72

<210> 385
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 385
cgccaucuua aguaguugag gcggacggug gcgugaguuc aaaggucacc aucagccaca 60

ccuacucaaa auggug 76

<210> 386
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 386
uuucggaccu ucggcgucgg gggggucggg ggcuuuacua aacagacucc gagaugccau 60

uggacacuga ggg 73

<210> 387
<211> 76
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 387
ccauuuuaag uaggugccgu ccagcacugc uguuccgggu uaaagggcau ccucggcgga 60

accuauacaa aauggc 76

<210> 388
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 388
cuacgucauc gaugacgugg ggaggcguac uaugaaacgc ggaaguaggc cccgcuacgu 60

caucaucacg ugg 73

<210> 389
<211> 73
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 389
ccauuuuaag uaaggcggaa gagcugcucu auauacaaaa uggcggagga gcacuuccgg 60

cuugcccaaa aug 73

<210> 390
<211> 75
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 390
ugccuacgua acaagucacg uggggagggu uggcguauaa cccggaaguc aauccuccca 60

cguggccugu cacgu 75

<210> 391
<211> 72
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 391
uaaguaaggc ggaaccaggc ugucaccccg ugucaaaggu caggggucag ccuuccgcuu 60

uacacaaaau gg 72

<210> 392
<211> 72
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 392
uaaguaaggc ggaaccaggc ugucaccccg ugucaaaggu caggggucag ccuuccgcuu 60

uacacaaaau gg 72

<210> 393
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 393
gcagccauuu uaagucagcu ucggggaggg ucacgcaaag uucaaagguc auccucaccg 60

gaacugguac aaaauggccg 80

<210> 394
<211> 74
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 394
ugcuacguca uaagugacgu agcugguguc ugcuguaaac acggaaguag gccccgccac 60

gucacuuguc acgu 74

<210> 395
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 395

aguagcugac gucaaggauu gac 23

<210> 396
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 396
caagucacgu ggaggggacc cg 22

<210> 397
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 397
aaguagcuga cgucaaggau ugacg 25

<210> 398
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 398
auaagucacg uggugggac ccg 23

<210> 399
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 399
uggggagggu uggcguauag cccgga 26

<210> 400
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 400
ccccccccgg gggggguuu gccc 24

<210> 401
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 401
aucagucacg uggggaagg cgugc 25

<210> 402
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 402
aaguaaggcg gaagcagcuc gg 22

<210> 403
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 403
agucacgugg ggaggguugg c 21

<210> 404
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 404
cccgaaggcg gguacccgag gu 22

<210> 405
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 405
uaucguagcc gacgcggacc ccg 23

<210> 406
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 406
auuuugugau acgcgcgucc ccuccc 26

<210> 407
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 407
aagucacgug acugggcagg u 21

<210> 408
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 408
ugacgucaaa gucacguggg ragggu 26

<210> 409
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 409
gaucgagcgu cccgugggcg ggu 23

<210> 410
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 410
ugacgucaaa gucacguggg gagggcgg 28

<210> 411
<211> 23
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 411
gaucgagcgu cccgugggcg ggu 23

<210> 412
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 412
gaucgagcgu cccgugggcg ggu 23

<210> 413
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 413
ugacgucaaa gucacguggg gagggcgg 28

<210> 414
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 414
acgugguaag ucacgugggg gcagcu 26

<210> 415

<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 415
auuuuauaau acgcgcgucc ccucc 25

<210> 416
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 416
aagggacagc cuuccggcuu gc 22

<210> 417
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 417
cauaagucac gugggacgg cugcu 25

<210> 418
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 418
uaaggcggaa caggcugaca cccc 24

<210> 419
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 419
uacgucauaa gucacguggg aagagcug 28

<210> 420
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 420
ucgggcgucc cgagggcggg ug 22

<210> 421
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 421
aaagugagug gggccagacu ucgcc 25

<210> 422
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 422

cucucacgug auggggcgu gc 22

<210> 423
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 423
ucuacuacgu acacuuccug gggugugu 28

<210> 424
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 424
uggcuguugu cacgugacuu ga 22

<210> 425
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 425
agacgccauu uuaggcccuc gcgg 24

<210> 426
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 426
ugucgugacg uuugagacac gugau 25

<210> 427
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 427
ugacguuuga gacacgugau gggggcgugc 30

<210> 428
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 428
agugaguggg gccagacuuc gc 22

<210> 429
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 429
ugugacguca aagucacgug gggagggcgg 30

<210> 430
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 430
aaaagaggaa guaugacgua gcggcgg 27

<210> 431
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 431
agcgagcggc cgagcgaccc g 21

<210> 432
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 432
uucagucacg uggggaggga acgc 24

<210> 433
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 433
aaaagaggaa gugugacgua gcgg 24

<210> 434
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 434
ugugacguca aagucacgug gggagggcgg 30

<210> 435
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 435
aaaagaggaa guaugacgug gcgg 24

<210> 436
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 436
ugacgucaaa gucacguggg gagggcgg 28

<210> 437
<211> 29
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 437
aaaaaagagg aagugugacg uagcggcgg 29

<210> 438
<211> 23
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 438
gaucgagcgu cccgugggcg ggu 23

<210> 439
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 439
ugugacguca aagucacgug gggaggaggg 30

<210> 440
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 440
uugguccuga cgucacuguc a 21

<210> 441
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 441
cgucacuguc acguggggag ggaacac 27

<210> 442

<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 442
uaaguaagca uggcgggcgg ugau 24

<210> 443
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 443
aaguaagcau ggcgggcggu ga 22

<210> 444
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 444
uaaguaguug aggcggacgg uggc 24

<210> 445
<211> 29
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 445
ucaugcuugc acaaaauggc ggacuuccg 29

<210> 446
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 446
aguagcugac gucaaggauu gac 23

<210> 447
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 447
aguagcugac gucaaggauu gac 23

<210> 448
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 448
acaagucacg ugggggggac ccg 23

<210> 449
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 449

aaguagcuga cgucaaggau ugacg 25

<210> 450
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 450
auaagucacg uggugggac ccg 23

<210> 451
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 451
aagucagcuc uggggaggcg ugacuu 26

<210> 452
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 452
aguagcugac gucaaggauu gac 23

<210> 453
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 453
caagucacgu ggaggggaca cg 22

<210> 454
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 454
uguuaagugg cugucgccga ggauuga 27

<210> 455
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 455
uaagucacau gacagggguc ca 22

<210> 456
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 456
cggagaggga gccacggagg ug 22

<210> 457
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 457
aagucagcgc uggggaggca uga 23

<210> 458
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 458
ucuuaagugg cugucgccga ggauugac 28

<210> 459
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 459
aagucacaug ucuaggguc cacu 24

<210> 460
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 460
aaguaaggcg gaagcagcug ucc 23

<210> 461
<211> 29
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 461
aucuuaagug gcugucgcug aggauugac 29

<210> 462
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 462
uaagucacau gacaggaguc cacu 24

<210> 463
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 463
aaguggcugu cgccgaggau ug 22

<210> 464
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 464
uaagucacau gacaggggguc ca 22

<210> 465
<211> 23
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 465
acaagucacg ugggggggac ccg 23

<210> 466
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 466
auuuuaggcu guugccgggc guuugacu 28

<210> 467
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 467
auaagucacg ugacagggag ggg 23

<210> 468
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 468
aaguagguga cguccaggac u 21

<210> 469

<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 469
cauaagucac gugggacgg cugu 24

<210> 470
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 470
uaaguaaggc ggaagagcuc uagcua 26

<210> 471
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 471
aguagcugac gucaaggauu gac 23

<210> 472
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 472
acaagucacg ugggggggac ccg 23

<210> 473
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 473
uaaguaagca uggcgggcgg cgac 24

<210> 474
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 474
aucgagacac guggugccag cagcu 25

<210> 475
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 475
ucauccucag ccggaacuua cacaaaaugg 30

<210> 476
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 476

auaugucacg uggggaggcc cugcug 26

<210> 477
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 477
aaguaaggcg gaagcagcuc cacuuu 26

<210> 478
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 478
aaguaaggcg gaaguuucuc cacu 24

<210> 479
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 479
uaaguaguug aggcggacgg uggc 24

<210> 480
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 480
uaaguaguug aggcggacgg ugg 23

<210> 481
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 481
gaccuucggc gucgggggg ucggggg 27

<210> 482
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 482
auccucggcg gaaccuaua 19

<210> 483
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 483
aucgaugacg uggggaggcg uacuau 26

<210> 484
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 484
uggcggagga gcacuuccgg cuug 24

<210> 485
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 485
aacaagucac guggggaggg uuggc 25

<210> 486
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 486
aggggucagc cuuccgcuuu a 21

<210> 487
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 487
aggggucagc cuuccgcuuu a 21

<210> 488
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 488
uaagucagcu ucggggaggg ucac 24

<210> 489
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 489
ucauaaguga cguagcuggu gucugcu 27

<210> 490
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 490
cauccucggc ggaagcuaca caa 23

<210> 491
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 491
ggccccguca cgugacuuac cac 23

<210> 492
<211> 24
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 492
ucauccucgg cggaagcuac acaa 24

<210> 493
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 493
ggccccguca cgugauuugu cac 23

<210> 494
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 494
ccgggucaua ggucacaccu acgucac 27

<210> 495
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 495
ggcugccgcc ccccccgggg aaaggggg 28

<210> 496

<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 496
auccucgucc acgugacugu ga 22

<210> 497
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 497
gagcacuucc ggcuugccca a 21

<210> 498
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 498
caauccucuu acguggccug 20

<210> 499
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 499
cgagguuaag ggccaauucg ggcu 24

<210> 500
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 500
gggcccccgc ggggcucucg gcg 23

<210> 501
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 501
gcggggcgug gccguaucag aaaaugg 27

<210> 502
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 502
ccucggucac guggccugu 19

<210> 503
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 503

ccucgucacg ugaccugacg ucacag 26

<210> 504
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 504
ccguccgcgg cgagagcgcg agcga 25

<210> 505
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 505
auccucguca cgugaccuga cgucacg 27

<210> 506
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 506
ccguccgcgg cgagagcgcg agcga 25

<210> 507
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 507
ccguccgcgg cgagagcgcg agcga 25

<210> 508
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 508
auccucguca cgugaccuga cgucacg 27

<210> 509
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 509
cugacccgcg ugacugguca cguga 25

<210> 510
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 510
cggggcgugg ccguauuaga aaaugg 26

<210> 511
<211> 29
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 511
aguaaggcgg aaccaggcug ucacccugu 29

<210> 512
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 512
uagcugaccc gcgugacuug ucac 24

<210> 513
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 513
ggucagccuc cgcuuugca 19

<210> 514
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 514
gcugacccgc guggcuuguc acgugagu 28

<210> 515
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 515
ggcccguccg cggcgagagc gcgag 25

<210> 516
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 516
uccgggugcg ucuguggggcc gccauuu 27

<210> 517
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 517
auccucgacc acgugacugu g 21

<210> 518
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 518
auaaaccaga ggggugacga augguagagu 30

<210> 519
<211> 23
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 519
caaagucacg uggugacggc cau 23

<210> 520
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 520
guaggcgcgu uuuaaugacg ucacgg 26

<210> 521
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 521
uaggcgcguu uuaaugacgu cacggcag 28

<210> 522
<211> 29
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 522
aucccugguc acgugacucu gacgucacg 29

<210> 523

<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 523
gcgugugggg gccgccauuu uagcuu 26

<210> 524
<211> 29
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 524
ucauccucgu cacgugaccu gacgucacg 29

<210> 525
<211> 29
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 525
cgccaucuug ugacuuccuu ccgcuuuuu 29

<210> 526
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 526
uagaguuuuu uccgcccguc cg 22

<210> 527
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 527
gucccugguc acgugauugu gac 23

<210> 528
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 528
cauuuuguga cuuccuuccg cuuuuu 26

<210> 529
<211> 29
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 529
ucauccucgu cacgugaccu gacgucacg 29

<210> 530
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 530

ccgccaucuu gugacuuccu uccgcuuuuu 30

<210> 531
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 531
auccucguca cgugaccuga cgucacg 27

<210> 532
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 532
cgccaucuug ugacuuccuu ccgcuuuuuc 30

<210> 533
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 533
ccguccgcgg cgagagcgcg agcga 25

<210> 534
<211> 29
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 534
ucauccucgu cacgugaccu gacgucacg 29

<210> 535
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 535
cuuagugacg ucacggcagc cau 23

<210> 536
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 536
gucccugguc acgugacaug acguc 25

<210> 537
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 537
cacagccggu caugcuugca caaa 24

<210> 538
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 538
acagccuguc augcuugcac aa 22

<210> 539
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 539
caccaucagc cacaccuacu caaa 24

<210> 540
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 540
cgggucgccg ccauauuugg ucacguga 28

<210> 541
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 541
cauccucggc ggaagcuaca caa 23

<210> 542
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 542
cauccucggc ggaagcuaca caa 23

<210> 543
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 543
ggccccguca cgugacuuac cac 23

<210> 544
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 544
ucauccucgg cggaagcuac acaa 24

<210> 545
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 545
ggccccguca cgugauuugu cac 23

<210> 546
<211> 25
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 546
gucauccuca ccauaacugg cacaa 25

<210> 547
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 547
cauccucggc ggaagcuaca caa 23

<210> 548
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 548
ggccccguca cgugacuuac cac 23

<210> 549
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 549
auccucgacg guaaccgcaa acaug 25

<210> 550

<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 550
ggccccgaca ugugacucgu c 21

<210> 551
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 551
agcacuuccg aauggcugag uuuucca 27

<210> 552
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 552
auccucaccg gaacugacac aa 22

<210> 553
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 553
cauccucggc gguaaccgca aagaug 26

<210> 554
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 554
uaggccccga caugugacuc gu 22

<210> 555
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 555
acagccuucc gcuuugcaca a 21

<210> 556
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 556
cauccucggc gguaaccgca aagaugg 27

<210> 557
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 557

uaggccccga caugugacuc guc 23

<210> 558
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 558
uccucggcgg uaaccgcaaa 20

<210> 559
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 559
ggccccgaca ugugacucgu c 21

<210> 560
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 560
ggccccguca cgugacuuac cac 23

<210> 561
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 561
ucaaacaccc agcgacacca aaaaaugg 28

<210> 562
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 562
ccucgcccac gugacuuacc ac 22

<210> 563
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 563
ccucggcgga accuauacaa 20

<210> 564
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 564
gccccgucac gugauuuacc ac 22

<210> 565
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 565
gcggcggagc acuuccgcuu ugcccaaa 28

<210> 566
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 566
cauccucggc ggaagcuaca caa 23

<210> 567
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 567
ggccccguca cgugacuuac cac 23

<210> 568
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 568
caccgcacuu ccgugcuugc acaaa 25

<210> 569
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 569
ucgcugacac acgugucuug ucac 24

<210> 570
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 570
cauuuuaagu aagcaccgcc uagggaugac 30

<210> 571
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 571
guaggccccg ucacguguca uaccac 26

<210> 572
<211> 25
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 572
ggcggggcac uuccggcuug cccaa 25

<210> 573
<211> 26
<212> RNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 573
cggcggagca cuuccggcuu gcccaa 26

<210> 574
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 574
caccaucagc cacaccuacu caaa 24

<210> 575
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 575
accaucagcc acaccuacuc aaa 23

<210> 576
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 576
gacuccgaga ugccauugga cacugagg 28

<210> 577

<211> 18
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 577
aguaggugcc guccagca 18

<210> 578
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 578
aaguaggccc cgcuacguca ucaucac 27

<210> 579
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 579
aaggcggaag agcugcucua uau 23

<210> 580
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 580
caauccuccc acguggccug ucac 24

<210> 581
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 581
aaggcggaac caggcuguca ccccgu 26

<210> 582
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 582
aaggcggaac caggcuguca ccccgu 26

<210> 583
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 583
cauccucacc ggaacuggua caaa 24

<210> 584
<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<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
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic primer"

<400> 690
attcgaatgg ctgagtttat gc 22

<210> 691

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 691
cacgaattag ccaagactgg gcac 24

<210> 692
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 692
gctcccactc ctgatttctg 20

<210> 693
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 693
ccttgactac ggtggtttca c 21

<210> 694
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 694
tgcaggcatt cgagggcttg tt 22

<210> 695
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 695
tttaaccccc tagtcccagg 20

<210> 696
<211> 56
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 696
tttgtgacac aagatggccg acttccttcc tctttagtct tccccaaaga agacaa 56

<210> 697
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 697
gaagcccacc aaaagcaatt 20

<210> 698
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 698

agttcccgtg tctatagtcg a 21

<210> 699
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic probe"

<400> 699
acttcgttac agagtccagg gg 22

<210> 700
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic primer"

<400> 700
agcaacaggt aatggaggac 20

<210> 701
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic primer"

<400> 701
tggaagctgg ggtctttaac 20

<210> 702
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic probe"

<400> 702
tctaccttag gtgcaaaggg cc 22

<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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="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>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 804
cggcccagcg gcggcgcgcg cgcttcgcgc gcgcgccggg gggctccgcc ccccccgcg 60

catgcgcggg gccccccccc gcgggggggct ccgccccccg gtcccccccc g 111

<210> 805
<211> 115
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 805
cggccgtgcg gcggcgcgcg cgcttcgcgc gcgcgccggg ggctgccgcc ccccccgcg 60

catgcgcgcg gggcccccccc ccgcgggggg ctccgccccc cggcccccccc ccccg 115

<210> 806
<211> 104
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="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>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 807
ggcggcggcg cgcgcgctac gcgcgcgcgc cggggagctc tgccccccgc cgcgcatgcg 60

cgcgggtccc ccccccgcgg ggggctccgc cccccggtcc cccccccg 108

<210> 808
<211> 37
<212> PRT
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Arginine-rich region"

<400> 808
Thr Val Val Arg Arg Arg Gly Arg Ser Pro Arg Arg Arg Thr Pro Ser
1 5 10 15

Pro Arg Arg Arg Arg Ser Gln Ser Pro Arg Arg Arg Arg Ser Gln Ser
20 25 30

Arg Glu Ser Gln Cys
35

<210> 809
<211> 27
<212> PRT
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Arginine-rich region"

<400> 809
Arg Arg Arg Tyr Ala Arg Pro Tyr Arg Arg Arg His Ile Arg Arg Tyr
1 5 10 15

Arg Arg Arg Arg Arg His Phe Arg Arg Arg Arg
20 25

<210> 810
<211> 71
<212> DNA

<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 810
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggcymtgg g 71

<210> 811
<211> 71
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Alphatorquevirus clade sequence "

<400> 811
cgggtgccgt aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 812
<211> 70
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Alphatorquevirus clade sequence "

<400> 812
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggcccggg 70

<210> 813
<211> 71
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Alphatorquevirus clade sequence "

<400> 813
cgggtgccgg aggtgagttt acacaccgaa gtcaaggggc aattcgggct caggactggc 60

cgggctttgg g 71

<210> 814
<211> 69
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Alphatorquevirus clade sequence "

<400> 814
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggaggccg 60

ggccatggg 69

<210> 815
<211> 71
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Alphatorquevirus clade sequence "

<400> 815
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggccccgg g 71

<210> 816
<211> 71
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Alphatorquevirus clade sequence "

<400> 816
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 817
<211> 71
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Alphatorquevirus clade sequence "

<400> 817
cgggtgccga aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 818
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 818
gcgctkcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 819
<211> 36
<212> DNA
<213> Torque teno virus

<400> 819
gcgcttcgcg cgccgcccac taggggggcgt tgcgcg 36

<210> 820
<211> 36
<212> DNA
<213> Torque teno virus

<400> 820
gcgctgcgcg cgccgcccag taggggggcgc aatgcg 36

<210> 821
<211> 36
<212> DNA
<213> Torque teno virus

<400> 821
gcgctgcgcg cgcggccccc gggggaggca ttgcct 36

<210> 822
<211> 36
<212> DNA
<213> Torque teno virus

<400> 822
gcgctgcgcg cgcgcgccgg gggggcgcca gcgccc 36

<210> 823
<211> 36
<212> DNA
<213> Torque teno virus

<400> 823
gcgcttcgcg cgcgcgccgg ggggctccgc cccccc 36

<210> 824
<211> 36
<212> DNA
<213> Anelloviridae sp.

<400> 824
gcgcttcgcg cgcgcgccgg ggggctccgc cccccc 36

<210> 825
<211> 36
<212> DNA
<213> Torque teno virus

<400> 825
gcgcttcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 826
<211> 36
<212> DNA
<213> Torque teno virus

<400> 826
gcgctacgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 827
<211> 36
<212> DNA
<213> Torque teno virus

<400> 827
gcgctacgcg cgcgcgccgg ggggctctgc cccccc 36

<210> 828
<211> 724
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<220>
<221> MOD_RES
<222> (36)..(36)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (170)..(170)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (246)..(246)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (276)..(276)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (360)..(360)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (370)..(370)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (434)..(434)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (682)..(682)
<223> Any amino acid

<400> 828
Met Ala Trp Gly Trp Trp Arg Arg Arg Arg Met Ala Trp Arg Arg Trp
1   5   10   15

Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg
            20   25   30

Arg Arg Pro Xaa Arg Arg Arg Arg Arg Arg Arg Arg Val Arg Arg Arg
        35   40   45

Arg Arg Gly Arg Trp Arg Arg Arg Arg Tyr Arg Arg Arg Trp Arg Arg
    50   55   60

Arg Arg Arg Arg Arg Arg Arg Lys Lys Leu Val Leu Thr Gln Trp Gln
65   70   75   80

Pro Asn Thr Val Arg Arg Cys Tyr Ile Arg Arg Gly Tyr Leu Pro Leu
                85   90   95

Ile Ile Cys Gly Glu Asn Thr Thr Ser Arg Asn Tyr Ala Thr His Ser
            100   105   110

Asp Asp Thr Pro Gln Gly Pro Phe Gly Gly Gly Met Ser Thr Thr Thr
        115   120   125

Phe Ser Leu Arg Val Leu Tyr Asp Glu Tyr Gln Arg Phe Met Asn Arg
    130   135   140

Trp Thr Tyr Ser Asn Glu Asp Leu Asp Leu Ala Arg Tyr Leu Gly Cys
145   150   155   160

Lys Phe Thr Phe Tyr Arg His Pro Asp Xaa Asp Phe Ile Val Gln Tyr
                165   170   175

Asn Thr Asn Pro Pro Phe Lys Asp Thr Lys Leu Thr Ala Pro Ser Ile
            180   185   190

His Pro Gly Met Leu Met Leu Ser Lys Arg Lys Ile Leu Ile Pro Ser
        195   200   205

```
Leu Lys Thr Arg Pro Lys Gly Lys His Tyr Val Lys Val Arg Ile Gly
    210                 215                 220

Pro Pro Lys Leu Phe Glu Asp Lys Trp Tyr Thr Gln Ser Asp Leu Cys
225                 230                 235                 240

Asp Val Pro Leu Val Xaa Leu Tyr Ala Thr Ala Ala Asp Leu Gln His
                245                 250                 255

Pro Phe Gly Ser Pro Gln Thr Asp Asn Pro Cys Val Thr Phe Gln Val
            260                 265                 270

Leu Gly Ser Xaa Tyr Asn Lys His Leu Ser Ile Ser Pro Thr Asn Asp
        275                 280                 285

Asn Leu Lys His Leu Leu Asn Asn Leu Tyr Thr Ser Gly Tyr Asn Thr
    290                 295                 300

Phe Gln Thr Ile Ala Gln Leu Lys Pro Thr Ile Asp Ala Lys Asn Gly
305                 310                 315                 320

Thr Thr Asn Thr Thr Asn Thr Thr Thr Thr Gly Thr Thr Phe Thr Thr
                325                 330                 335

Thr Lys Asn Asp Ser Trp Tyr Gly Gly Asn Thr Tyr Asn Pro Asn Ile
            340                 345                 350

Lys Lys Leu Lys Lys Ile Arg Xaa Tyr Phe Lys Lys Leu Thr Lys Lys
        355                 360                 365

Leu Xaa Pro Thr Tyr Thr Thr Pro Thr Asp Tyr Leu Glu Tyr His Leu
    370                 375                 380

Gly Ile Tyr Ser Pro Ile Phe Ile Ser Pro Gly Arg Ser Asn Phe Glu
385                 390                 395                 400

Phe Pro Gly Ala Tyr Thr Asp Ile Thr Tyr Asn Pro Leu Thr Asp Lys
                405                 410                 415

Gly Val Gly Asn Met Val Trp Ile Gln Tyr Leu Thr Lys Pro Asp Thr
            420                 425                 430
```

```
Ile Xaa Asp Lys Thr Gln Ser Lys Cys Leu Ile Glu Asp Leu Pro Leu
        435                 440                 445

Met Ala Ala Leu Tyr Gly Tyr Val Asp Phe Cys Glu Lys Glu Thr Gly
    450                 455                 460

Asp Gln Asp Ile Glu Thr Asn Gly Arg Val Leu Ile Arg Cys Pro Tyr
465                 470                 475                 480

Thr Lys Pro Pro Leu Tyr Asp Lys Thr Asp Pro Asn Lys Gly Phe Val
                485                 490                 495

Pro Tyr Ser Thr Asn Ile Gly Asn Gly Lys Met Pro Gly Gly Ser Gly
            500                 505                 510

Tyr Val Pro Ile Tyr Trp Arg Ala Arg Met Tyr Pro Thr Leu Phe His
        515                 520                 525

Gln Lys Glu Val Leu Glu Asp Ile Val Gln Ser Gly Pro Phe Ala Tyr
    530                 535                 540

Lys Asp Asp Leu Lys Ser Ala Thr Leu Thr Ala Lys Tyr Lys Phe Lys
545                 550                 555                 560

Phe Lys Trp Gly Gly Asn Pro Ile Ser Gln Gln Val Val Arg Asn Pro
                565                 570                 575

Cys Lys Asp Ser Gly Ser Ser Pro Pro Gly Arg Gly Pro Arg Ser Val
            580                 585                 590

Gln Val Val Asp Pro Lys Tyr Met Gly Pro Glu Tyr Thr Phe His Ser
        595                 600                 605

Trp Asp Trp Arg Arg Gly Leu Phe Gly Glu Lys Ala Ile Lys Arg Met
    610                 615                 620

Ser Glu Gln Pro Thr Asp Asp Glu Leu Phe Pro Ala Gly Pro Lys Arg
625                 630                 635                 640
```

Pro Arg Arg Asp Pro Pro Glu Glu Gln Glu Glu Ser Ser Leu Phe Leu
645 650 655

Arg Arg Arg Pro Pro Trp Glu Glu Ser Ser Gln Glu Glu Ser Ser Ser
660 665 670

Glu Glu Glu Glu Gln Glu Glu Gln Thr Xaa Gln Gln Gln Leu Arg Gln
675 680 685

Gln Leu Arg Glu Gln Arg Gln Leu Arg Val Gln Leu Gln Leu Leu Phe
690 695 700

Gln Gln Leu Leu Lys Thr Gln Ala Gly Leu His Ile Asn Pro Leu Leu
705 710 715 720

Leu Ser Gln Ala

<210> 829
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="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
<211> 3648
<212> DNA
<213> Torque teno virus

<400> 830
cgtcactaac cacgtgactc ccacaggcca accacagtgt acgtgattca cttcctggga 60

gtggtttaca ttataatata agcaactgca cttccgaatg gctgagtttt ccacgcccgt 120

ccgcagcgag aacaccacgg aggggagtcc gcgcgtcccg tgggcgggtg ccgaaggtga 180

gtttacacac cgcagtcaag gggcaattcg ggcacgggac tggccgggct atgggcaagg 240

ctcttaaaaa gctatgtttc ttggtaggcc gtaccgaaag aaaaggaaac tgctactgct 300

accactgcat tctacaccga aaactagccg ggttatgagc tggtctaggc ctgtacataa 360

tgccacaggc attgaaagaa actggtggga gtcctgtctt agatcccacg caagttcttg 420

tggctgcggt aattttgtta atcatattaa tgtactggct aatcggtatg gctttgctgg 480

ttccacggag acgccgggta atcctcggcc gaggcccccg gtactgagct ccaccaccag 540

cactcctacc gatcaatcca gaccagctct accatggcat ggggatactg gtggagaagg 600

cgcttctgga gaccccgcag gagatggaga acgtggcgcc gcagaaggag actacggccc 660

agaagatcta gacgcacttt tcgacgcact cgacgaagag taaggaggcg acggtggggg 720

aggcgtgcac gcaggcgggg atggcgacgc aggacttata ttagagccag gcgacgcagg 780

agacgaaaaa gacttgtact gactcagtgg catcccgcag ttagaagaaa atgtaaaatt 840

acaggctaca tgcctatagt atactgtgga catggcagag ctagttttaa ctatgcctgg 900

cactctgatg actgtataaa acaaccacta ccctttggag gctcactatc tacagtgtcc 960

ttcaacctaa aagtactatt tgacgaaaac caaagaggac taaacaaatg gagctaccca 1020

aatgaccaac tagacctcgc cagatacaaa ggctgtagac taacatttta cagaaaaaaa 1080

aacacagact acatagctca atatgacata tcagaacctt atcaactaga caaatatagc 1140

tgtgcaaact atcacccctc aaaaatgatg tttgcaaaaa acaaaatttt aattcctagc 1200

tatgatacaa aacctagagg cagacaaaga gttagagtta gaatagggcc ccctaaacta 1260

tttacagaca agtggtacag tcaatcagac ttatgcaagg taaaccttgt gtcacttgcg 1320

gtttctgcgg cttcctttct ccacccattc ggctcaccac aaactgccaa cttttgtgca 1380

accttccagg tgctgcaacc gttctactac caggctatag gcattagttc tacaaaacac 1440

tcagaagtta tagacatttt atataagaaa aatacatact ggcaaagcaa cattacctct 1500

tggtttttaa ctaatgttaa aaacccaaaa aatatgtcca caaaaatgtt tgaggacatt 1560

aatgttaaat caaacaaaga cagtaattat gactggtttc catttacccc atacactaca 1620

gaaaactatt caaaaattca aaatgcagct caagaatact ggaaatattt aactagtgac 1680

cacccacaag ctactaatag caatgaaggc ctagtacaac catggactaa tgccactata 1740

aaacaatatg aataccacct cggtatgttt agtcctatat ttataggacc taccagagct 1800

aaaactaaat ttaaaacagc atactttgac tgcacttata acccactact agacaaagga 1860

atgggaaaca gaatatggta tcaatacgca accaaagctg acacacaaat atcaaaaaca 1920

gggtgctact gcatgttaga agacattcca atatatgcag cattttatgg atacgtagac 1980

tttatagaaa tggaaatagg taaaggacaa gacattaaag agaacggact tatttgctgc 2040

atatgtagat acacagaccc cccaatgtac aatgaacaac atccagacat gggatttgta 2100

ttttataaca ctaactttgg aaatggaaaa tggatagatg gacggggcga catacctact 2160

tactggatgc aaagatggag acctgttgta ttatttcaaa ctgatgttat tagagactta 2220

gtagaaactg gacctttag ttacaaagat gacctagcaa atacctcact gactatgaaa 2280

tatgaattct attttacctg gggcggaaac caggcgtacc accagacaat caaaaaccct 2340

tgtaaagacg aaggtaccgg accccataga cagcctagag acgtacaagt tacggacccg 2400

acaaccgtgg gacctgaata tgtgttccac gcgtgggact ggagacgggg cttccttagc 2460

gagcgagctc tcagacgcat gttcgaaaaa cctctcaact atgatgagta ttctaaaaaa 2520

ccaaaaagac ctagaatatt tcctccaaca gaaacagagt cccgaaacca agagctcgaa 2580

gaaagctcgc tttcagagga agaaaagtcg ctactctcca cagaagagat ccagaaagag 2640

gagatacagc gacagttcaa gcgacagctc aagcgacagc tgcgcctcgg gcagcagctc 2700

aaactcctcc aacaacaact cctcaagacg caagcgggcc tgcacctaaa ccccctttca 2760

tatttcccgc aataaataaa gtgtacctgt tcccagacag agctccaaaa cctaaaccca 2820

cctctggaga ctgggaaaca gagtatgcag cttgcagtgc ctttgacaga cccgctagaa 2880

ccaaccttag ctcacccccc tactacccag gagtacctac tccctggcaa gtaaaattca 2940

gccttaaatt tcaataaagt gcatttttac tacagctggg ccgtgggagt ttcacttgtc 3000

ggtgtctacc tcttaaggtc actaagcact ccgagcgcag cgaggagtgc gacccttaac 3060

cctgggtcaa cgccttcgga gccgcgcgct acgccttcgg ctgcgcgcgg cacctcagac 3120

ccccgctcgt gctgacgcgc ttgcgcgcgt cagaccactt cgggctcgcg ggggtcggga 3180

actttgctaa cagactccga ggtgccattg gacacagagt gggcgttcag caacgaaagt 3240

gagtggggcc agacttcgcc ataaggcctt tatcttcttg ccatttgtca gtataagggg 3300

ttgccatagg cttcggcctc aattttaggc cttccggact accaaaatgg ccgatttagt 3360

gacgtcacgg cggccatttt aagtaaggcg gaagtaactc cactatttac aaaatggcgg 3420

cggagcactt ccggcttgcc caaaatggcg gcaaaaaaca tccgggtcaa aggtcgttac 3480

cacgtcacaa gtcacgtggg agggtggtgc tgtaaacccg gaagcaatcc tctcacgtgg 3540

ctagtcacgt gactaacacg tcacacccgc cattttgttt tacaaaatgg ccgacttcct 3600

tccgcttttt taaaaataac ggctcagcgg cggcgcgcgc gctacgcg 3648

<210> 831
<211> 122
<212> PRT
<213> Torque teno virus

<400> 831
Met Ser Trp Ser Arg Pro Val His Asn Ala Thr Gly Ile Glu Arg Asn
1 5 10 15

Trp Trp Glu Ser Cys Leu Arg Ser His Ala Ser Ser Cys Gly Cys Gly
20 25 30

Asn Phe Val Asn His Ile Asn Val Leu Ala Asn Arg Tyr Gly Phe Ala
35 40 45

Gly Ser Thr Glu Thr Pro Gly Asn Pro Arg Pro Arg Pro Pro Val Leu
50 55 60

Ser Ser Thr Thr Ser Thr Pro Thr Asp Gln Ser Arg Pro Ala Leu Pro
65 70 75 80

Trp His Gly Asp Thr Gly Gly Glu Gly Ala Ser Gly Asp Pro Ala Gly
85 90 95

Asp Gly Glu Arg Gly Ala Ala Glu Gly Asp Tyr Gly Pro Glu Asp Leu
100 105 110

Asp Ala Leu Phe Asp Ala Leu Asp Glu Glu
        115   120

<210> 832
<211> 262
<212> PRT
<213> Torque teno virus

<400> 832
Met Ser Trp Ser Arg Pro Val His Asn Ala Thr Gly Ile Glu Arg Asn
1   5   10   15

Trp Trp Glu Ser Cys Leu Arg Ser His Ala Ser Ser Cys Gly Cys Gly
            20   25   30

Asn Phe Val Asn His Ile Asn Val Leu Ala Asn Arg Tyr Gly Phe Ala
        35   40   45

Gly Ser Thr Glu Thr Pro Gly Asn Pro Arg Pro Arg Pro Pro Val Leu
    50   55   60

Ser Ser Thr Thr Ser Thr Pro Thr Asp Gln Ser Arg Pro Ala Leu Pro
65   70   75   80

Trp His Gly Asp Thr Gly Gly Glu Gly Ala Ser Gly Asp Pro Ala Gly
                85   90   95

Asp Gly Glu Arg Gly Ala Ala Glu Gly Asp Tyr Gly Pro Glu Asp Leu
            100   105   110

Asp Ala Leu Phe Asp Ala Leu Asp Glu Glu Gln Ser Lys Thr Leu Val
        115   120   125

Lys Thr Lys Val Pro Asp Pro Ile Asp Ser Leu Glu Thr Tyr Lys Leu
    130   135   140

Arg Thr Arg Gln Pro Trp Asp Leu Asn Met Cys Ser Thr Arg Gly Thr
145   150   155   160

Gly Asp Gly Ala Ser Leu Ala Ser Glu Leu Ser Asp Ala Cys Ser Lys
                165   170   175

```
Asn Leu Ser Thr Met Met Ser Ile Leu Lys Asn Gln Lys Asp Leu Glu
            180                 185                 190

Tyr Phe Leu Gln Gln Lys Gln Ser Pro Glu Thr Lys Ser Ser Lys Lys
        195                 200                 205

Ala Arg Phe Gln Arg Lys Lys Ser Arg Tyr Ser Pro Gln Lys Arg Ser
    210                 215                 220

Arg Lys Arg Arg Tyr Ser Asp Ser Ser Ser Asp Ser Ser Ser Asp Ser
225                 230                 235                 240

Cys Ala Ser Gly Ser Ser Ser Asn Ser Ser Asn Asn Asn Ser Ser Arg
                245                 250                 255

Arg Lys Arg Ala Cys Thr
            260
```

<210> 833
<211> 256
<212> PRT
<213> Torque teno virus

<400> 833

```
Met Ser Trp Ser Arg Pro Val His Asn Ala Thr Gly Ile Glu Arg Asn
1               5                   10                  15

Trp Trp Glu Ser Cys Leu Arg Ser His Ala Ser Ser Cys Gly Cys Gly
            20                  25                  30

Asn Phe Val Asn His Ile Asn Val Leu Ala Asn Arg Tyr Gly Phe Ala
        35                  40                  45

Gly Ser Thr Glu Thr Pro Gly Asn Pro Arg Pro Arg Pro Pro Val Leu
    50                  55                  60

Ser Ser Thr Thr Ser Thr Pro Thr Asp Gln Ser Arg Pro Ala Leu Pro
65                  70                  75                  80

Trp His Gly Asp Thr Gly Gly Glu Gly Ala Ser Gly Asp Pro Ala Gly
                85                  90                  95
```

```
Asp Gly Glu Arg Gly Ala Ala Glu Gly Asp Tyr Gly Pro Glu Asp Leu
            100                 105                 110

Asp Ala Leu Phe Asp Ala Leu Asp Glu Glu Asn Arg Val Pro Lys Pro
        115                 120                 125

Arg Ala Arg Arg Lys Leu Ala Phe Arg Gly Arg Lys Val Ala Thr Leu
    130                 135                 140

His Arg Arg Asp Pro Glu Arg Gly Asp Thr Ala Thr Val Gln Ala Thr
145                 150                 155                 160

Ala Gln Ala Thr Ala Ala Pro Arg Ala Ala Ala Gln Thr Pro Pro Thr
                165                 170                 175

Thr Thr Pro Gln Asp Ala Ser Gly Pro Ala Pro Lys Pro Pro Phe Ile
            180                 185                 190

Phe Pro Ala Ile Asn Lys Val Tyr Leu Phe Pro Asp Arg Ala Pro Lys
        195                 200                 205

Pro Lys Pro Thr Ser Gly Asp Trp Glu Thr Glu Tyr Ala Ala Cys Ser
    210                 215                 220

Ala Phe Asp Arg Pro Ala Arg Thr Asn Leu Ser Ser Pro Pro Tyr Tyr
225                 230                 235                 240

Pro Gly Val Pro Thr Pro Trp Gln Val Lys Phe Ser Leu Lys Phe Gln
                245                 250                 255
```

<210> 834
<211> 178
<212> PRT
<213> Torque teno virus

<400> 834

```
Met Ser Trp Ser Arg Pro Val His Asn Ala Thr Gly Ile Glu Arg Asn
1               5                   10                  15

Trp Trp Glu Ser Cys Leu Arg Ser His Ala Ser Ser Cys Gly Cys Gly
            20                  25                  30
```

```
Asn Phe Val Asn His Ile Asn Val Leu Ala Asn Arg Asn Arg Val Pro
        35                  40                  45

Lys Pro Arg Ala Arg Arg Lys Leu Ala Phe Arg Gly Arg Lys Val Ala
    50                  55                  60

Thr Leu His Arg Arg Asp Pro Glu Arg Gly Asp Thr Ala Thr Val Gln
65                  70                  75                  80

Ala Thr Ala Gln Ala Thr Ala Ala Pro Arg Ala Ala Ala Gln Thr Pro
                85                  90                  95

Pro Thr Thr Thr Pro Gln Asp Ala Ser Gly Pro Ala Pro Lys Pro Pro
            100                 105                 110

Phe Ile Phe Pro Ala Ile Asn Lys Val Tyr Leu Phe Pro Asp Arg Ala
        115                 120                 125

Pro Lys Pro Lys Pro Thr Ser Gly Asp Trp Glu Thr Glu Tyr Ala Ala
    130                 135                 140

Cys Ser Ala Phe Asp Arg Pro Ala Arg Thr Asn Leu Ser Ser Pro Pro
145                 150                 155                 160

Tyr Tyr Pro Gly Val Pro Thr Pro Trp Gln Val Lys Phe Ser Leu Lys
                165                 170                 175

Phe Gln
```

<210> 835
<211> 733
<212> PRT
<213> Torque teno virus

<400> 835

```
Met Ala Trp Gly Tyr Trp Trp Arg Arg Arg Phe Trp Arg Pro Arg Arg
1               5                   10                  15

Arg Trp Arg Thr Trp Arg Arg Arg Arg Arg Leu Arg Pro Arg Arg Ser
            20                  25                  30
```

```
Arg Arg Thr Phe Arg Arg Thr Arg Arg Arg Val Arg Arg Arg Arg Trp
        35                  40                  45

Gly Arg Arg Ala Arg Arg Arg Gly Trp Arg Arg Arg Thr Tyr Ile Arg
    50                  55                  60

Ala Arg Arg Arg Arg Arg Arg Lys Arg Leu Val Leu Thr Gln Trp His
65                  70                  75                  80

Pro Ala Val Arg Arg Lys Cys Lys Ile Thr Gly Tyr Met Pro Ile Val
                85                  90                  95

Tyr Cys Gly His Gly Arg Ala Ser Phe Asn Tyr Ala Trp His Ser Asp
            100                 105                 110

Asp Cys Ile Lys Gln Pro Leu Pro Phe Gly Gly Ser Leu Ser Thr Val
        115                 120                 125

Ser Phe Asn Leu Lys Val Leu Phe Asp Glu Asn Gln Arg Gly Leu Asn
    130                 135                 140

Lys Trp Ser Tyr Pro Asn Asp Gln Leu Asp Leu Ala Arg Tyr Lys Gly
145                 150                 155                 160

Cys Arg Leu Thr Phe Tyr Arg Lys Lys Asn Thr Asp Tyr Ile Ala Gln
                165                 170                 175

Tyr Asp Ile Ser Glu Pro Tyr Gln Leu Asp Lys Tyr Ser Cys Ala Asn
            180                 185                 190

Tyr His Pro Ser Lys Met Met Phe Ala Lys Asn Lys Ile Leu Ile Pro
        195                 200                 205

Ser Tyr Asp Thr Lys Pro Arg Gly Arg Gln Arg Val Arg Val Arg Ile
    210                 215                 220

Gly Pro Pro Lys Leu Phe Thr Asp Lys Trp Tyr Ser Gln Ser Asp Leu
225                 230                 235                 240
```

```
Cys Lys Val Asn Leu Val Ser Leu Ala Val Ser Ala Ala Ser Phe Leu
                245                 250                 255

His Pro Phe Gly Ser Pro Gln Thr Ala Asn Phe Cys Ala Thr Phe Gln
            260                 265                 270

Val Leu Gln Pro Phe Tyr Tyr Gln Ala Ile Gly Ile Ser Ser Thr Lys
        275                 280                 285

His Ser Glu Val Ile Asp Ile Leu Tyr Lys Lys Asn Thr Tyr Trp Gln
    290                 295                 300

Ser Asn Ile Thr Ser Trp Phe Leu Thr Asn Val Lys Asn Pro Lys Asn
305                 310                 315                 320

Met Ser Thr Lys Met Phe Glu Asp Ile Asn Val Lys Ser Asn Lys Asp
                325                 330                 335

Ser Asn Tyr Asp Trp Phe Pro Phe Thr Pro Tyr Thr Thr Glu Asn Tyr
            340                 345                 350

Ser Lys Ile Gln Asn Ala Ala Gln Glu Tyr Trp Lys Tyr Leu Thr Ser
        355                 360                 365

Asp His Pro Gln Ala Thr Asn Ser Asn Glu Gly Leu Val Gln Pro Trp
    370                 375                 380

Thr Asn Ala Thr Ile Lys Gln Tyr Glu Tyr His Leu Gly Met Phe Ser
385                 390                 395                 400

Pro Ile Phe Ile Gly Pro Thr Arg Ala Lys Thr Lys Phe Lys Thr Ala
                405                 410                 415

Tyr Phe Asp Cys Thr Tyr Asn Pro Leu Leu Asp Lys Gly Met Gly Asn
            420                 425                 430

Arg Ile Trp Tyr Gln Tyr Ala Thr Lys Ala Asp Thr Gln Ile Ser Lys
        435                 440                 445

Thr Gly Cys Tyr Cys Met Leu Glu Asp Ile Pro Ile Tyr Ala Ala Phe
    450                 455                 460
```

```
Tyr Gly Tyr Val Asp Phe Ile Glu Met Glu Ile Gly Lys Gly Gln Asp
465                 470                 475                 480

Ile Lys Glu Asn Gly Leu Ile Cys Cys Ile Cys Arg Tyr Thr Asp Pro
                485                 490                 495

Pro Met Tyr Asn Glu Gln His Pro Asp Met Gly Phe Val Phe Tyr Asn
            500                 505                 510

Thr Asn Phe Gly Asn Gly Lys Trp Ile Asp Gly Arg Gly Asp Ile Pro
        515                 520                 525

Thr Tyr Trp Met Gln Arg Trp Arg Pro Val Val Leu Phe Gln Thr Asp
    530                 535                 540

Val Ile Arg Asp Leu Val Glu Thr Gly Pro Phe Ser Tyr Lys Asp Asp
545                 550                 555                 560

Leu Ala Asn Thr Ser Leu Thr Met Lys Tyr Glu Phe Tyr Phe Thr Trp
                565                 570                 575

Gly Gly Asn Gln Ala Tyr His Gln Thr Ile Lys Asn Pro Cys Lys Asp
            580                 585                 590

Glu Gly Thr Gly Pro His Arg Gln Pro Arg Asp Val Gln Val Thr Asp
        595                 600                 605

Pro Thr Thr Val Gly Pro Glu Tyr Val Phe His Ala Trp Asp Trp Arg
    610                 615                 620

Arg Gly Phe Leu Ser Glu Arg Ala Leu Arg Arg Met Phe Glu Lys Pro
625                 630                 635                 640

Leu Asn Tyr Asp Glu Tyr Ser Lys Lys Pro Lys Arg Pro Arg Ile Phe
                645                 650                 655

Pro Pro Thr Glu Thr Glu Ser Arg Asn Gln Glu Leu Glu Glu Ser Ser
            660                 665                 670
```

Leu Ser Glu Glu Glu Lys Ser Leu Leu Ser Thr Glu Glu Ile Gln Lys
675 680 685

Glu Glu Ile Gln Arg Gln Phe Lys Arg Gln Leu Lys Arg Gln Leu Arg
690 695 700

Leu Gly Gln Gln Leu Lys Leu Leu Gln Gln Gln Leu Leu Lys Thr Gln
705 710 715 720

Ala Gly Leu His Leu Asn Pro Leu Ser Tyr Phe Pro Gln
725 730

<210> 836
<211> 191
<212> PRT
<213> Torque teno virus

<400> 836
Met Ala Trp Gly Tyr Trp Trp Arg Arg Arg Phe Trp Arg Pro Arg Arg
1 5 10 15

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

Arg Arg Thr Phe Arg Arg Thr Arg Arg Arg Thr Ile Lys Asn Pro Cys
35 40 45

Lys Asp Glu Gly Thr Gly Pro His Arg Gln Pro Arg Asp Val Gln Val
50 55 60

Thr Asp Pro Thr Thr Val Gly Pro Glu Tyr Val Phe His Ala Trp Asp
65 70 75 80

Trp Arg Arg Gly Phe Leu Ser Glu Arg Ala Leu Arg Arg Met Phe Glu
85 90 95

Lys Pro Leu Asn Tyr Asp Glu Tyr Ser Lys Lys Pro Lys Arg Pro Arg
100 105 110

Ile Phe Pro Pro Thr Glu Thr Glu Ser Arg Asn Gln Glu Leu Glu Glu
115 120 125

Ser Ser Leu Ser Glu Glu Glu Lys Ser Leu Leu Ser Thr Glu Glu Ile
130 135 140

Gln Lys Glu Glu Ile Gln Arg Gln Phe Lys Arg Gln Leu Lys Arg Gln
145 150 155 160

Leu Arg Leu Gly Gln Gln Leu Lys Leu Leu Gln Gln Gln Leu Leu Lys
165 170 175

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

<210> 837
<211> 107
<212> PRT
<213> Torque teno virus

<400> 837
Met Ala Trp Gly Tyr Trp Trp Arg Arg Arg Phe Trp Arg Pro Arg Arg
1 5 10 15

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

Arg Arg Thr Phe Arg Arg Thr Arg Arg Arg Lys Gln Ser Pro Glu Thr
35 40 45

Lys Ser Ser Lys Lys Ala Arg Phe Gln Arg Lys Lys Ser Arg Tyr Ser
50 55 60

Pro Gln Lys Arg Ser Arg Lys Arg Arg Tyr Ser Asp Ser Ser Ser Asp
65 70 75 80

Ser Ser Ser Asp Ser Cys Ala Ser Gly Ser Ser Ser Asn Ser Ser Asn
85 90 95

Asn Asn Ser Ser Arg Arg Lys Arg Ala Cys Thr
100 105

<210> 838
<211> 3704
<212> DNA
<213> Torque teno virus

<400> 838
ccccgaagtc cgtcactaac cacgtgactc ccacaggcca atcagatgct atgtcgtgca 60

cttcctgggc tgtgtctacg tcctcatata agtaactgca cttccgaatg gctgagtttt 120

ccacgcccgt ccgcagcggc agcaccacgg agggtgatcc ccgcgtcccg tgggcgggtg 180

ccggaggtga gtttacacac cgcagtcaag gggcaattcg ggcacgggac tggccgggct 240

atgggcaagg ctcttaaaaa gctatgttct tcggtaggtg ctggagaaag aaaaggaaag 300

tgcttctgca agatctgtca actccaccga aaaaacctgc tatgagtgtg tggcttcctc 360

ccatagacaa tgttaccgag cgtgagagga gctggctctc tagcattctt cagtctcaca 420

gagctttttg tgggtgccat gatgctatct atcatcttag cagtctggct gctcgcttta 480

atatgcaacc agggccgtcg ccgggtggtg attctaggcc gccgcgaccg ccactaagac 540

gcctgcccgc gctcccgggt cccagagacc cccctagcga caccaacaac cgcaggtcat 600

ggcctactgg ggatggtgga gacggaggcg ctggccaagg cgcaggtgga ggcgctaccg 660

ctaccgaaga agactaccgc gccgaagacc tagacgagct gtacgccgcc ctcgaaggag 720

acgagtaagg aggcgccgcg gtagggggtg gtacagaggg cgacgctact cccgcagacg 780

gtacagacgt agatatgtga ggcgaaagag aaagactcta gtttggagac agtggcagcc 840

tcaaaatatc agaaaatgca ggatcagggg cataattccc atcctgatat gcggacacgg 900

gaggggggcc agaaactatg cgctccacag cgacgacata acccccaga acacccccctt 960

cgggggagga ctgagcacca cctcctggag cctaaaagtg ctatatgacc agcacaccag 1020

gggactcaac aggtggtctg ccagtaacga gagcctagac cttgccagat acaatggctg 1080

tagtttcact ttctacagag acaaaaagac tgactttata gtgacctatg acacctctgc 1140

tccctacaaa ctagacaaat acagctcccc cagctaccac ccagggtcca tgatgctcat 1200

gacaaaacac aaaatcctga tccccagttt tgacacaaaa cccaaaggtc ctgccaaaat 1260

tagagtcaga atcaagcccc ccaaaatgtt cttagataaa tggtacactc aagacgacct 1320

ctgttccgtt aatcttgtgt cacttgcggt tagcgcagct tcctttacac atccgttctg 1380

cccaccacta actgacactc cttgtgtaac gctgcaggtg ttgaaagact tctactacac 1440

aaccataggc tactcctcta atgcagacaa agtagagtct gtattcacta acactctcta 1500

caaacactgc tgctactatc agtcctttct caccactcaa tttatagcca aaatcactcg 1560

cacaccagat ggacaaccag tagccacatt ctctcctcct acctctttcc ctggcacaac 1620

tgtaacaaaa agttccatag aatcatttaa ccaatgggta acttccacag gtacaagtgg 1680

ctggctaaca aatgcaaacc aacactttca tttctgtaac tataaaccag atgccacaaa 1740

gctaaaatgg ctcagacagt actactttga ctgggaaaca tacaaattag cagatgtaaa 1800

gccagacggc cttacaccct cagtaaactg gtatgagtac agaataggcc tctttagtcc 1860

tattttcctg agcccttca gatctagcag tctagacttt cccagagcct accaggatgt 1920

gaactacaac cccctggtag acaaaggagt gggcaacatc atatggttcc aatacaacac 1980

aaaaccagac acacagctgt cagtacccag ctgcaagtgt gtcatagaag acaaacccct 2040

atgggcagcc ttctatggct acagtgactt tgtacaacaa gagataggag actacacaga 2100

cgcagaggcc gtgggcttcg tctgtgtcat ctgtccatac accaaacccc ctctaaaaaa 2160

cccagacaac cccatgcaag ggttcatatt ctatgacagc ctttttggca atggcaagtg 2220

gatagatggc acggggcacg tcccccttta ctggcagagc aggtggaggc cagagatgct 2280

cttccaagaa aacaccatga gagacatcac actatctggg cccttcagct acaaggacga 2340

ctataagaac tgtgtactga cttgcaaata caaatttaac tttcgattcg ggggcaatct 2400

tctccacgaa cagacgatca gaaacccatg ccccacggac ggacatccca gtaccggtag 2460

acagcctaga gacgtacaag tggttgaccc gatcaaagtg ggcccccggt tcgtgttcca 2520

ctcctgggac tggcgcagag gctaccttag cccagcagct ctcaaaagaa ttggagagca 2580

accgctcgat tatgaagctt attcgtaccg cccaaagaga cctagaatct ttcctcccac 2640

agaaggagac cagctcgccc gaagtcgaga agaagactca ttttcagagg aagaaagtcc 2700

ccatatctcg ttcgaagagg ggcaggaacc gaaagcccag gcggtacagc agcacctcct 2760

ccgacacctc agaaagcagc gagaactccg aaagcgactc cgagccctgt tccaaagcct 2820

ccaaaagacg caggcgggtc tccacgtaaa tccattatta ttcaaccagc ctgcaatcag 2880

gttctgatgt tcccagagat ggggcctaag ccagctccca ctgcccaaga ctggcagtgc 2940

gaatacgaga catgtaagca ctgggataga ccccccagaa agtttctcac agacccccct 3000

ttctatccct gggcccctac tacttacaat gtatctttca agctaaactt caaataaact 3060

aggccgtggg agtctcactt gtcggtgtct acctcttaag gtcactaagc actccgagcg 3120

tcagcgagga gtgcgaccct tccccctggt gcaacgccct cggcggccgc gcgctacgcc 3180

ttcggctgcg cgcggcacct cggacccccg ctcgtgctga cgcgctcgcg cgcgtcagac 3240

cacttcgggc tcgcgggggt cgggaaattt gctaaacaga ctccgagttg ccattggaca 3300

caggagctgt gaatcagtaa cgaaagtgag tggggccaga cttcgccata aggcctttat 3360

cttcttgcca tttgtccgtg aggaggggtc gccaagacgc ggaccccgtt ttcggacctt 3420

ccgaactacc aaaatggccg attcagtgac gtcacggcag ccattttgtg taagcaccgc 3480

ccaggacaga cgtcacagtt caaaggtcat cctcgagcgg aacttacaga aaatggcggt 3540

caattgcttc cgggtcaaag gtcacgccta cgtcataagt cacgtggtgg aggctactgc 3600

gcatacacgg aagtaggccc cgccacgtga ccgaccacgt gggtgctgcg tcacggccgc 3660

cattttgtat cacaaaatgg ccgacttcct tcctcttttt caaa 3704

<210> 839
<211> 128
<212> PRT
<213> Torque teno virus

<400> 839
Met Ser Val Trp Leu Pro Pro Ile Asp Asn Val Thr Glu Arg Glu Arg
1 5 10 15

Ser Trp Leu Ser Ser Ile Leu Gln Ser His Arg Ala Phe Cys Gly Cys
20 25 30

His Asp Ala Ile Tyr His Leu Ser Ser Leu Ala Ala Arg Phe Asn Met
35 40 45

Gln Pro Gly Pro Ser Pro Gly Gly Asp Ser Arg Pro Pro Arg Pro Pro
50 55 60

Leu Arg Arg Leu Pro Ala Leu Pro Gly Pro Arg Asp Pro Pro Ser Asp
65 70 75 80

Thr Asn Asn Arg Arg Ser Trp Pro Thr Gly Asp Gly Gly Asp Gly Gly
85 90 95

Ala Gly Gln Gly Ala Gly Gly Gly Ala Thr Ala Thr Glu Glu Asp Tyr
100 105 110

Arg Ala Glu Asp Leu Asp Glu Leu Tyr Ala Ala Leu Glu Gly Asp Glu

115 120 125

<210> 840
<211> 272
<212> PRT
<213> Torque teno virus

<400> 840
Met Ser Val Trp Leu Pro Pro Ile Asp Asn Val Thr Glu Arg Glu Arg
1 5 10 15

Ser Trp Leu Ser Ser Ile Leu Gln Ser His Arg Ala Phe Cys Gly Cys
20 25 30

His Asp Ala Ile Tyr His Leu Ser Ser Leu Ala Ala Arg Phe Asn Met
35 40 45

Gln Pro Gly Pro Ser Pro Gly Gly Asp Ser Arg Pro Pro Arg Pro Pro
50 55 60

Leu Arg Arg Leu Pro Ala Leu Pro Gly Pro Arg Asp Pro Pro Ser Asp
65 70 75 80

Thr Asn Asn Arg Arg Ser Trp Pro Thr Gly Asp Gly Gly Asp Gly Gly
85 90 95

Ala Gly Gln Gly Ala Gly Gly Gly Ala Thr Ala Thr Glu Glu Asp Tyr
100 105 110

Arg Ala Glu Asp Leu Asp Glu Leu Tyr Ala Ala Leu Glu Gly Asp Glu
115 120 125

Arg Ser Glu Thr His Ala Pro Arg Thr Asp Ile Pro Val Pro Val Asp
130 135 140

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

Ser Cys Ser Thr Pro Gly Thr Gly Ala Glu Ala Thr Leu Ala Gln Gln
165 170 175

Leu Ser Lys Glu Leu Glu Ser Asn Arg Ser Ile Met Lys Leu Ile Arg

```
            180                 185                 190

Thr Ala Gln Arg Asp Leu Glu Ser Phe Leu Pro Gln Lys Glu Thr Ser
        195                 200                 205

Ser Pro Glu Val Glu Lys Lys Thr His Phe Gln Arg Lys Lys Val Pro
    210                 215                 220

Ile Ser Arg Ser Lys Arg Gly Arg Asn Arg Lys Pro Arg Arg Tyr Ser
225                 230                 235                 240

Ser Thr Ser Ser Asp Thr Ser Glu Ser Ser Glu Asn Ser Glu Ser Asp
                245                 250                 255

Ser Glu Pro Cys Ser Lys Ala Ser Lys Arg Arg Arg Arg Val Ser Thr
            260                 265                 270
```

<210> 841
<211> 265
<212> PRT
<213> Torque teno virus

<400> 841

```
Met Ser Val Trp Leu Pro Pro Ile Asp Asn Val Thr Glu Arg Glu Arg
1               5                   10                  15

Ser Trp Leu Ser Ser Ile Leu Gln Ser His Arg Ala Phe Cys Gly Cys
            20                  25                  30

His Asp Ala Ile Tyr His Leu Ser Ser Leu Ala Ala Arg Phe Asn Met
        35                  40                  45

Gln Pro Gly Pro Ser Pro Gly Gly Asp Ser Arg Pro Pro Arg Pro Pro
    50                  55                  60

Leu Arg Arg Leu Pro Ala Leu Pro Gly Pro Arg Asp Pro Pro Ser Asp
65                  70                  75                  80

Thr Asn Asn Arg Arg Ser Trp Pro Thr Gly Asp Gly Gly Asp Gly Gly
                85                  90                  95

Ala Gly Gln Gly Ala Gly Gly Gly Ala Thr Ala Thr Glu Glu Asp Tyr
```

```
            100                 105                 110

Arg Ala Glu Asp Leu Asp Glu Leu Tyr Ala Ala Leu Glu Gly Asp Glu
        115                 120                 125

Arg Arg Pro Ala Arg Pro Lys Ser Arg Arg Arg Leu Ile Phe Arg Gly
    130                 135                 140

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

Gly Gly Thr Ala Ala Pro Pro Pro Thr Pro Gln Lys Ala Ala Arg Thr
                165                 170                 175

Pro Lys Ala Thr Pro Ser Pro Val Pro Lys Pro Pro Lys Asp Ala Gly
            180                 185                 190

Gly Ser Pro Arg Lys Ser Ile Ile Ile Gln Pro Ala Cys Asn Gln Val
        195                 200                 205

Leu Met Phe Pro Glu Met Gly Pro Lys Pro Ala Pro Thr Ala Gln Asp
    210                 215                 220

Trp Gln Cys Glu Tyr Glu Thr Cys Lys His Trp Asp Arg Pro Pro Arg
225                 230                 235                 240

Lys Phe Leu Thr Asp Pro Pro Phe Tyr Pro Trp Ala Pro Thr Thr Tyr
                245                 250                 255

Asn Val Ser Phe Lys Leu Asn Phe Lys
            260                 265


<210> 842
<211> 762
<212> PRT
<213> Torque teno virus

<400> 842
Met Ala Tyr Trp Gly Trp Trp Arg Arg Arg Arg Trp Pro Arg Arg Arg
1               5                   10                  15

Trp Arg Arg Tyr Arg Tyr Arg Arg Arg Leu Pro Arg Arg Arg Pro Arg
```

```
            20                  25                  30

Arg Ala Val Arg Arg Pro Arg Arg Arg Arg Val Arg Arg Arg Arg Gly
        35                  40                  45

Arg Gly Trp Tyr Arg Gly Arg Arg Tyr Ser Arg Arg Arg Tyr Arg Arg
    50                  55                  60

Arg Tyr Val Arg Arg Lys Arg Lys Thr Leu Val Trp Arg Gln Trp Gln
65                  70                  75                  80

Pro Gln Asn Ile Arg Lys Cys Arg Ile Arg Gly Ile Ile Pro Ile Leu
                85                  90                  95

Ile Cys Gly His Gly Arg Gly Ala Arg Asn Tyr Ala Leu His Ser Asp
            100                 105                 110

Asp Ile Thr Pro Gln Asn Thr Pro Phe Gly Gly Gly Leu Ser Thr Thr
        115                 120                 125

Ser Trp Ser Leu Lys Val Leu Tyr Asp Gln His Thr Arg Gly Leu Asn
    130                 135                 140

Arg Trp Ser Ala Ser Asn Glu Ser Leu Asp Leu Ala Arg Tyr Asn Gly
145                 150                 155                 160

Cys Ser Phe Thr Phe Tyr Arg Asp Lys Lys Thr Asp Phe Ile Val Thr
                165                 170                 175

Tyr Asp Thr Ser Ala Pro Tyr Lys Leu Asp Lys Tyr Ser Ser Pro Ser
            180                 185                 190

Tyr His Pro Gly Ser Met Met Leu Met Thr Lys His Lys Ile Leu Ile
        195                 200                 205

Pro Ser Phe Asp Thr Lys Pro Lys Gly Pro Ala Lys Ile Arg Val Arg
    210                 215                 220

Ile Lys Pro Pro Lys Met Phe Leu Asp Lys Trp Tyr Thr Gln Asp Asp
225                 230                 235                 240
```

```
Leu Cys Ser Val Asn Leu Val Ser Leu Ala Val Ser Ala Ala Ser Phe
                245                 250                 255

Thr His Pro Phe Cys Pro Pro Leu Thr Asp Thr Pro Cys Val Thr Leu
            260                 265                 270

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

Ala Asp Lys Val Glu Ser Val Phe Thr Asn Thr Leu Tyr Lys His Cys
    290                 295                 300

Cys Tyr Tyr Gln Ser Phe Leu Thr Thr Gln Phe Ile Ala Lys Ile Thr
305                 310                 315                 320

Arg Thr Pro Asp Gly Gln Pro Val Ala Thr Phe Ser Pro Pro Thr Ser
                325                 330                 335

Phe Pro Gly Thr Thr Val Thr Lys Ser Ser Ile Glu Ser Phe Asn Gln
            340                 345                 350

Trp Val Thr Ser Thr Gly Thr Ser Gly Trp Leu Thr Asn Ala Asn Gln
        355                 360                 365

His Phe His Phe Cys Asn Tyr Lys Pro Asp Ala Thr Lys Leu Lys Trp
    370                 375                 380

Leu Arg Gln Tyr Tyr Phe Asp Trp Glu Thr Tyr Lys Leu Ala Asp Val
385                 390                 395                 400

Lys Pro Asp Gly Leu Thr Pro Ser Val Asn Trp Tyr Glu Tyr Arg Ile
                405                 410                 415

Gly Leu Phe Ser Pro Ile Phe Leu Ser Pro Phe Arg Ser Ser Ser Leu
            420                 425                 430

Asp Phe Pro Arg Ala Tyr Gln Asp Val Asn Tyr Asn Pro Leu Val Asp
        435                 440                 445

Lys Gly Val Gly Asn Ile Ile Trp Phe Gln Tyr Asn Thr Lys Pro Asp
```

```
    450                 455                 460

Thr Gln Leu Ser Val Pro Ser Cys Lys Cys Val Ile Glu Asp Lys Pro
465                 470                 475                 480

Leu Trp Ala Ala Phe Tyr Gly Tyr Ser Asp Phe Val Gln Gln Glu Ile
                485                 490                 495

Gly Asp Tyr Thr Asp Ala Glu Ala Val Gly Phe Val Cys Val Ile Cys
            500                 505                 510

Pro Tyr Thr Lys Pro Pro Leu Lys Asn Pro Asp Asn Pro Met Gln Gly
        515                 520                 525

Phe Ile Phe Tyr Asp Ser Leu Phe Gly Asn Gly Lys Trp Ile Asp Gly
    530                 535                 540

Thr Gly His Val Pro Leu Tyr Trp Gln Ser Arg Trp Arg Pro Glu Met
545                 550                 555                 560

Leu Phe Gln Glu Asn Thr Met Arg Asp Ile Thr Leu Ser Gly Pro Phe
                565                 570                 575

Ser Tyr Lys Asp Asp Tyr Lys Asn Cys Val Leu Thr Cys Lys Tyr Lys
            580                 585                 590

Phe Asn Phe Arg Phe Gly Gly Asn Leu Leu His Glu Gln Thr Ile Arg
        595                 600                 605

Asn Pro Cys Pro Thr Asp Gly His Pro Ser Thr Gly Arg Gln Pro Arg
    610                 615                 620

Asp Val Gln Val Val Asp Pro Ile Lys Val Gly Pro Arg Phe Val Phe
625                 630                 635                 640

His Ser Trp Asp Trp Arg Arg Gly Tyr Leu Ser Pro Ala Ala Leu Lys
                645                 650                 655

Arg Ile Gly Glu Gln Pro Leu Asp Tyr Glu Ala Tyr Ser Tyr Arg Pro
            660                 665                 670
```

```
Lys Arg Pro Arg Ile Phe Pro Pro Thr Glu Gly Asp Gln Leu Ala Arg
        675                 680                 685

Ser Arg Glu Glu Asp Ser Phe Ser Glu Glu Glu Ser Pro His Ile Ser
    690                 695                 700

Phe Glu Glu Gly Gln Glu Pro Lys Ala Gln Ala Val Gln Gln His Leu
705                 710                 715                 720

Leu Arg His Leu Arg Lys Gln Arg Glu Leu Arg Lys Arg Leu Arg Ala
                725                 730                 735

Leu Phe Gln Ser Leu Gln Lys Thr Gln Ala Gly Leu His Val Asn Pro
            740                 745                 750

Leu Leu Phe Asn Gln Pro Ala Ile Arg Phe
        755                 760
```

<210> 843
<211> 199
<212> PRT
<213> Torque teno virus

<400> 843

```
Met Ala Tyr Trp Gly Trp Trp Arg Arg Arg Arg Trp Pro Arg Arg Arg
1               5                   10                  15

Trp Arg Arg Tyr Arg Tyr Arg Arg Arg Leu Pro Arg Arg Arg Pro Arg
            20                  25                  30

Arg Ala Val Arg Arg Pro Arg Arg Arg Arg Thr Ile Arg Asn Pro Cys
        35                  40                  45

Pro Thr Asp Gly His Pro Ser Thr Gly Arg Gln Pro Arg Asp Val Gln
    50                  55                  60

Val Val Asp Pro Ile Lys Val Gly Pro Arg Phe Val Phe His Ser Trp
65                  70                  75                  80

Asp Trp Arg Arg Gly Tyr Leu Ser Pro Ala Ala Leu Lys Arg Ile Gly
                85                  90                  95
```

Glu Gln Pro Leu Asp Tyr Glu Ala Tyr Ser Tyr Arg Pro Lys Arg Pro
100 105 110

Arg Ile Phe Pro Pro Thr Glu Gly Asp Gln Leu Ala Arg Ser Arg Glu
115 120 125

Glu Asp Ser Phe Ser Glu Glu Glu Ser Pro His Ile Ser Phe Glu Glu
130 135 140

Gly Gln Glu Pro Lys Ala Gln Ala Val Gln Gln His Leu Leu Arg His
145 150 155 160

Leu Arg Lys Gln Arg Glu Leu Arg Lys Arg Leu Arg Ala Leu Phe Gln
165 170 175

Ser Leu Gln Lys Thr Gln Ala Gly Leu His Val Asn Pro Leu Leu Phe
180 185 190

Asn Gln Pro Ala Ile Arg Phe
195

<210> 844
<211> 110
<212> PRT
<213> Torque teno virus

<400> 844
Met Ala Tyr Trp Gly Trp Trp Arg Arg Arg Arg Trp Pro Arg Arg Arg
1 5 10 15

Trp Arg Arg Tyr Arg Tyr Arg Arg Arg Leu Pro Arg Arg Arg Pro Arg
20 25 30

Arg Ala Val Arg Arg Pro Arg Arg Arg Arg Lys Glu Thr Ser Ser Pro
35 40 45

Glu Val Glu Lys Lys Thr His Phe Gln Arg Lys Lys Val Pro Ile Ser
50 55 60

Arg Ser Lys Arg Gly Arg Asn Arg Lys Pro Arg Arg Tyr Ser Ser Thr
65 70 75 80

Ser Ser Asp Thr Ser Glu Ser Ser Glu Asn Ser Glu Ser Asp Ser Glu
85 90 95

Pro Cys Ser Lys Ala Ser Lys Arg Arg Arg Arg Val Ser Thr
100 105 110

<210> 845
<211> 3653
<212> DNA
<213> Torque teno virus

<220>
<221> modified_base
<222> (2949)..(2966)
<223> a, c, t, g, unknown or other

<400> 845
ccaaccagag tctatgtcgt gcacttcctg ggcatggtct acgtaataat ataaagcggt 60

gcacttccga atggctgagt tttccacgcc cgtccgcagc gagatcgcga cggaggagcg 120

atcgagcgtc ccgagggcgg gtgccggagg tgagtttaca caccgcagtc aaggggcaat 180

tcgggctcgg gactggccgg gctatgggca aggctcttaa aaagccatgt ttctcggtaa 240

actttacagg cagaaaagga aactgctact gcagcctgtg cgtgctccac agacgccatc 300

ttccatgagc tctacctggc gagtgccccg cggcgatgtc tccgcccgcg agctatgttg 360

gtaccgctca gttcgagaga gccacgatgc tttttgtggc tgtcgtgatc ctgtttttca 420

tctttctcgt ctggctgcac gttctaacca tcagggacct ccgacgcccc ccacggacga 480

gcgcccgtcg gcgtctaccc cagtgaggcg cctgctgccg ctgccctcct accccggcga 540

gggtccccag gctagatggc ctggtgggga tggagaaggc gctggtggcg cccgcggagg 600

cgctggagat ggcggcgccc gcgcaggcga agaagagtac cggcccgaag acctcgacga 660

gctgttcgac gctatcgaac aagaacagta aggagacgga ggcgagggtg gcggaggggc 720

tacaggcgcc gttacagact gagacgctac cgtagaaggg gcaggcgacg caaaaaaata 780

gtactgactc agtggaaccc ccagactgtc agaaagtgct ttatcagagg actgatgcca 840

gtactatggg cgggcatggg cacgggggggc cacaactacg ccgtccgctc agatgacttt 900

gtggtagaca gaggcttcgg gggctccttc gccacagaaa ctttctccct gagggtcctc 960

tttgaccagt accagagagg atttaatagg tggtctcaca ccaacgaaga cctagacctg 1020

gcccgctaca cgggctgcaa atggacattt tacagacacc aagacacaga ctttatagtg 1080

tactttacaa acaatccccc catgaaaacc aaccagcaca cagcccctct cacaactcca 1140

ggcatgctca tgaggagcaa gtataaaata ctagtgccca gttttaaaac aagaccaaag 1200

ggcagaaaaa cagtgtcagt gagagttaga ccccccaaac tgtttcagga caaatggtat 1260

actcaacagg acctctgtcc agtacccctc gtccaactga acgtgaccgc agcggatttc 1320

acacatccgt tcggctcacc actaactgac acgccttgca taagattcca agttttaggg 1380

aacttataca acaagtgcct aaatatagat cttccgcaat ttgatgagga cggtgagata 1440

ctcacttcaa caccttataa cagagaaaac aaagaagatc ttaaaaagct ttataaaact 1500

ctatttgtag atgaacacgc aggcaattat tggcagacat tcttaaccaa cacaatggta 1560

aagtcacaca tagatgcaaa ccaagcaaag acatacgatc aagaaaaaac tgctgcagaa 1620

caaggtaaag accccttccc aacaaaccca ccaaaagacc aattcactac ctggaacaag 1680

aaactagtag accctagaga cagcaacttt ctctttgcca catatcaccc aaaaaacatt 1740

aaaaaagcta taaaaaccat gagagacaac aactttgctc tcaccacagg caaaaatgac 1800

atatatggag actacaccgc ggcctacacc agaaacaccc acatgctaga ctactaccta 1860

ggcttttata gccccatatt tctttccagc ggtaggtcca acacagagtt ctggaccgcc 1920

tacagagaca tagtatataa tcccctctta gacaaaggca caggcaacat gatctggttc 1980

caatatcaca caaaaacaga caatatatac aaaaaaccag agtgccactg ggagatacta 2040

gacatgcccc tgtgggccct ctgcaacggg tatgtagagt acctagagag ccaaataaag 2100

tacggggaca tcctagtaga gggcaaagtc ctcatcagat gcccctacac caaacccgca 2160

ctggtagacc ccaataacag cctagctggt tacgtggtat tcaacaccac cttcggccag 2220

ggaaaatgga tagatggcaa aggctacatc cccctacacg agaggagcaa gtggtacgtc 2280

atgctcagat accagaccga cgtactccat gacatagtga cttgtggacc ctggcagtac 2340

agagacgata acaaaaactc tcagctaata gccaagtaca gattcaagtt ctactgggga 2400

ggtaacatgg tacattctca ggtcatcaga aacccgtgca aagacaccca agtatccgga 2460

ccccgtcgac agcctcgcga agtacaagtc gttgacccgc aactcattac gccgccgtgg 2520

gtcctccact cgttcgacca gagacgagga atgtttactg caggagctat caaacgtctg 2580

ctcaagcaac caatacctgg cgagtatgct cctacaccac tcagggtccc gctcctcttt 2640

ccctcctcag agttccagcg agagggagaa gatgcagaaa gcggctcagg ttcaccaccc 2700

aagagaccgc gactctggca ggaagaggcc aaccagacgc aaacggagtc ctcggagggg 2760

ccggcggaga cgacgaggga gctcctcgag cgaaagctca gagagcagcg agtcctcaac 2820

ctccaactcc agcatgtcgc agtacaactc gccaaaaccc aagcgaacct ccacataaac 2880

cccctattat actcccagcc ttaaacaaag tgtatctatt ccccccctgac aagcccactc 2940

ccatacagnn nnnnnnnnnn nnnnnnaaca cagagttcga agcctgccag gccttcgaca 3000

gaccacctag aaaatacctc tcagacacac ctacctaccc ttggctcccc gtccccaatc 3060

ctgaaataaa ggtcagcttt aagctcggtt tcaaatctta caaggccgtg ggagtttcac 3120

tggtcggtgt ctacctctta aggtcactaa gcactccgag cgtcagcgag gagtgcgacc 3180

cttccccctg gtgcaacgcc ctcggcggcc gcgcgctacg ccttcggctg cgcgcggcac 3240

ctcggacccc cgctcgtgct gacgcgctcg cgcgcgtcag accacttcgg gctcgcgggg 3300

gtcgggaatt ttgctaaaca gactccgagt tgccattgga cactgtagct gtgaatcagt 3360

aacgaaagtg agtggggcca gacttcgcca taaggccttt atcttcttgc cattggtccg 3420

tgtagggggt cgccataggc ttcgggttcg gttttaggcc ttccggacta caaaaatggc 3480

ggatttagtg acgtcacggc cgccatttta agtaggtgcc gtccaggact gctgttccgg 3540

gtcacagggc atcctcggcg gaacttacac aaaatggcgg tcaaaaacat ccgggtcaaa 3600

ggtcgcagct acgtcataag tcacgtgcag gggtcctgct gcgtcatatg cgg 3653

<210> 846
<211> 128
<212> PRT
<213> Torque teno virus

<400> 846
Met Ser Ser Thr Trp Arg Val Pro Arg Gly Asp Val Ser Ala Arg Glu
1 5 10 15

Leu Cys Trp Tyr Arg Ser Val Arg Glu Ser His Asp Ala Phe Cys Gly
20 25 30

Cys Arg Asp Pro Val Phe His Leu Ser Arg Leu Ala Ala Arg Ser Asn
35 40 45

```
His Gln Gly Pro Pro Thr Pro Pro Thr Asp Glu Arg Pro Ser Ala Ser
    50                  55                  60

Thr Pro Val Arg Arg Leu Leu Pro Leu Pro Ser Tyr Pro Gly Glu Gly
65                  70                  75                  80

Pro Gln Ala Arg Trp Pro Gly Gly Asp Gly Glu Gly Ala Gly Gly Ala
                85                  90                  95

Arg Gly Gly Ala Gly Asp Gly Gly Ala Arg Ala Gly Glu Glu Glu Tyr
            100                 105                 110

Arg Pro Glu Asp Leu Asp Glu Leu Phe Asp Ala Ile Glu Gln Glu Gln
        115                 120                 125
```

<210> 847
<211> 279
<212> PRT
<213> Torque teno virus

<400> 847

```
Met Ser Ser Thr Trp Arg Val Pro Arg Gly Asp Val Ser Ala Arg Glu
1               5                   10                  15

Leu Cys Trp Tyr Arg Ser Val Arg Glu Ser His Asp Ala Phe Cys Gly
            20                  25                  30

Cys Arg Asp Pro Val Phe His Leu Ser Arg Leu Ala Ala Arg Ser Asn
        35                  40                  45

His Gln Gly Pro Pro Thr Pro Pro Thr Asp Glu Arg Pro Ser Ala Ser
    50                  55                  60

Thr Pro Val Arg Arg Leu Leu Pro Leu Pro Ser Tyr Pro Gly Glu Gly
65                  70                  75                  80

Pro Gln Ala Arg Trp Pro Gly Gly Asp Gly Glu Gly Ala Gly Gly Ala
                85                  90                  95

Arg Gly Gly Ala Gly Asp Gly Gly Ala Arg Ala Gly Glu Glu Glu Tyr
            100                 105                 110
```

```
Arg Pro Glu Asp Leu Asp Glu Leu Phe Asp Ala Ile Glu Gln Glu Gln
        115                 120                 125

Ser Ser Glu Thr Arg Ala Lys Thr Pro Lys Tyr Pro Asp Pro Val Asp
    130                 135                 140

Ser Leu Ala Lys Tyr Lys Ser Leu Thr Arg Asn Ser Leu Arg Arg Arg
145                 150                 155                 160

Gly Ser Ser Thr Arg Ser Thr Arg Asp Glu Glu Cys Leu Leu Gln Glu
                165                 170                 175

Leu Ser Asn Val Cys Ser Ser Asn Gln Tyr Leu Ala Ser Met Leu Leu
            180                 185                 190

His His Ser Gly Ser Arg Ser Ser Phe Pro Pro Gln Ser Ser Ser Glu
        195                 200                 205

Arg Glu Lys Met Gln Lys Ala Ala Gln Val His His Pro Arg Asp Arg
    210                 215                 220

Asp Ser Gly Arg Lys Arg Pro Thr Arg Arg Lys Arg Ser Pro Arg Arg
225                 230                 235                 240

Gly Arg Arg Arg Arg Arg Gly Ser Ser Ser Ser Glu Ser Ser Glu Ser
                245                 250                 255

Ser Glu Ser Ser Thr Ser Asn Ser Ser Met Ser Gln Tyr Asn Ser Pro
            260                 265                 270

Lys Pro Lys Arg Thr Ser Thr
        275
```

<210> 848
<211> 378
<212> PRT
<213> Torque teno virus

<220>
<221> MOD_RES
<222> (257)..(262)
<223> Any amino acid

<400> 848
Met Ser Ser Thr Trp Arg Val Pro Arg Gly Asp Val Ser Ala Arg Glu
1   5   10   15

Leu Cys Trp Tyr Arg Ser Val Arg Glu Ser His Asp Ala Phe Cys Gly
            20   25   30

Cys Arg Asp Pro Val Phe His Leu Ser Arg Leu Ala Ala Arg Ser Asn
        35   40   45

His Gln Gly Pro Pro Thr Pro Pro Thr Asp Glu Arg Pro Ser Ala Ser
    50   55   60

Thr Pro Val Arg Arg Leu Leu Pro Leu Pro Ser Tyr Pro Gly Glu Gly
65   70   75   80

Pro Gln Ala Arg Trp Pro Gly Gly Asp Gly Glu Gly Ala Gly Gly Ala
                85   90   95

Arg Gly Gly Ala Gly Asp Gly Gly Ala Arg Ala Gly Glu Glu Glu Tyr
            100   105   110

Arg Pro Glu Asp Leu Asp Glu Leu Phe Asp Ala Ile Glu Gln Glu Gln
        115   120   125

Ser Tyr Gln Thr Ser Ala Gln Ala Thr Asn Thr Trp Arg Val Cys Ser
    130   135   140

Tyr Thr Thr Gln Gly Pro Ala Pro Leu Ser Leu Leu Arg Val Pro Ala
145   150   155   160

Arg Gly Arg Arg Cys Arg Lys Arg Leu Arg Phe Thr Thr Gln Glu Thr
                165   170   175

Ala Thr Leu Ala Gly Arg Gly Gln Pro Asp Ala Asn Gly Val Leu Gly
            180   185   190

Gly Ala Gly Gly Asp Asp Glu Gly Ala Pro Arg Ala Lys Ala Gln Arg
        195   200   205

```
Ala Ala Ser Pro Gln Pro Pro Thr Pro Ala Cys Arg Ser Thr Thr Arg
    210                 215                 220

Gln Asn Pro Ser Glu Pro Pro His Lys Pro Pro Ile Ile Leu Pro Ala
225                 230                 235                 240

Leu Asn Lys Val Tyr Leu Phe Pro Pro Asp Lys Pro Thr Pro Ile Gln
                245                 250                 255

Xaa Xaa Xaa Xaa Xaa Xaa Asn Thr Glu Phe Glu Ala Cys Gln Ala Phe
            260                 265                 270

Asp Arg Pro Pro Arg Lys Tyr Leu Ser Asp Thr Pro Thr Tyr Pro Trp
        275                 280                 285

Leu Pro Val Pro Asn Pro Glu Ile Lys Val Ser Phe Lys Leu Gly Phe
    290                 295                 300

Lys Ser Tyr Lys Ala Val Gly Val Ser Leu Val Gly Val Tyr Leu Leu
305                 310                 315                 320

Arg Ser Leu Ser Thr Pro Ser Val Ser Glu Glu Cys Asp Pro Ser Pro
                325                 330                 335

Trp Cys Asn Ala Leu Gly Gly Arg Ala Leu Arg Leu Arg Leu Arg Ala
            340                 345                 350

Ala Pro Arg Thr Pro Ala Arg Ala Asp Ala Leu Ala Arg Val Arg Pro
        355                 360                 365

Leu Arg Ala Arg Gly Gly Arg Glu Phe Cys
    370                 375
```

<210> 849
<211> 269
<212> PRT
<213> Torque teno virus

<220>
<221> MOD_RES
<222> (148)..(153)
<223> Any amino acid

<400> 849
Met Ser Ser Thr Trp Arg Val Pro Arg Gly Asp Val Ser Ala Arg Glu
1   5   10   15

Leu Cys Trp Ser Tyr Gln Thr Ser Ala Gln Ala Thr Asn Thr Trp Arg
            20   25   30

Val Cys Ser Tyr Thr Thr Gln Gly Pro Ala Pro Leu Ser Leu Leu Arg
        35   40   45

Val Pro Ala Arg Gly Arg Arg Cys Arg Lys Arg Leu Arg Phe Thr Thr
    50   55   60

Gln Glu Thr Ala Thr Leu Ala Gly Arg Gly Gln Pro Asp Ala Asn Gly
65   70   75   80

Val Leu Gly Gly Ala Gly Gly Asp Asp Glu Gly Ala Pro Arg Ala Lys
                85   90   95

Ala Gln Arg Ala Ala Ser Pro Gln Pro Pro Thr Pro Ala Cys Arg Ser
            100   105   110

Thr Thr Arg Gln Asn Pro Ser Glu Pro Pro His Lys Pro Pro Ile Ile
        115   120   125

Leu Pro Ala Leu Asn Lys Val Tyr Leu Phe Pro Pro Asp Lys Pro Thr
    130   135   140

Pro Ile Gln Xaa Xaa Xaa Xaa Xaa Xaa Asn Thr Glu Phe Glu Ala Cys
145   150   155   160

Gln Ala Phe Asp Arg Pro Pro Arg Lys Tyr Leu Ser Asp Thr Pro Thr
                165   170   175

Tyr Pro Trp Leu Pro Val Pro Asn Pro Glu Ile Lys Val Ser Phe Lys
            180   185   190

Leu Gly Phe Lys Ser Tyr Lys Ala Val Gly Val Ser Leu Val Gly Val
        195   200   205

```
Tyr Leu Leu Arg Ser Leu Ser Thr Pro Ser Val Ser Glu Glu Cys Asp
    210                 215                 220

Pro Ser Pro Trp Cys Asn Ala Leu Gly Gly Arg Ala Leu Arg Leu Arg
225                 230                 235                 240

Leu Arg Ala Ala Pro Arg Thr Pro Ala Arg Ala Asp Ala Leu Ala Arg
                245                 250                 255

Val Arg Pro Leu Arg Ala Arg Gly Gly Arg Glu Phe Cys
            260                 265
```

<210> 850
<211> 782
<212> PRT
<213> Torque teno virus

<400> 850

```
Met Ala Trp Trp Gly Trp Arg Arg Arg Trp Trp Arg Pro Arg Arg Arg
1               5                   10                  15

Trp Arg Trp Arg Arg Pro Arg Arg Arg Arg Arg Val Pro Ala Arg Arg
            20                  25                  30

Pro Arg Arg Ala Val Arg Arg Tyr Arg Thr Arg Thr Val Arg Arg Arg
        35                  40                  45

Arg Arg Gly Trp Arg Arg Gly Tyr Arg Arg Arg Tyr Arg Leu Arg Arg
    50                  55                  60

Tyr Arg Arg Arg Gly Arg Arg Arg Lys Lys Ile Val Leu Thr Gln Trp
65                  70                  75                  80

Asn Pro Gln Thr Val Arg Lys Cys Phe Ile Arg Gly Leu Met Pro Val
                85                  90                  95

Leu Trp Ala Gly Met Gly Thr Gly Gly His Asn Tyr Ala Val Arg Ser
            100                 105                 110

Asp Asp Phe Val Val Asp Arg Gly Phe Gly Gly Ser Phe Ala Thr Glu
        115                 120                 125
```

```
Thr Phe Ser Leu Arg Val Leu Phe Asp Gln Tyr Gln Arg Gly Phe Asn
    130                 135                 140

Arg Trp Ser His Thr Asn Glu Asp Leu Asp Leu Ala Arg Tyr Thr Gly
145                 150                 155                 160

Cys Lys Trp Thr Phe Tyr Arg His Gln Asp Thr Asp Phe Ile Val Tyr
                165                 170                 175

Phe Thr Asn Asn Pro Pro Met Lys Thr Asn Gln His Thr Ala Pro Leu
            180                 185                 190

Thr Thr Pro Gly Met Leu Met Arg Ser Lys Tyr Lys Ile Leu Val Pro
        195                 200                 205

Ser Phe Lys Thr Arg Pro Lys Gly Arg Lys Thr Val Ser Val Arg Val
    210                 215                 220

Arg Pro Pro Lys Leu Phe Gln Asp Lys Trp Tyr Thr Gln Gln Asp Leu
225                 230                 235                 240

Cys Pro Val Pro Leu Val Gln Leu Asn Val Thr Ala Ala Asp Phe Thr
                245                 250                 255

His Pro Phe Gly Ser Pro Leu Thr Asp Thr Pro Cys Ile Arg Phe Gln
            260                 265                 270

Val Leu Gly Asn Leu Tyr Asn Lys Cys Leu Asn Ile Asp Leu Pro Gln
        275                 280                 285

Phe Asp Glu Asp Gly Glu Ile Leu Thr Ser Thr Pro Tyr Asn Arg Glu
    290                 295                 300

Asn Lys Glu Asp Leu Lys Lys Leu Tyr Lys Thr Leu Phe Val Asp Glu
305                 310                 315                 320

His Ala Gly Asn Tyr Trp Gln Thr Phe Leu Thr Asn Thr Met Val Lys
                325                 330                 335

Ser His Ile Asp Ala Asn Gln Ala Lys Thr Tyr Asp Gln Glu Lys Thr
            340                 345                 350
```

```
Ala Ala Glu Gln Gly Lys Asp Pro Phe Pro Thr Asn Pro Pro Lys Asp
        355                 360                 365

Gln Phe Thr Thr Trp Asn Lys Lys Leu Val Asp Pro Arg Asp Ser Asn
    370                 375                 380

Phe Leu Phe Ala Thr Tyr His Pro Lys Asn Ile Lys Lys Ala Ile Lys
385                 390                 395                 400

Thr Met Arg Asp Asn Asn Phe Ala Leu Thr Thr Gly Lys Asn Asp Ile
                405                 410                 415

Tyr Gly Asp Tyr Thr Ala Ala Tyr Thr Arg Asn Thr His Met Leu Asp
            420                 425                 430

Tyr Tyr Leu Gly Phe Tyr Ser Pro Ile Phe Leu Ser Ser Gly Arg Ser
        435                 440                 445

Asn Thr Glu Phe Trp Thr Ala Tyr Arg Asp Ile Val Tyr Asn Pro Leu
    450                 455                 460

Leu Asp Lys Gly Thr Gly Asn Met Ile Trp Phe Gln Tyr His Thr Lys
465                 470                 475                 480

Thr Asp Asn Ile Tyr Lys Lys Pro Glu Cys His Trp Glu Ile Leu Asp
                485                 490                 495

Met Pro Leu Trp Ala Leu Cys Asn Gly Tyr Val Glu Tyr Leu Glu Ser
            500                 505                 510

Gln Ile Lys Tyr Gly Asp Ile Leu Val Glu Gly Lys Val Leu Ile Arg
        515                 520                 525

Cys Pro Tyr Thr Lys Pro Ala Leu Val Asp Pro Asn Asn Ser Leu Ala
    530                 535                 540

Gly Tyr Val Val Phe Asn Thr Thr Phe Gly Gln Gly Lys Trp Ile Asp
545                 550                 555                 560
```

```
Gly Lys Gly Tyr Ile Pro Leu His Glu Arg Ser Lys Trp Tyr Val Met
                565                 570                 575

Leu Arg Tyr Gln Thr Asp Val Leu His Asp Ile Val Thr Cys Gly Pro
            580                 585                 590

Trp Gln Tyr Arg Asp Asp Asn Lys Asn Ser Gln Leu Ile Ala Lys Tyr
        595                 600                 605

Arg Phe Lys Phe Tyr Trp Gly Gly Asn Met Val His Ser Gln Val Ile
    610                 615                 620

Arg Asn Pro Cys Lys Asp Thr Gln Val Ser Gly Pro Arg Arg Gln Pro
625                 630                 635                 640

Arg Glu Val Gln Val Val Asp Pro Gln Leu Ile Thr Pro Pro Trp Val
                645                 650                 655

Leu His Ser Phe Asp Gln Arg Arg Gly Met Phe Thr Ala Gly Ala Ile
            660                 665                 670

Lys Arg Leu Leu Lys Gln Pro Ile Pro Gly Glu Tyr Ala Pro Thr Pro
        675                 680                 685

Leu Arg Val Pro Leu Leu Phe Pro Ser Ser Glu Phe Gln Arg Glu Gly
    690                 695                 700

Glu Asp Ala Glu Ser Gly Ser Gly Ser Pro Pro Lys Arg Pro Arg Leu
705                 710                 715                 720

Trp Gln Glu Glu Ala Asn Gln Thr Gln Thr Glu Ser Ser Glu Gly Pro
                725                 730                 735

Ala Glu Thr Thr Arg Glu Leu Leu Glu Arg Lys Leu Arg Glu Gln Arg
            740                 745                 750

Val Leu Asn Leu Gln Leu Gln His Val Ala Val Gln Leu Ala Lys Thr
        755                 760                 765

Gln Ala Asn Leu His Ile Asn Pro Leu Leu Tyr Ser Gln Pro
    770                 775                 780
```

<210> 851
<211> 20
<212> PRT
<213> Torque teno midi virus

<400> 851
Leu Arg Tyr Asn Pro Ala Ala Asp Gln Gly Pro Gly Asn Ile Val Thr
1   5   10   15

Ile Ile Asp Leu
            20

<210> 852
<211> 204
<212> PRT
<213> Torque teno virus

<400> 852
Met Ala Trp Trp Gly Trp Arg Arg Arg Trp Trp Arg Pro Arg Arg Arg
1   5   10   15

Trp Arg Trp Arg Arg Pro Arg Arg Arg Arg Arg Val Pro Ala Arg Arg
            20   25   30

Pro Arg Arg Ala Val Arg Arg Tyr Arg Thr Arg Thr Val Ile Arg Asn
        35   40   45

Pro Cys Lys Asp Thr Gln Val Ser Gly Pro Arg Arg Gln Pro Arg Glu
    50   55   60

Val Gln Val Val Asp Pro Gln Leu Ile Thr Pro Pro Trp Val Leu His
65   70   75   80

Ser Phe Asp Gln Arg Arg Gly Met Phe Thr Ala Gly Ala Ile Lys Arg
                85   90   95

Leu Leu Lys Gln Pro Ile Pro Gly Glu Tyr Ala Pro Thr Pro Leu Arg
            100   105   110

Val Pro Leu Leu Phe Pro Ser Ser Glu Phe Gln Arg Glu Gly Glu Asp
        115   120   125

```
Ala Glu Ser Gly Ser Gly Ser Pro Pro Lys Arg Pro Arg Leu Trp Gln
    130                 135                 140

Glu Glu Ala Asn Gln Thr Gln Thr Glu Ser Ser Glu Gly Pro Ala Glu
145                 150                 155                 160

Thr Thr Arg Glu Leu Leu Glu Arg Lys Leu Arg Glu Gln Arg Val Leu
                165                 170                 175

Asn Leu Gln Leu Gln His Val Ala Val Gln Leu Ala Lys Thr Gln Ala
            180                 185                 190

Asn Leu His Ile Asn Pro Leu Leu Tyr Ser Gln Pro
        195                 200
```

<210> 853
<211> 148
<212> PRT
<213> Torque teno virus

<400> 853

```
Met Ala Trp Trp Gly Trp Arg Arg Arg Trp Trp Arg Pro Arg Arg Arg
1               5                   10                  15

Trp Arg Trp Arg Arg Pro Arg Arg Arg Arg Arg Val Pro Ala Arg Arg
            20                  25                  30

Pro Arg Arg Ala Val Arg Arg Tyr Arg Thr Arg Thr Glu Leu Ser Asn
        35                  40                  45

Val Cys Ser Ser Asn Gln Tyr Leu Ala Ser Met Leu Leu His His Ser
    50                  55                  60

Gly Ser Arg Ser Ser Phe Pro Pro Gln Ser Ser Ser Glu Arg Glu Lys
65                  70                  75                  80

Met Gln Lys Ala Ala Gln Val His His Pro Arg Asp Arg Asp Ser Gly
                85                  90                  95

Arg Lys Arg Pro Thr Arg Arg Lys Arg Ser Pro Arg Arg Gly Arg Arg
            100                 105                 110
```

Arg Arg Arg Gly Ser Ser Ser Ser Glu Ser Ser Glu Ser Ser Glu Ser
        115   120   125

Ser Thr Ser Asn Ser Ser Met Ser Gln Tyr Asn Ser Pro Lys Pro Lys
    130   135   140

Arg Thr Ser Thr
145

<210> 854
<211> 3742
<212> DNA
<213> Torque teno virus

<400> 854
aaagtgctac gtcactaacc acgtgacacc cacaggccaa ccgaatgcta tgtcgtgcac 60

ttcctgggcc gggtctacgt cctcatataa ctacctgcac ttccgaatgg ctgagttttc 120

cacgcccgtc cgcagcggtg aagccacgga gggagatcag cgcgtcccga gggcgggtgc 180

cgaaggtgag tttacacacc gaagtcaagg ggcaattcgg gctcgggact ggccgggcta 240

tgggcaaggc tctgaaaaaa gcatgtttat tggcaggcat tacagaaaga aaagggcgct 300

gccactgtgt gctgtgcgat caacaaagaa ggcttgcaaa ctactaatag taatgtggac 360

cccacctcgc aatgaccaac agtaccttaa ctggcaatgg tactcaagta tacttagctc 420

ccacgctgct atgtgcgggt gtcccgacgt tgttgctcat tttaatcatc ttgcttctgt 480

gcttcgcgcc ccgcaaaatc cacccccacc cggtccccag cgaaacctgc ccctccgacg 540

gctgccggct ctcccggctg cgccagaggc gcccggagat agagcaccat ggcctatggc 600

tggtggcgcc ggaggagaag acggtggcgc aggtggagac gcagaccatg gaggcgccgc 660

tggaggacca gaagacgcag acctgttaga cgccgtggcc gccgcagaaa cgtaaggaga 720

cgccgcagag gagggaggtg gaggaggagg tacaggagat ggaaaagaaa gggcagacgc 780

agaaaaaaag ctaaaataat aataagacaa tggcaaccta actacagaag gagatgtaac 840

atagtaggct atattcctgt actgatatgt ggcgaaaata ctgtcagcag aaactatgcc 900

acacactcag acgatactaa ctacccagga ccctttgggg ggggtatgac tacagacaaa 960

tttaccttaa gaattctgta tgacgagtac aaaaggttta tgaactattg gacagcatct 1020

aatgaagacc tagacctctg tagatatcta ggagtaaacc tgtacttttt tagacaccca 1080

gaagtagact ttattataaa aataaatacc atgccccctt ttctagacac agaactaaca 1140

gctcctagca tacacccagg aatgctagcc ttagacaaaa gagcaagatg gatacctagc 1200

ttaaaatcta gaccaggaaa aaaacactat attaaaataa gagtaggggc gcctaaaatg 1260

ttcacagata aatggtaccc ccaaacagat ctttgtgaca tggtgctgct aactgtctat 1320

gcaaccgcag cggatatgca atatccgttc ggctcaccac taactgactc tgtggttgtg 1380

aacttccagg ttctgcaatc catgtatgat gaaaccatta gcatattacc agatcaaaag 1440

gagaaaagaa taacgctgct cactagtata gccttttata acaccacaca aactatagcc 1500

caattaaagc catttataga tgcaggcaat atgacttcaa ctacaacagc aacaacatgg 1560

ggatcataca taaacacaac caaatttaat acagcagcca ctacaacata cacataccca 1620

ggcagtacta caactacagt aactatgtta acttgtaatg actcctggta cagaggaaca 1680

gtatataacg accaaattaa aaatttacca aaggaagcag ctcaattata cttaaaagca 1740

acaaaaacct tactaggaaa caccttcaca aatgacgacc acacactaga ataccatgga 1800

ggactgtaca gctcaatttg gctgtccccc ggcagatctt actttgaaac accaggagca 1860

tacacagaca taaaatacaa cccatttaca gacagaggag aaggaaacat gctatggata 1920

gactggctaa gcaaaaaaaa tatgaactat gacaaactac aaagtaaatg tttaatatca 1980

gacctacctt tatgggcagc agcatatgga tatttagaat tttgtgcaaa aagtacagga 2040

gaccaaaata tacacatgaa tgccagacta ctaataagaa gtccctttac agacccccaa 2100

ctactagtac acacaaaccc cacaaaaggc tttgttccct actctttaaa ctttggaaat 2160

ggtaaaatgc caggaggtag tagtaatgtt cctattagaa tgagagctaa atggtatcca 2220

acattgtttc accagcaaga agtactagag gccttagcac agtcaggccc ctttgcatac 2280

cactcagaca ttaaaaaagt atctctgggt atgaaatacc gttttaagtg gatctggggt 2340

ggaaaccccg ttcgccaaca ggttgttaga aatccctgca aagactccca ctcctcggtc 2400

aatagagtcc ctagaagctt acaaatcgtt gacccgaaat acaactcacc ggaactcaca 2460

ttccatacgt gggacttcag acgtggcctc tttggccaga aagctattga gagaatgcaa 2520

caacaaccaa caactactga catttttca gcaggccgca agagacccag gagggacacc 2580

gaggtgtacc actccagcca agaaggggag caaaaagaaa gcttactttt ccccccagtc 2640

aagctcctca gacgagtccc cccgtgggaa gactcgcagc aggaggaaag cgggtcgcaa 2700

agctcagagg aagagacgca gaccgtctcc cagcagctca agcagcagct gcagcaacag 2760

cgaatcctgg gagtcaaact catactcctg ttcaaccaag tccaaaaaat ccaacaaaat 2820

caagatatca accctacctt gttaccaagg gggggggatc tagcatcctt atttcaaata 2880

gcaccataaa catgtttgga gaccccaaac cttacaaccc ttccagtaat gactggaaag 2940

aggagtatga ggcctgtaga atatgggaca gacccccaag aggcaatcta agagacaccc 3000

ccttttaccc ctgggccccc aaagaaaacc agtaccgtgt aaactttaaa cttggatttc 3060

aataaagcta ggccgtggga ctttcacttg tcggtgtctg cttataaaag taaccaagca 3120

ctccgagcga agcgaggagt gcgacccttg ggggctcaac gacttcggag ccgcgcgtta 3180

agccttcggc tgcgcgcggc acctcagacc cccgctcgtg ctgacacgct tgcgcgtgtc 3240

agaccacttc gggctcgcgg gggtcgggaa atttattaaa cagactccga gttgccattg 3300

gacacagtag tctatgaaca gcaacgaaag tgagtggggc cagacttcgc cataaggcct 3360

ttatcttctt gccatttgtc agtatagagg gtcgccatag gcttcggtct ccattttaac 3420

ctgtaaaaac taccaaaatg gccgttccag tgacgtgaca gccgccattt taagtagctg 3480

acgtcaagga ttgacgtaaa ggttaaaggt catcctcggc ggaagctaca caaaatggtg 3540

gacaacatct tccgggtcaa aggtcgtgca cacgtcaaaa gtcacgtggt ggggacccgc 3600

tgtaacccgg aagtaggccc cgtcacgtga tttgtcacgt gtgtacacgt cacagccgcc 3660

attttgtttt acaaaatggc tgacttcctt cctctttttt gaaaaaaggc gccaaaaaag 3720

gctccgcccc ccggcccccc cc 3742

<210> 855
<211> 120
<212> PRT
<213> Torque teno virus

<400> 855
Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1 5 10 15

Tyr Ser Ser Ile Leu Ser Ser His Ala Ala Met Cys Gly Cys Pro Asp
20 25 30

Val Val Ala His Phe Asn His Leu Ala Ser Val Leu Arg Ala Pro Gln
35 40 45

Asn Pro Pro Pro Pro Gly Pro Gln Arg Asn Leu Pro Leu Arg Arg Leu
    50   55   60

Pro Ala Leu Pro Ala Ala Pro Glu Ala Pro Gly Asp Arg Ala Pro Trp
65   70   75   80

Pro Met Ala Gly Gly Ala Gly Gly Glu Asp Gly Gly Ala Gly Gly Asp
                85   90   95

Ala Asp His Gly Gly Ala Ala Gly Gly Pro Glu Asp Ala Asp Leu Leu
            100   105   110

Asp Ala Val Ala Ala Ala Glu Thr
        115   120

<210> 856
<211> 286
<212> PRT
<213> Torque teno virus

<400> 856
Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1   5   10   15

Tyr Ser Ser Ile Leu Ser Ser His Ala Ala Met Cys Gly Cys Pro Asp
            20   25   30

Val Val Ala His Phe Asn His Leu Ala Ser Val Leu Arg Ala Pro Gln
        35   40   45

Asn Pro Pro Pro Pro Gly Pro Gln Arg Asn Leu Pro Leu Arg Arg Leu
    50   55   60

Pro Ala Leu Pro Ala Ala Pro Glu Ala Pro Gly Asp Arg Ala Pro Trp
65   70   75   80

Pro Met Ala Gly Gly Ala Gly Gly Glu Asp Gly Gly Ala Gly Gly Asp
                85   90   95

Ala Asp His Gly Gly Ala Ala Gly Gly Pro Glu Asp Ala Asp Leu Leu
            100   105   110

```
Asp Ala Val Ala Ala Ala Glu Thr Leu Leu Glu Ile Pro Ala Lys Thr
        115                 120                 125

Pro Thr Pro Arg Ser Ile Glu Ser Leu Glu Ala Tyr Lys Ser Leu Thr
    130                 135                 140

Arg Asn Thr Thr His Arg Asn Ser His Ser Ile Arg Gly Thr Ser Asp
145                 150                 155                 160

Val Ala Ser Leu Ala Arg Lys Leu Leu Arg Glu Cys Asn Asn Asn Gln
                165                 170                 175

Gln Leu Leu Thr Phe Phe Gln Gln Ala Ala Arg Asp Pro Gly Gly Thr
            180                 185                 190

Pro Arg Cys Thr Thr Pro Ala Lys Lys Gly Ser Lys Lys Lys Ala Tyr
        195                 200                 205

Phe Ser Pro Gln Ser Ser Ser Ser Asp Glu Ser Pro Arg Gly Lys Thr
    210                 215                 220

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

Pro Ser Pro Ser Ser Ser Ser Ser Ser Cys Ser Asn Ser Glu Ser Trp
                245                 250                 255

Glu Ser Asn Ser Tyr 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> 857
<211> 289
<212> PRT
<213> Torque teno virus

<400> 857
Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1               5                   10                  15
```

```
Tyr Ser Ser Ile Leu Ser Ser His Ala Ala Met Cys Gly Cys Pro Asp
            20                  25                  30

Val Val Ala His Phe Asn His Leu Ala Ser Val Leu Arg Ala Pro Gln
        35                  40                  45

Asn Pro Pro Pro Pro Gly Pro Gln Arg Asn Leu Pro Leu Arg Arg Leu
    50                  55                  60

Pro Ala Leu Pro Ala Ala Pro Glu Ala Pro Gly Asp Arg Ala Pro Trp
65                  70                  75                  80

Pro Met Ala Gly Gly Ala Gly Gly Glu Asp Gly Gly Ala Gly Gly Asp
                85                  90                  95

Ala Asp His Gly Gly Ala Ala Gly Gly Pro Glu Asp Ala Asp Leu Leu
            100                 105                 110

Asp Ala Val Ala Ala Ala Glu Thr Pro Gln Glu Thr Gln Glu Gly His
        115                 120                 125

Arg Gly Val Pro Leu Gln Pro Arg Arg Gly Ala Lys Arg Lys Leu Thr
    130                 135                 140

Phe Pro Pro Ser Gln Ala Pro Gln Thr Ser Pro Pro Val Gly Arg Leu
145                 150                 155                 160

Ala Ala Gly Gly Lys Arg Val Ala Lys Leu Arg Gly Arg Asp Ala Asp
                165                 170                 175

Arg Leu Pro Ala Ala Gln Ala Ala Ala Ala Ala Thr Ala Asn Pro Gly
            180                 185                 190

Ser Gln Thr His 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 Ile Ser Asn Ser Thr Ile Asn Met Phe Gly Asp Pro Lys Pro Tyr
```

225 230 235 240

Asn Pro Ser Ser Asn Asp Trp Lys Glu Glu Tyr Glu Ala Cys Arg Ile
245 250 255

Trp Asp Arg Pro Pro Arg Gly Asn Leu Arg Asp Thr Pro Phe Tyr Pro
260 265 270

Trp Ala Pro Lys Glu Asn Gln Tyr Arg Val Asn Phe Lys Leu Gly Phe
275 280 285

Gln

<210> 858
<211> 196
<212> PRT
<213> Torque teno virus

<400> 858
Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1 5 10 15

Tyr Ser Ser Ile Leu Ser Ser His Ala Ala Met Pro Gln Glu Thr Gln
20 25 30

Glu Gly His Arg Gly Val Pro Leu Gln Pro Arg Arg Gly Ala Lys Arg
35 40 45

Lys Leu Thr Phe Pro Pro Ser Gln Ala Pro Gln Thr Ser Pro Pro Val
50 55 60

Gly Arg Leu Ala Ala Gly Gly Lys Arg Val Ala Lys Leu Arg Gly Arg
65 70 75 80

Asp Ala Asp Arg Leu Pro Ala Ala Gln Ala Ala Ala Ala Ala Thr Ala
85 90 95

Asn Pro Gly Ser Gln Thr His Thr Pro Val Gln Pro Ser Pro Lys Asn
100 105 110

Pro Thr Lys Ser Arg Tyr Gln Pro Tyr Leu Val Thr Lys Gly Gly Gly

```
        115                 120                 125

Ser Ser Ile Leu Ile Ser Asn Ser Thr Ile Asn Met Phe Gly Asp Pro
    130                 135                 140

Lys Pro Tyr Asn Pro Ser Ser Asn Asp Trp Lys Glu Glu Tyr Glu Ala
145                 150                 155                 160

Cys Arg Ile Trp Asp Arg Pro Pro Arg Gly Asn Leu Arg Asp Thr Pro
                165                 170                 175

Phe Tyr Pro Trp Ala Pro Lys Glu Asn Gln Tyr Arg Val Asn Phe Lys
            180                 185                 190

Leu Gly Phe Gln
        195


<210> 859
<211> 766
<212> PRT
<213> Torque teno virus

<400> 859
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

Val Arg Arg Arg Gly Arg Arg Arg Asn Val Arg Arg Arg Arg Arg Gly
        35                  40                  45

Gly Arg Trp Arg Arg Arg Tyr Arg Arg Trp Lys Arg Lys Gly Arg Arg
    50                  55                  60

Arg Lys Lys Ala Lys Ile Ile Ile Arg Gln Trp Gln Pro Asn Tyr Arg
65                  70                  75                  80

Arg Arg Cys Asn Ile Val Gly Tyr Ile Pro Val Leu Ile Cys Gly Glu
                85                  90                  95

Asn Thr Val Ser Arg Asn Tyr Ala Thr His Ser Asp Asp Thr Asn Tyr
```

```
            100                 105                 110

Pro Gly Pro Phe Gly Gly Gly Met Thr Thr Asp Lys Phe Thr Leu Arg
        115                 120                 125

Ile Leu Tyr Asp Glu Tyr Lys Arg Phe Met Asn Tyr Trp Thr Ala Ser
    130                 135                 140

Asn Glu Asp Leu Asp Leu Cys Arg Tyr Leu Gly Val Asn Leu Tyr Phe
145                 150                 155                 160

Phe Arg His Pro Glu Val Asp Phe Ile Ile Lys Ile Asn Thr Met Pro
                165                 170                 175

Pro Phe Leu Asp Thr Glu Leu Thr Ala Pro Ser Ile His Pro Gly Met
            180                 185                 190

Leu Ala Leu Asp Lys Arg Ala Arg Trp Ile Pro Ser Leu Lys Ser Arg
        195                 200                 205

Pro Gly Lys Lys His Tyr Ile Lys Ile Arg Val Gly Ala Pro Lys Met
    210                 215                 220

Phe Thr Asp Lys Trp Tyr Pro Gln Thr Asp Leu Cys Asp Met Val Leu
225                 230                 235                 240

Leu Thr Val Tyr Ala Thr Ala Ala Asp Met Gln Tyr Pro Phe Gly Ser
                245                 250                 255

Pro Leu Thr Asp Ser Val Val Val Asn Phe Gln Val Leu Gln Ser Met
            260                 265                 270

Tyr Asp Glu Thr Ile Ser Ile Leu Pro Asp Gln Lys Glu Lys Arg Ile
        275                 280                 285

Thr Leu Leu Thr Ser Ile Ala Phe Tyr Asn Thr Thr Gln Thr Ile Ala
    290                 295                 300

Gln Leu Lys Pro Phe Ile Asp Ala Gly Asn Met Thr Ser Thr Thr Thr
305                 310                 315                 320
```

```
Ala Thr Thr Trp Gly Ser Tyr Ile Asn Thr Thr Lys Phe Asn Thr Ala
                325                 330                 335

Ala Thr Thr Thr Tyr Thr Tyr Pro Gly Ser Thr Thr Thr Thr Val Thr
            340                 345                 350

Met Leu Thr Cys Asn Asp Ser Trp Tyr Arg Gly Thr Val Tyr Asn Asp
        355                 360                 365

Gln Ile Lys Asn Leu Pro Lys Glu Ala Ala Gln Leu Tyr Leu Lys Ala
    370                 375                 380

Thr Lys Thr Leu Leu Gly Asn Thr Phe Thr Asn Asp Asp His Thr Leu
385                 390                 395                 400

Glu Tyr His Gly Gly Leu Tyr Ser Ser Ile Trp Leu Ser Pro Gly Arg
                405                 410                 415

Ser Tyr Phe Glu Thr Pro Gly Ala Tyr Thr Asp Ile Lys Tyr Asn Pro
            420                 425                 430

Phe Thr Asp Arg Gly Glu Gly Asn Met Leu Trp Ile Asp Trp Leu Ser
        435                 440                 445

Lys Lys Asn Met Asn Tyr Asp Lys Leu Gln Ser Lys Cys Leu Ile Ser
    450                 455                 460

Asp Leu Pro Leu Trp Ala Ala Ala Tyr Gly Tyr Leu Glu Phe Cys Ala
465                 470                 475                 480

Lys Ser Thr Gly Asp Gln Asn Ile His Met Asn Ala Arg Leu Leu Ile
                485                 490                 495

Arg Ser Pro Phe Thr Asp Pro Gln Leu Leu Val His Thr Asn Pro Thr
            500                 505                 510

Lys Gly Phe Val Pro Tyr Ser Leu Asn Phe Gly Asn Gly Lys Met Pro
        515                 520                 525

Gly Gly Ser Ser Asn Val Pro Ile Arg Met Arg Ala Lys Trp Tyr Pro
```

```
    530                 535                 540

Thr Leu Phe His Gln Gln Glu Val Leu Glu Ala Leu Ala Gln Ser Gly
545                 550                 555                 560

Pro Phe Ala Tyr His Ser Asp Ile Lys Lys Val Ser Leu Gly Met Lys
                565                 570                 575

Tyr Arg Phe Lys Trp Ile Trp Gly Gly Asn Pro Val Arg Gln Gln Val
            580                 585                 590

Val Arg Asn Pro Cys Lys Asp Ser His Ser Ser Val Asn Arg Val Pro
        595                 600                 605

Arg Ser Leu Gln Ile Val Asp Pro Lys Tyr Asn Ser Pro Glu Leu Thr
    610                 615                 620

Phe His Thr Trp Asp Phe Arg Arg Gly Leu Phe Gly Gln Lys Ala Ile
625                 630                 635                 640

Glu Arg Met Gln Gln Gln Pro Thr Thr Thr Asp Ile Phe Ser Ala Gly
                645                 650                 655

Arg Lys Arg Pro Arg Arg Asp Thr Glu Val Tyr His Ser Ser Gln Glu
            660                 665                 670

Gly Glu Gln Lys Glu Ser Leu Leu Phe Pro Pro Val Lys Leu Leu Arg
        675                 680                 685

Arg Val Pro Pro Trp Glu Asp Ser Gln Gln Glu Glu Ser Gly Ser Gln
    690                 695                 700

Ser Ser Glu Glu Glu Thr Gln Thr Val Ser Gln Gln Leu Lys Gln Gln
705                 710                 715                 720

Leu Gln Gln Gln Arg Ile Leu Gly Val Lys Leu Ile Leu Leu Phe Asn
                725                 730                 735

Gln Val Gln Lys Ile Gln Gln Asn Gln Asp Ile Asn Pro Thr Leu Leu
            740                 745                 750
```

```
Pro Arg Gly Gly Asp Leu Ala Ser Leu Phe Gln Ile Ala Pro
        755                 760                 765


<210> 860
<211> 216
<212> PRT
<213> Torque teno virus

<400> 860
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


Val Arg Arg Arg Gly Arg Arg Arg Asn Val Val Arg Asn Pro Cys Lys
        35                  40                  45


Asp Ser His Ser Ser Val Asn Arg Val Pro Arg Ser Leu Gln Ile Val
    50                  55                  60


Asp Pro Lys Tyr Asn Ser Pro Glu Leu Thr Phe His Thr Trp Asp Phe
65                  70                  75                  80


Arg Arg Gly Leu Phe Gly Gln Lys Ala Ile Glu Arg Met Gln Gln Gln
                85                  90                  95


Pro Thr Thr Thr Asp Ile Phe Ser Ala Gly Arg Lys Arg Pro Arg Arg
            100                 105                 110


Asp Thr Glu Val Tyr His Ser Ser Gln Glu Gly Glu Gln Lys Glu Ser
        115                 120                 125


Leu Leu Phe Pro Pro Val Lys Leu Leu Arg Arg Val Pro Pro Trp Glu
    130                 135                 140


Asp Ser Gln Gln Glu Glu Ser Gly Ser Gln Ser Ser Glu Glu Glu Thr
145                 150                 155                 160


Gln Thr Val Ser Gln Gln Leu Lys Gln Gln Leu Gln Gln Gln Arg Ile
                165                 170                 175
```

Leu Gly Val Lys Leu Ile 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

Ala Ser Leu Phe Gln Ile Ala Pro
210 215

<210> 861
<211> 143
<212> PRT
<213> Torque teno virus

<400> 861
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

Val Arg Arg Arg Gly Arg Arg Arg Asn Ala Ala Arg Asp Pro Gly Gly
35 40 45

Thr Pro Arg Cys Thr Thr Pro Ala Lys Lys Gly Ser Lys Lys Lys Ala
50 55 60

Tyr Phe Ser Pro Gln Ser Ser Ser Ser Asp Glu Ser Pro Arg Gly Lys
65 70 75 80

Thr Arg Ser Arg Arg Lys Ala Gly Arg Lys Ala Gln Arg Lys Arg Arg
85 90 95

Arg Pro Ser Pro Ser Ser Ser Ser Ser Ser Cys Ser Asn Ser Glu Ser
100 105 110

Trp Glu Ser Asn Ser Tyr 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> 862
<211> 3553
<212> DNA
<213> Torque teno virus

<220>
<221> modified_base
<222> (1842)..(1842)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (3125)..(3176)
<223> a, c, t, g, unknown or other

<400> 862
atacctcatc atataaagcg gcgcacttcc gaatggctga gttttccacg cccgtccgca 60

gcgagatcgc gacggaggag cgatcgagcg tcccgagggc gggtgccgga ggtgagttta 120

cacaccgcag tcaaggggca attcgggctc gggactggcc gggctatggg gcaagactct 180

taaaaaagcc atgtttctcg gtaaacttta cagaaagaaa agggcactgt cactgctacg 240

cgtgcgagct ccagaggcga aaccacctgc tatgagttgg agacccccgg tgcacaaccc 300

caatgggatc gagagaaacc tgtgggaggc attctttcgc atgcatgctt cagcttgtgg 360

ttgtggcgat cttgttggcc atcttactgt actggctggt cggtatggtg ctcctcctcg 420

tcccccggcc cccggcgctc ccagaccacc gctgatacgc cagctggccc ttccggcgcc 480

ccccgccgat cctcaacagg ctaacccaca atggcctggt ggggacggtg gagaagatgg 540

cgctggaggc cccgccgctg gcggcgccgt cgcagacgcc gagtaccaag aagacgagct 600

caacgccctg ttcgacgccg tcgagcaaga agagtaagga ggaggcgatg ggggaggcgg 660

aggtggagac gggggtacag acgcagactg agactaagac gcagacgcag acgaaagcga 720

aagatagtac taactcagtg gaatcccgcc aaagtgcgga ggtgtactat taagggagtt 780

ctgcccatga tcctgtgcgg ggccgggcgc tcggggttta actacggact gcacagcgac 840

gactacactg tacagaagcc ccttggccag aacccccacg ggggcggcat gagtacagtg 900

acttttagcc tacaggtgct ctatgaccag taccagaggt ttatgaacaa gtggtcgtac 960

tccaacgacc agctagacct cgccaggtac tttggctgca ccttctggtt ctacagacac 1020

ccagaggtgg actttgtagc tcagtttgac aacgttcccc ccatgaaaat ggacgagaac 1080

acagccccaa acactcatcc ctctttctta ctacagaaca aacacaaggt taaaattccc 1140

agctttaaaa caaagccttt tggtaaaaaa agagttagag ttacagtagg gccccccaaa 1200

ctgtttgaag ataagtggta cagccaacat gacttgtgta aggtgcccct agtcagttgg 1260

cggttaaccg cagctgactt caggtttccg ttctgctcac cacaaactga caacccttgc 1320

tacaccttcc aggtattgca tgaagagtat tacccagtaa taggcacttc tgctttagaa 1380

aacggcagta actacaatag ctcagctata acagccttag aaaaattctt atatgaaaaa 1440

tgcacacact atcaaacatt tgccacagac accagactta atcctcagcg accagtgtca 1500

tctacaaatg caaacaaaac atacaccccc tcaggctccc aagaaacaat agtgtggggg 1560

cagtcagatt ttaatttatt taaaaagcac acagacagca actatggcta ctgcacctac 1620

tgtcctacca atgacttagc tacaaaaatt aaaaagtaca gagacaaaag attcgactgg 1680

ctaacaaaca tgccagtaac aaacacctgc cacataaatg ccaccttcgc ccgaggcaaa 1740

attaaagaat gggagtacca cctagggtgg ttctcaaaca tctttatagg caacctgaga 1800

cacaacctag cattccgggc cgcatacata gacatcacct anacagacaa gggagaaggc 1860

aacattatct ggttccagta cctcactaaa cccaccacag agtacataga agcccaagca 1920

aagtgctcca tcacaaacat acccctgtat gctgcttttt atggctacga agactacctc 1980

cagagaacac taggccccta ccaagatgta gaaaccctag gtataatctg tgttaaatgt 2040

ccctacacag atccccctct agttcacaag tctacagata aaaagaactg gggctacgtg 2100

ttctacgacg tgcactttgg caacggaaag accccagagg gactgggcca ggtgcaccct 2160

tactggatgc agaggtggag accctacgta cagtttcaga aagacactat gaacaaaata 2220

gccaggacgg gaccgttcag ctacagagac gagacgcctt ccatcaccct gaccgccggg 2280

tacaagtttc attttaactg ggggggcgac tctatatttc cacagattat taaaaacccc 2340

tgcccagaca gcggggtacg accttcatcc agtagagagc gtcgctcagt acaagtcgtt 2400

agcccgctca caatggggcc agagtacata ttccaccggt gggactggcg acgggggttc 2460

tttaatcaaa aagctctcaa aagaatgctt gaaaaatcaa ttaatgatgg agagtatcca 2520

acaggcccaa aggtccctcg atggtttccc ccactcgaca accaagagca agaaggcgcc 2580

tcaggttcag aggagacaag gtcgcagtcc tcgcaagaag aagccgctca agaagccctc 2640

caagaagtcc aagaggcgtc gctacagcag cacctcctcc agcagtaccg agagcagcga 2700

cggatcggaa agcaactcca actcgtcatg ctgcagctca ccaagacgca gagcaacctg 2760

cacataaacc cccgtgttct tggccatgca taaataaagt ctacatgttt ccccccgaca 2820

agcccatgcc catacacggg taccacgggt gggagacgga gtaccaggcc tgcaaggcct 2880

tcaacaggcc ccccagaaac tacctttcag acaaacccat ctacccttgg ctccctcgcc 2940

ccgaacccga aataatagtg agctttaggt tcggtttcaa ataaacaagg ccgcaaataa 3000

acaaggccgt gggagtttca ctggtcggtg tctacctctt aaggtcacta agcactccga 3060

gcgttagcga ggagtgcgac ccttcccccc ggtgccacgc cctcggcggc cgcgcgctac 3120

gcctnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnntgaa 3180

tcagtaacga aagtgagtgg ggccagactt cgccataagg cctttatctt cttgccattg 3240

gtccgtgtgg ggagtcgcca taggcttcgg gctcggtttt aggccttccg gactacaaaa 3300

accgccattt tagtgacgtc acggcggcca ttttaagtaa gcatggcggg cggtgacgta 3360

caagttgaaa ggtcaccgcg cttccgtgtt tactcaaaat ggtggccaac tgcttccggg 3420

tcaaaggtcg gcggccacgt cataagtcac gtgggagggc tgcgtcacaa acacggaagt 3480

ggctgtccca cgtgacttgt cacgtgattg ctacgtcacg gccgccattt tagttcacaa 3540

aatggcggac ttc 3553

<210> 863
<211> 121
<212> PRT
<213> Torque teno virus

<400> 863
Met Ser Trp Arg Pro Pro Val His Asn Pro Asn Gly Ile Glu Arg Asn
1 5 10 15

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

Asp Leu Val Gly His Leu Thr Val Leu Ala Gly Arg Tyr Gly Ala Pro
35 40 45

Pro Arg Pro Pro Ala Pro Gly Ala Pro Arg Pro Pro Leu Ile Arg Gln
50 55 60

Leu Ala Leu Pro Ala Pro Pro Ala Asp Pro Gln Gln Ala Asn Pro Gln

65 70 75 80

Trp Pro Gly Gly Asp Gly Gly Glu Asp Gly Ala Gly Gly Pro Ala Ala
85 90 95

Gly Gly Ala Val Ala Asp Ala Glu Tyr Gln Glu Asp Glu Leu Asn Ala
100 105 110

Leu Phe Asp Ala Val Glu Gln Glu Glu
115 120

<210> 864
<211> 267
<212> PRT
<213> Torque teno virus

<400> 864
Met Ser Trp Arg Pro Pro Val His Asn Pro Asn Gly Ile Glu Arg Asn
1 5 10 15

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

Asp Leu Val Gly His Leu Thr Val Leu Ala Gly Arg Tyr Gly Ala Pro
35 40 45

Pro Arg Pro Pro Ala Pro Gly Ala Pro Arg Pro Pro Leu Ile Arg Gln
50 55 60

Leu Ala Leu Pro Ala Pro Pro Ala Asp Pro Gln Gln Ala Asn Pro Gln
65 70 75 80

Trp Pro Gly Gly Asp Gly Gly Glu Asp Gly Ala Gly Gly Pro Ala Ala
85 90 95

Gly Gly Ala Val Ala Asp Ala Glu Tyr Gln Glu Asp Glu Leu Asn Ala
100 105 110

Leu Phe Asp Ala Val Glu Gln Glu Glu Leu Leu Lys Thr Pro Ala Gln
115 120 125

Thr Ala Gly Tyr Asp Leu His Pro Val Glu Ser Val Ala Gln Tyr Lys

```
    130                 135                 140

Ser Leu Ala Arg Ser Gln Trp Gly Gln Ser Thr Tyr Ser Thr Gly Gly
145                 150                 155                 160

Thr Gly Asp Gly Gly Ser Leu Ile Lys Lys Leu Ser Lys Glu Cys Leu
                165                 170                 175

Lys Asn Gln Leu Met Met Glu Ser Ile Gln Gln Ala Gln Arg Ser Leu
            180                 185                 190

Asp Gly Phe Pro His Ser Thr Thr Lys Ser Lys Lys Ala Pro Gln Val
        195                 200                 205

Gln Arg Arg Gln Gly Arg Ser Pro Arg Lys Lys Lys Pro Leu Lys Lys
    210                 215                 220

Pro Ser Lys Lys Ser Lys Arg Arg Arg Tyr Ser Ser Thr Ser Ser Ser
225                 230                 235                 240

Ser Thr Glu Ser Ser Asp Gly Ser Glu Ser Asn Ser Asn Ser Ser Cys
                245                 250                 255

Cys Ser Ser Pro Arg Arg Arg Ala Thr Cys Thr
            260                 265


<210> 865
<211> 273
<212> PRT
<213> Torque teno virus

<400> 865
Met Ser Trp Arg Pro Pro Val His Asn Pro Asn Gly Ile Glu Arg Asn
1               5                   10                  15

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

Asp Leu Val Gly His Leu Thr Val Leu Ala Gly Arg Tyr Gly Ala Pro
        35                  40                  45

Pro Arg Pro Pro Ala Pro Gly Ala Pro Arg Pro Pro Leu Ile Arg Gln
```

```
    50                  55                  60

Leu Ala Leu Pro Ala Pro Pro Ala Asp Pro Gln Gln Ala Asn Pro Gln
65                  70                  75                  80

Trp Pro Gly Gly Asp Gly Gly Glu Asp Gly Ala Gly Gly Pro Ala Ala
                85                  90                  95

Gly Gly Ala Val Ala Asp Ala Glu Tyr Gln Glu Asp Glu Leu Asn Ala
            100                 105                 110

Leu Phe Asp Ala Val Glu Gln Glu Glu Pro Lys Gly Pro Ser Met Val
        115                 120                 125

Ser Pro Thr Arg Gln Pro Arg Ala Arg Arg Arg Leu Arg Phe Arg Gly
    130                 135                 140

Asp Lys Val Ala Val Leu Ala Arg Arg Ser Arg Ser Arg Ser Pro Pro
145                 150                 155                 160

Arg Ser Pro Arg Gly Val Ala Thr Ala Ala Pro Pro Pro Ala Val Pro
                165                 170                 175

Arg Ala Ala Thr Asp Arg Lys Ala Thr Pro Thr Arg His Ala Ala Ala
            180                 185                 190

His Gln Asp Ala Glu Gln Pro Ala His Lys Pro Pro Cys Ser Trp Pro
        195                 200                 205

Cys Ile Asn Lys Val Tyr Met Phe Pro Pro Asp Lys Pro Met Pro Ile
    210                 215                 220

His Gly Tyr His Gly Trp Glu Thr Glu Tyr Gln Ala Cys Lys Ala Phe
225                 230                 235                 240

Asn Arg Pro Pro Arg Asn Tyr Leu Ser Asp Lys Pro Ile Tyr Pro Trp
                245                 250                 255

Leu Pro Arg Pro Glu Pro Glu Ile Ile Val Ser Phe Arg Phe Gly Phe
            260                 265                 270
```

Lys

<210> 866
<211> 196
<212> PRT
<213> Torque teno virus

<400> 866
Met Ser Trp Arg Pro Pro Val His Asn Pro Asn Gly Ile Glu Arg Asn
1 5 10 15

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

Asp Leu Val Gly His Leu Thr Val Leu Ala Gly Arg Pro Lys Gly Pro
        35 40 45

Ser Met Val Ser Pro Thr Arg Gln Pro Arg Ala Arg Arg Arg Leu Arg
    50 55 60

Phe Arg Gly Asp Lys Val Ala Val Leu Ala Arg Arg Ser Arg Ser Arg
65 70 75 80

Ser Pro Pro Arg Ser Pro Arg Gly Val Ala Thr Ala Ala Pro Pro Pro
                85 90 95

Ala Val Pro Arg Ala Ala Thr Asp Arg Lys Ala Thr Pro Thr Arg His
            100 105 110

Ala Ala Ala His Gln Asp Ala Glu Gln Pro Ala His Lys Pro Pro Cys
        115 120 125

Ser Trp Pro Cys Ile Asn Lys Val Tyr Met Phe Pro Pro Asp Lys Pro
    130 135 140

Met Pro Ile His Gly Tyr His Gly Trp Glu Thr Glu Tyr Gln Ala Cys
145 150 155 160

Lys Ala Phe Asn Arg Pro Pro Arg Asn Tyr Leu Ser Asp Lys Pro Ile
                165 170 175

```
Tyr Pro Trp Leu Pro Arg Pro Glu Pro Glu Ile Ile Val Ser Phe Arg
            180                 185                 190


Phe Gly Phe Lys
        195


<210> 867
<211> 760
<212> PRT
<213> Torque teno virus


<220>
<221> MOD_RES
<222> (444)..(444)
<223> Any amino acid

<400> 867
Met Ala Trp Trp Gly Arg Trp Arg Arg Trp Arg Trp Arg Pro Arg Arg
1               5                   10                  15


Trp Arg Arg Arg Arg Arg Arg Arg Val Pro Arg Arg Arg Ala Gln Arg
            20                  25                  30


Pro Val Arg Arg Arg Arg Ala Arg Arg Val Arg Arg Arg Arg Trp Gly
        35                  40                  45


Arg Arg Arg Trp Arg Arg Gly Tyr Arg Arg Arg Leu Arg Leu Arg Arg
    50                  55                  60


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


Lys Val Arg Arg Cys Thr Ile Lys Gly Val Leu Pro Met Ile Leu Cys
                85                  90                  95


Gly Ala Gly Arg Ser Gly Phe Asn Tyr Gly Leu His Ser Asp Asp Tyr
            100                 105                 110


Thr Val Gln Lys Pro Leu Gly Gln Asn Pro His Gly Gly Gly Met Ser
        115                 120                 125


Thr Val Thr Phe Ser Leu Gln Val Leu Tyr Asp Gln Tyr Gln Arg Phe
```

```
    130                 135                 140

Met Asn Lys Trp Ser Tyr Ser Asn Asp Gln Leu Asp Leu Ala Arg Tyr
145                 150                 155                 160

Phe Gly Cys Thr Phe Trp Phe Tyr Arg His Pro Glu Val Asp Phe Val
                165                 170                 175

Ala Gln Phe Asp Asn Val Pro Pro Met Lys Met Asp Glu Asn Thr Ala
            180                 185                 190

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

Ile Pro Ser Phe Lys Thr Lys Pro Phe Gly Lys Lys Arg Val Arg Val
    210                 215                 220

Thr Val Gly Pro Pro Lys Leu Phe Glu Asp Lys Trp Tyr Ser Gln His
225                 230                 235                 240

Asp Leu Cys Lys Val Pro Leu Val Ser Trp Arg Leu Thr Ala Ala Asp
                245                 250                 255

Phe Arg Phe Pro Phe Cys Ser Pro Gln Thr Asp Asn Pro Cys Tyr Thr
            260                 265                 270

Phe Gln Val Leu His Glu Glu Tyr Tyr Pro Val Ile Gly Thr Ser Ala
        275                 280                 285

Leu Glu Asn Gly Ser Asn Tyr Asn Ser Ser Ala Ile Thr Ala Leu Glu
    290                 295                 300

Lys Phe Leu Tyr Glu Lys Cys Thr His Tyr Gln Thr Phe Ala Thr Asp
305                 310                 315                 320

Thr Arg Leu Asn Pro Gln Arg Pro Val Ser Ser Thr Asn Ala Asn Lys
                325                 330                 335

Thr Tyr Thr Pro Ser Gly Ser Gln Glu Thr Ile Val Trp Gly Gln Ser
            340                 345                 350
```

```
Asp Phe Asn Leu Phe Lys Lys His Thr Asp Ser Asn Tyr Gly Tyr Cys
        355                 360                 365

Thr Tyr Cys Pro Thr Asn Asp Leu Ala Thr Lys Ile Lys Lys Tyr Arg
    370                 375                 380

Asp Lys Arg Phe Asp Trp Leu Thr Asn Met Pro Val Thr Asn Thr Cys
385                 390                 395                 400

His Ile Asn Ala Thr Phe Ala Arg Gly Lys Ile Lys Glu Trp Glu Tyr
                405                 410                 415

His Leu Gly Trp Phe Ser Asn Ile Phe Ile Gly Asn Leu Arg His Asn
            420                 425                 430

Leu Ala Phe Arg Ala Ala Tyr Ile Asp Ile Thr Xaa Thr Asp Lys Gly
        435                 440                 445

Glu Gly Asn Ile Ile Trp Phe Gln Tyr Leu Thr Lys Pro Thr Thr Glu
    450                 455                 460

Tyr Ile Glu Ala Gln Ala Lys Cys Ser Ile Thr Asn Ile Pro Leu Tyr
465                 470                 475                 480

Ala Ala Phe Tyr Gly Tyr Glu Asp Tyr Leu Gln Arg Thr Leu Gly Pro
                485                 490                 495

Tyr Gln Asp Val Glu Thr Leu Gly Ile Ile Cys Val Lys Cys Pro Tyr
            500                 505                 510

Thr Asp Pro Pro Leu Val His Lys Ser Thr Asp Lys Lys Asn Trp Gly
        515                 520                 525

Tyr Val Phe Tyr Asp Val His Phe Gly Asn Gly Lys Thr Pro Glu Gly
    530                 535                 540

Leu Gly Gln Val His Pro Tyr Trp Met Gln Arg Trp Arg Pro Tyr Val
545                 550                 555                 560

Gln Phe Gln Lys Asp Thr Met Asn Lys Ile Ala Arg Thr Gly Pro Phe
```

```
                565                 570                 575

Ser Tyr Arg Asp Glu Thr Pro Ser Ile Thr Leu Thr Ala Gly Tyr Lys
            580                 585                 590

Phe His Phe Asn Trp Gly Gly Asp Ser Ile Phe Pro Gln Ile Ile Lys
        595                 600                 605

Asn Pro Cys Pro Asp Ser Gly Val Arg Pro Ser Ser Ser Arg Glu Arg
    610                 615                 620

Arg Ser Val Gln Val Val Ser Pro Leu Thr Met Gly Pro Glu Tyr Ile
625                 630                 635                 640

Phe His Arg Trp Asp Trp Arg Arg Gly Phe Phe Asn Gln Lys Ala Leu
                645                 650                 655

Lys Arg Met Leu Glu Lys Ser Ile Asn Asp Gly Glu Tyr Pro Thr Gly
            660                 665                 670

Pro Lys Val Pro Arg Trp Phe Pro Pro Leu Asp Asn Gln Glu Gln Glu
        675                 680                 685

Gly Ala Ser Gly Ser Glu Glu Thr Arg Ser Gln Ser Ser Gln Glu Glu
    690                 695                 700

Ala Ala Gln Glu Ala Leu Gln Glu Val Gln Glu Ala Ser Leu Gln Gln
705                 710                 715                 720

His Leu Leu Gln Gln Tyr Arg Glu Gln Arg Arg Ile Gly Lys Gln Leu
                725                 730                 735

Gln Leu Val Met Leu Gln Leu Thr Lys Thr Gln Ser Asn Leu His Ile
            740                 745                 750

Asn Pro Arg Val Leu Gly His Ala
        755                 760

<210> 868
<211> 196
<212> PRT
```

<213> Torque teno virus

<400> 868

```
Met Ala Trp Trp Gly Arg Trp Arg Arg Trp Arg Trp Arg Pro Arg Arg
1               5                  10                  15

Trp Arg Arg Arg Arg Arg Arg Arg Val Pro Arg Arg Arg Ala Gln Arg
            20                  25                  30

Pro Val Arg Arg Arg Arg Ala Arg Arg Ile Ile Lys Asn Pro Cys Pro
        35                  40                  45

Asp Ser Gly Val Arg Pro Ser Ser Ser Arg Glu Arg Arg Ser Val Gln
    50                  55                  60

Val Val Ser Pro Leu Thr Met Gly Pro Glu Tyr Ile Phe His Arg Trp
65                  70                  75                  80

Asp Trp Arg Arg Gly Phe Phe Asn Gln Lys Ala Leu Lys Arg Met Leu
                85                  90                  95

Glu Lys Ser Ile Asn Asp Gly Glu Tyr Pro Thr Gly Pro Lys Val Pro
            100                 105                 110

Arg Trp Phe Pro Pro Leu Asp Asn Gln Glu Gln Glu Gly Ala Ser Gly
        115                 120                 125

Ser Glu Glu Thr Arg Ser Gln Ser Ser Gln Glu Glu Ala Ala Gln Glu
    130                 135                 140

Ala Leu Gln Glu Val Gln Glu Ala Ser Leu Gln Gln His Leu Leu Gln
145                 150                 155                 160

Gln Tyr Arg Glu Gln Arg Arg Ile Gly Lys Gln Leu Gln Leu Val Met
                165                 170                 175

Leu Gln Leu Thr Lys Thr Gln Ser Asn Leu His Ile Asn Pro Arg Val
            180                 185                 190

Leu Gly His Ala
        195
```

<210> 869
<211> 121
<212> PRT
<213> Torque teno virus

<400> 869
Met Ala Trp Trp Gly Arg Trp Arg Arg Trp Arg Trp Arg Pro Arg Arg
1   5   10   15

Trp Arg Arg Arg Arg Arg Arg Arg Val Pro Arg Arg Arg Ala Gln Arg
            20   25   30

Pro Val Arg Arg Arg Arg Ala Arg Arg Ala Gln Arg Ser Leu Asp Gly
        35   40   45

Phe Pro His Ser Thr Thr Lys Ser Lys Lys Ala Pro Gln Val Gln Arg
    50   55   60

Arg Gln Gly Arg Ser Pro Arg Lys Lys Lys Pro Leu Lys Lys Pro Ser
65   70   75   80

Lys Lys Ser Lys Arg Arg Arg Tyr Ser Ser Thr Ser Ser Ser Ser Thr
                85   90   95

Glu Ser Ser Asp Gly Ser Glu Ser Asn Ser Asn Ser Ser Cys Cys Ser
            100   105   110

Ser Pro Arg Arg Arg Ala Thr Cys Thr
        115   120

<210> 870
<211> 3896
<212> DNA
<213> Torque teno virus

<220>
<221> modified_base
<222> (1276)..(1276)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (3198)..(3655)
<223> a, c, t, g, unknown or other

<400> 870
taaacttcct cttttaatag gaaaccacaa aatttgcatt gccgaccaca aacgcatatg 60

caaatttact tccccaaaaa ctcaaccaca aaatttgcat tgccgcccac aaacgtctac 120

tttaaccaca tcctctaaca tgttagaaac tccacccaac tacttcatta gtatacagca 180

tcacaaggga ggagccaaac aactatataa ccaagtgtac ttccgaatgg ctgagtttat 240

gccgccagac ggagacggga tcgcgacgga ggagcgatcg agcgtcccga gggcgggtgc 300

cggaggtgag tttacacacc gcagtcaagg ggcaattcgg gctcgggact ggccgggcta 360

tgggcaaggc tcttaaaaaa gccatgtttc tcggtcgacc ttacagaaag aaaagggcac 420

tgtcactgct acgcgtgcga gctccagagg cgaaaccacc tgctatgagc tggaggcccc 480

cggtgcacaa ccctaatggg atccagagaa acctgtggga ggcattcttt cgcatgcatg 540

ctgcagcttg tggttgtggc gatcttgttg gccatattac tgtactggct ggtcggtatg 600

gtgctcctcc tcgtcccccg gcccccgggg ctcccagacc accgctgata cgccagctgg 660

cccttccggc gcccccgcc gatcctcaac aggctaaccc acaatggcct ggtggggacg 720

gtggagaaga tggcgctgga ggccccgccg ctggcggcgc cgtcgcagac gccgagtacc 780

aagaagacga gctcaacgcc ctgttcgacg ccgtcgagca agaagagtaa ggaggaggcg 840

atgggggagg cggaggtgga gacgggggta cagacgcaga ctaagactga gacgcagacg 900

cagacgaaag aaaataagac tgactcagtg gaacccagcc aaagtcagga gatgtactat 960

taagggggtg ctacccatga tcttatgcgg cgccggccgc tcggggttta actatggact 1020

gcacagcgac gactacacgg tgcagaaacc cctggggcag aacccccacg ggggcggcat 1080

gagcacagta acttttagcc tacaagtact atttgaccag taccagaggt ttatgaaccg 1140

gtggtcgtac tccaacgacc agctagacct cgccaggtac tttggctgca ccttctactt 1200

ttacagacac cctgaaattg actttgtagc tcagtatgac aatgtacccc caatgaaaat 1260

ggacgagaac acggcnccta acactcaccc ctcttttcta ctacaaaaca aacgcaaaat 1320

taaaatcccc agctttaaaa ccaagccatt tggcagaaaa agagtaaaag taacagtggg 1380

gccccccaaa ctgtttgaag ataaatggta cagccagcat gacttgtgta aggtgcccct 1440

agtcagttgg cggttaaccg catgtgactt caggtttccg ttctgctcac cactaactga 1500

caacccttgc tacaccttcc aggtattgca tgaaaactat tacccagtca taggcacttc 1560

ctctttagaa aacggtacaa actacaataa cactgctata actacccttg agacatggct 1620

atatggaaaa tgcacacact atcaaacatt tgccacagac accagactta atccacagag 1680

acctgtatct tcaagtaatg caaatgaaac ttatactcct agtggttcta aagaatcaat 1740

aatatgggga cagtctgact gggcaaactt taaaaagaac acagacagca actatggcta 1800

ctgttcctac tgcccctcaa atggcactaa cggaacagta gataaaatta aaaaatacag 1860

agaccaaaga tttagatggc ttacagaaat gccagtacct aacacctgtc acatacatgc 1920

caccttcgcc cgaggcacta ttaaatactg ggagtaccac ctaggctggt actcaaacat 1980

atttattggc aacctcagac acaacttagc cttcagacca gcctacatag acattaccta 2040

caatcccatc actgacaaag gagagggcaa cattatctgg ttccagtacc tcactaagcc 2100

caccacagaa tacatagaaa cccaggcaaa atgcaccatt actaacattc ccctttatgc 2160

tgctttctat ggctacgaag actacctcca gagaacacta ggcccctacc aagatgtaga 2220

aaccctaggc ataatctgtg ttaaatgtcc ctacacagat ccccctctag ttcacaaaga 2280

caaaagtaaa accaactggg gctacgtatt ctacgacgcc cactttggca acggaaagac 2340

cccagaggga ctaggccaag tacaccctta ctggatgcag agatggagac cctatgtaca 2400

gtttcaaaaa gacaccatgc acaaaatatc cagaacggga cccttcagct acagagacga 2460

cacgccttcc atcaccctca ctgccgaata caagtttcgt tttaactggg ggggcgactc 2520

tatatttcca cagattatta aaaacccctg cccagacacc ggggttcgac cttcaaccgg 2580

tagagaccgt cgctcagtac aagtcgttag cccgctcaca atgggacccc agtttatatt 2640

ccactcatgg gactggagac gggggttctt taatcaaaaa actctcaaaa gaatgcttga 2700

aaaaccagtt aatgatggag aatatccaac aggcccaaag gtgcctcgat ggtttccccc 2760

actcgacaac caagagcaag aaggcgtctc agatacagag acgacaacct cgcagtcctc 2820

gcaagaagaa gccgctcaag aagccctcca agaagtccaa gaggcgtcgc tacagcagca 2880

cctcctccag cagtaccgag agcagcgaag aatcggaaag caactccaac tcgtcatgct 2940

ccaactcacc aagacgcaga gcaacctgca cataaatccc cgtgtccttg gccatgcata 3000

aataaagtgt acatgtttcc ccccgaaaag ccaatgccca tacacggcta ccacgggtgg 3060

gagacagagt atcaggcctg caaggccttt gacaggcccc ctagaaacta cctatcagac 3120

aaacccatct acccctggct tccccgctcc caaccagaat ttaaagtgag ttttaagctt 3180

ggctgtcaat aaacaagnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3300

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3420

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3480

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3540

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3600

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnngttta 3660

cacaaaatgg tggccaagtc cttccgggtg aaaggtcggc gcctacgtca taagtcacgt 3720

ggggagggct gcgtcacaac caggaagcaa tcctcaccac gtgatttgtc acgtgatcgc 3780

tacgtcacgg ccgccatttt agtttacaaa atggcggact tccttcctct ttttcaaaaa 3840

taacggccct gcggcggcgc gcgcgctgcg cgcgcgcgcc gggggctgcc gcccca 3896

<210> 871
<211> 121
<212> PRT
<213> Torque teno virus

<400> 871
Met Ser Trp Arg Pro Pro Val His Asn Pro Asn Gly Ile Gln Arg Asn
1 5 10 15

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

Asp Leu Val Gly His Ile Thr Val Leu Ala Gly Arg Tyr Gly Ala Pro
35 40 45

Pro Arg Pro Pro Ala Pro Gly Ala Pro Arg Pro Pro Leu Ile Arg Gln
50 55 60

Leu Ala Leu Pro Ala Pro Pro Ala Asp Pro Gln Gln Ala Asn Pro Gln
65 70 75 80

Trp Pro Gly Gly Asp Gly Gly Glu Asp Gly Ala Gly Gly Pro Ala Ala
85 90 95

Gly Gly Ala Val Ala Asp Ala Glu Tyr Gln Glu Asp Glu Leu Asn Ala
100 105 110

Leu Phe Asp Ala Val Glu Gln Glu Glu
115 120

<210> 872
<211> 267
<212> PRT
<213> Torque teno virus

<400> 872
Met Ser Trp Arg Pro Pro Val His Asn Pro Asn Gly Ile Gln Arg Asn
1 5 10 15

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

Asp Leu Val Gly His Ile Thr Val Leu Ala Gly Arg Tyr Gly Ala Pro
35 40 45

Pro Arg Pro Pro Ala Pro Gly Ala Pro Arg Pro Pro Leu Ile Arg Gln
50 55 60

Leu Ala Leu Pro Ala Pro Pro Ala Asp Pro Gln Gln Ala Asn Pro Gln
65 70 75 80

Trp Pro Gly Gly Asp Gly Gly Glu Asp Gly Ala Gly Gly Pro Ala Ala
85 90 95

Gly Gly Ala Val Ala Asp Ala Glu Tyr Gln Glu Asp Glu Leu Asn Ala
100 105 110

Leu Phe Asp Ala Val Glu Gln Glu Glu Leu Leu Lys Thr Pro Ala Gln
115 120 125

Thr Pro Gly Phe Asp Leu Gln Pro Val Glu Thr Val Ala Gln Tyr Lys
130 135 140

Ser Leu Ala Arg Ser Gln Trp Asp Pro Ser Leu Tyr Ser Thr His Gly
145 150 155 160

```
Thr Gly Asp Gly Gly Ser Leu Ile Lys Lys Leu Ser Lys Glu Cys Leu
                165                 170                 175

Lys Asn Gln Leu Met Met Glu Asn Ile Gln Gln Ala Gln Arg Cys Leu
            180                 185                 190

Asp Gly Phe Pro His Ser Thr Thr Lys Ser Lys Lys Ala Ser Gln Ile
        195                 200                 205

Gln Arg Arg Gln Pro Arg Ser Pro Arg Lys Lys Lys Pro Leu Lys Lys
    210                 215                 220

Pro Ser Lys Lys Ser Lys Arg Arg Arg Tyr Ser Ser Thr Ser Ser Ser
225                 230                 235                 240

Ser Thr Glu Ser Ser Glu Glu Ser Glu Ser Asn Ser Asn Ser Ser Cys
                245                 250                 255

Ser Asn Ser Pro Arg Arg Arg Ala Thr Cys Thr
            260                 265
```

<210> 873
<211> 277
<212> PRT
<213> Torque teno virus

```
<400> 873
Met Ser Trp Arg Pro Pro Val His Asn Pro Asn Gly Ile Gln Arg Asn
1               5                   10                  15

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

Asp Leu Val Gly His Ile Thr Val Leu Ala Gly Arg Tyr Gly Ala Pro
        35                  40                  45

Pro Arg Pro Pro Ala Pro Gly Ala Pro Arg Pro Pro Leu Ile Arg Gln
    50                  55                  60

Leu Ala Leu Pro Ala Pro Pro Ala Asp Pro Gln Gln Ala Asn Pro Gln
65                  70                  75                  80
```

```
Trp Pro Gly Gly Asp Gly Gly Glu Asp Gly Ala Gly Gly Pro Ala Ala
                85                  90                  95

Gly Gly Ala Val Ala Asp Ala Glu Tyr Gln Glu Asp Glu Leu Asn Ala
            100                 105                 110

Leu Phe Asp Ala Val Glu Gln Glu Glu Ile Ser Asn Arg Pro Lys Gly
        115                 120                 125

Ala Ser Met Val Ser Pro Thr Arg Gln Pro Arg Ala Arg Arg Arg Leu
    130                 135                 140

Arg Tyr Arg Asp Asp Asn Leu Ala Val Leu Ala Arg Arg Ser Arg Ser
145                 150                 155                 160

Arg Ser Pro Pro Arg Ser Pro Arg Gly Val Ala Thr Ala Ala Pro Pro
                165                 170                 175

Pro Ala Val Pro Arg Ala Ala Lys Asn Arg Lys Ala Thr Pro Thr Arg
            180                 185                 190

His Ala Pro Thr His Gln Asp Ala Glu Gln Pro Ala His Lys Ser Pro
        195                 200                 205

Cys Pro Trp Pro Cys Ile Asn Lys Val Tyr Met Phe Pro Pro Glu Lys
    210                 215                 220

Pro Met Pro Ile His Gly Tyr His Gly Trp Glu Thr Glu Tyr Gln Ala
225                 230                 235                 240

Cys Lys Ala Phe Asp Arg Pro Pro Arg Asn Tyr Leu Ser Asp Lys Pro
                245                 250                 255

Ile Tyr Pro Trp Leu Pro Arg Ser Gln Pro Glu Phe Lys Val Ser Phe
            260                 265                 270

Lys Leu Gly Cys Gln
        275
```

<210> 874

<211> 200
<212> PRT
<213> Torque teno virus

<400> 874
Met Ser Trp Arg Pro Pro Val His Asn Pro Asn Gly Ile Gln Arg Asn
1 5 10 15

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

Asp Leu Val Gly His Ile Thr Val Leu Ala Gly Arg Ile Ser Asn Arg
        35 40 45

Pro Lys Gly Ala Ser Met Val Ser Pro Thr Arg Gln Pro Arg Ala Arg
    50 55 60

Arg Arg Leu Arg Tyr Arg Asp Asp Asn Leu Ala Val Leu Ala Arg Arg
65 70 75 80

Ser Arg Ser Arg Ser Pro Pro Arg Ser Pro Arg Gly Val Ala Thr Ala
                85 90 95

Ala Pro Pro Pro Ala Val Pro Arg Ala Ala Lys Asn Arg Lys Ala Thr
            100 105 110

Pro Thr Arg His Ala Pro Thr His Gln Asp Ala Glu Gln Pro Ala His
        115 120 125

Lys Ser Pro Cys Pro Trp Pro Cys Ile Asn Lys Val Tyr Met Phe Pro
    130 135 140

Pro Glu Lys Pro Met Pro Ile His Gly Tyr His Gly Trp Glu Thr Glu
145 150 155 160

Tyr Gln Ala Cys Lys Ala Phe Asp Arg Pro Pro Arg Asn Tyr Leu Ser
                165 170 175

Asp Lys Pro Ile Tyr Pro Trp Leu Pro Arg Ser Gln Pro Glu Phe Lys
            180 185 190

Val Ser Phe Lys Leu Gly Cys Gln

195 200

<210> 875
<211> 765
<212> PRT
<213> Torque teno virus

<400> 875
Met Ala Trp Trp Gly Arg Trp Arg Arg Trp Arg Trp Arg Pro Arg Arg
1 5 10 15

Trp Arg Arg Arg Arg Arg Arg Arg Val Pro Arg Arg Arg Ala Gln Arg
20 25 30

Pro Val Arg Arg Arg Arg Ala Arg Arg Val Arg Arg Arg Arg Trp Gly
35 40 45

Arg Arg Arg Trp Arg Arg Gly Tyr Arg Arg Arg Leu Arg Leu Arg Arg
50 55 60

Arg Arg Arg Arg Lys Lys Ile Arg Leu Thr Gln Trp Asn Pro Ala Lys
65 70 75 80

Val Arg Arg Cys Thr Ile Lys Gly Val Leu Pro Met Ile Leu Cys Gly
85 90 95

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

Val Gln Lys Pro Leu Gly Gln Asn Pro His Gly Gly Gly Met Ser Thr
115 120 125

Val Thr Phe Ser Leu Gln Val Leu Phe Asp Gln Tyr Gln Arg Phe Met
130 135 140

Asn Arg Trp Ser Tyr Ser Asn Asp Gln Leu Asp Leu Ala Arg Tyr Phe
145 150 155 160

Gly Cys Thr Phe Tyr Phe Tyr Arg His Pro Glu Ile Asp Phe Val Ala
165 170 175

Gln Tyr Asp Asn Val Pro Pro Met Lys Met Asp Glu Asn Thr Ala Pro

```
            180                 185                 190

Asn Thr His Pro Ser Phe Leu Leu Gln Asn Lys Arg Lys Ile Lys Ile
        195                 200                 205

Pro Ser Phe Lys Thr Lys Pro Phe Gly Arg Lys Arg Val Lys Val Thr
    210                 215                 220

Val Gly Pro Pro Lys Leu Phe Glu Asp Lys Trp Tyr Ser Gln His Asp
225                 230                 235                 240

Leu Cys Lys Val Pro Leu Val Ser Trp Arg Leu Thr Ala Cys Asp Phe
                245                 250                 255

Arg Phe Pro Phe Cys Ser Pro Leu Thr Asp Asn Pro Cys Tyr Thr Phe
            260                 265                 270

Gln Val Leu His Glu Asn Tyr Tyr Pro Val Ile Gly Thr Ser Ser Leu
        275                 280                 285

Glu Asn Gly Thr Asn Tyr Asn Asn Thr Ala Ile Thr Thr Leu Glu Thr
    290                 295                 300

Trp Leu Tyr Gly Lys Cys Thr His Tyr Gln Thr Phe Ala Thr Asp Thr
305                 310                 315                 320

Arg Leu Asn Pro Gln Arg Pro Val Ser Ser Ser Asn Ala Asn Glu Thr
                325                 330                 335

Tyr Thr Pro Ser Gly Ser Lys Glu Ser Ile Ile Trp Gly Gln Ser Asp
            340                 345                 350

Trp Ala Asn Phe Lys Lys Asn Thr Asp Ser Asn Tyr Gly Tyr Cys Ser
        355                 360                 365

Tyr Cys Pro Ser Asn Gly Thr Asn Gly Thr Val Asp Lys Ile Lys Lys
    370                 375                 380

Tyr Arg Asp Gln Arg Phe Arg Trp Leu Thr Glu Met Pro Val Pro Asn
385                 390                 395                 400
```

```
Thr Cys His Ile His Ala Thr Phe Ala Arg Gly Thr Ile Lys Tyr Trp
                405                 410                 415

Glu Tyr His Leu Gly Trp Tyr Ser Asn Ile Phe Ile Gly Asn Leu Arg
            420                 425                 430

His Asn Leu Ala Phe Arg Pro Ala Tyr Ile Asp Ile Thr Tyr Asn Pro
        435                 440                 445

Ile Thr Asp Lys Gly Glu Gly Asn Ile Ile Trp Phe Gln Tyr Leu Thr
    450                 455                 460

Lys Pro Thr Thr Glu Tyr Ile Glu Thr Gln Ala Lys Cys Thr Ile Thr
465                 470                 475                 480

Asn Ile Pro Leu Tyr Ala Ala Phe Tyr Gly Tyr Glu Asp Tyr Leu Gln
                485                 490                 495

Arg Thr Leu Gly Pro Tyr Gln Asp Val Glu Thr Leu Gly Ile Ile Cys
            500                 505                 510

Val Lys Cys Pro Tyr Thr Asp Pro Pro Leu Val His Lys Asp Lys Ser
        515                 520                 525

Lys Thr Asn Trp Gly Tyr Val Phe Tyr Asp Ala His Phe Gly Asn Gly
    530                 535                 540

Lys Thr Pro Glu Gly Leu Gly Gln Val His Pro Tyr Trp Met Gln Arg
545                 550                 555                 560

Trp Arg Pro Tyr Val Gln Phe Gln Lys Asp Thr Met His Lys Ile Ser
                565                 570                 575

Arg Thr Gly Pro Phe Ser Tyr Arg Asp Asp Thr Pro Ser Ile Thr Leu
            580                 585                 590

Thr Ala Glu Tyr Lys Phe Arg Phe Asn Trp Gly Gly Asp Ser Ile Phe
        595                 600                 605

Pro Gln Ile Ile Lys Asn Pro Cys Pro Asp Thr Gly Val Arg Pro Ser
```

```
    610                 615                 620

Thr Gly Arg Asp Arg Arg Ser Val Gln Val Val Ser Pro Leu Thr Met
625                 630                 635                 640

Gly Pro Gln Phe Ile Phe His Ser Trp Asp Trp Arg Arg Gly Phe Phe
                645                 650                 655

Asn Gln Lys Thr Leu Lys Arg Met Leu Glu Lys Pro Val Asn Asp Gly
            660                 665                 670

Glu Tyr Pro Thr Gly Pro Lys Val Pro Arg Trp Phe Pro Pro Leu Asp
        675                 680                 685

Asn Gln Glu Gln Glu Gly Val Ser Asp Thr Glu Thr Thr Thr Ser Gln
    690                 695                 700

Ser Ser Gln Glu Glu Ala Ala Gln Glu Ala Leu Gln Glu Val Gln Glu
705                 710                 715                 720

Ala Ser Leu Gln Gln His Leu Leu Gln Gln Tyr Arg Glu Gln Arg Arg
                725                 730                 735

Ile Gly Lys Gln Leu Gln Leu Val Met Leu Gln Leu Thr Lys Thr Gln
            740                 745                 750

Ser Asn Leu His Ile Asn Pro Arg Val Leu Gly His Ala
        755                 760                 765


<210> 876
<211> 196
<212> PRT
<213> Torque teno virus

<400> 876
Met Ala Trp Trp Gly Arg Trp Arg Arg Trp Arg Trp Arg Pro Arg Arg
1               5                   10                  15

Trp Arg Arg Arg Arg Arg Arg Arg Val Pro Arg Arg Arg Ala Gln Arg
            20                  25                  30

Pro Val Arg Arg Arg Arg Ala Arg Arg Ile Ile Lys Asn Pro Cys Pro
```

```
        35                  40                  45

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

Val Val Ser Pro Leu Thr Met Gly Pro Gln Phe Ile Phe His Ser Trp
65                  70                  75                  80

Asp Trp Arg Arg Gly Phe Phe Asn Gln Lys Thr Leu Lys Arg Met Leu
                85                  90                  95

Glu Lys Pro Val Asn Asp Gly Glu Tyr Pro Thr Gly Pro Lys Val Pro
            100                 105                 110

Arg Trp Phe Pro Pro Leu Asp Asn Gln Glu Gln Glu Gly Val Ser Asp
        115                 120                 125

Thr Glu Thr Thr Thr Ser Gln Ser Ser Gln Glu Glu Ala Ala Gln Glu
    130                 135                 140

Ala Leu Gln Glu Val Gln Glu Ala Ser Leu Gln Gln His Leu Leu Gln
145                 150                 155                 160

Gln Tyr Arg Glu Gln Arg Arg Ile Gly Lys Gln Leu Gln Leu Val Met
                165                 170                 175

Leu Gln Leu Thr Lys Thr Gln Ser Asn Leu His Ile Asn Pro Arg Val
            180                 185                 190

Leu Gly His Ala
        195


<210> 877
<211> 125
<212> PRT
<213> Torque teno virus

<400> 877
Met Ala Trp Trp Gly Arg Trp Arg Arg Trp Arg Trp Arg Pro Arg Arg
1               5                   10                  15

Trp Arg Arg Arg Arg Arg Arg Arg Val Pro Arg Arg Arg Ala Gln Arg
```

```
            20                  25                  30

Pro Val Arg Arg Arg Arg Ala Arg Arg Asn Ile Gln Gln Ala Gln Arg
        35                  40                  45

Cys Leu Asp Gly Phe Pro His Ser Thr Thr Lys Ser Lys Lys Ala Ser
    50                  55                  60

Gln Ile Gln Arg Arg Gln Pro Arg Ser Pro Arg Lys Lys Lys Pro Leu
65                  70                  75                  80

Lys Lys Pro Ser Lys Lys Ser Lys Arg Arg Arg Tyr Ser Ser Thr Ser
                85                  90                  95

Ser Ser Ser Thr Glu Ser Ser Glu Glu Ser Glu Ser Asn Ser Asn Ser
            100                 105                 110

Ser Cys Ser Asn Ser Pro Arg Arg Arg Ala Thr Cys Thr
        115                 120                 125
```

<210> 878
<211> 3264
<212> DNA
<213> Torque teno midi virus

<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 gtttttttcaa 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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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
        115

<210> 886
<211> 3176
<212> DNA
<213> Torque teno midi virus

<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 tgggggaggc 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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno virus

<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 ctacccccccg 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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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
<211> 3828
<212> DNA
<213> Torque teno virus

<400> 903
gtgctacgtc actaacctac gtgtccgtct cccataggcc ggacaccgta tacgtcatac 60

acttcctggg catggtctac gtgataatat aagtggctgc acttccgaat ggctgagttt 120

tccacgcccg tccgcagcga ggacgccacg gagggggatc cgcgcgtccc gagggcgggt 180

gccggaggtg agtttacaca ccgcagtcaa ggggcaattc gggctcggga ctggccgggc 240

tatgggcaag gctcttaaaa atgcactttt ctaggtgcag tagaaagaaa aggacattgt 300

cactgctacc actgtaccat tcacagaaag ctaggccatc tgtgacaggt atgtggagac 360

ccccgactcg aaatgcgttc aatattcaac gtgactggtt ctacagttgc tttcactccc 420

acgcttctat gtgcggctgt gctgatttta ttggtcattt caatcatatc gctgctatgc 480

tcggccgtcc ggaagaccag aaccctcctc cgccacccgg ggctctgaga cccctacccg 540

ctctcccggc ctcttccgag gcacccggtg atcgagcgcc atggcctatg ggtggtggcg 600

gaggcgacgg aggcgcccgt ggtggaggag gagatggcgc cgctggagac gccgtcggag 660

accccgcaga cgccgacctc gtcgccgcta tcgacgccgc agaacagtaa ggaggcgcgg 720

caggggggagg tggactagag cacacaggag atggcgccgc aagggaaaac gcagtcgcaa 780

aaaaaagatt attataagac aatggcagcc caactacact cgcagatgca acatagtggg 840

ctacatgcct ctactaatat gtggggaaaa tactgttgct acaaactatg ccacccactc 900

agacgacagc tactaccccg gacccttttgg ggggggaatg actacagaca aatttactct 960

aagaatactg tatgatgagt acaaaaggtt catgaactac tggacctctt caaacgagga 1020

cctagaccta tgtagatacc tgggatgcac tctatatgtg tttagacacc cagaagtaga 1080

ctttataatc attataaata cctctcctcc attcctagac acagaaataa cagggcctag 1140

catacaccca ggtatgatgg cccttaacaa aagaagcaga tggataccta gcataaaaaa 1200

cagaccaggc agaaagcact atataaagat taaagtagga gccccccgaa tgttcacaga 1260

taagtggtac ccccaaacag acctctgtga catgacactc ctaacgatct ttgccagtgc 1320

ggcggatatg caatatccgt tcggctcacc actaactgac accatagttg tgtcattcca 1380

agttctgcaa tccatgtaca acgactgcct gagtgtactt cctgataatt ttgcagagac 1440

atcaggcaaa ggcacccaac tacatgagaa cataatacaa catctgccct actacaacac 1500

cacacaaaca caagcacaat ttaaaagatt tatagaaaac atgaatgcaa caaatggaga 1560

caatatatgg gcaagctaca taaacacaac caagttctca tccgcaaaca ctccaaagaa 1620

tgacacaggc ataggaggcc cttacactac atattcagac tcatggtaca aaggcacagt 1680

atacaatgac aaaattaaaa ccataccaat aaaagcaagc aagttatact acgagcaaac 1740

caaaaacctc attggcatta cattcactgg atccacacac agactccatt actgtggagg 1800

cctatactcc tccgtatggc tatcagcagg tagatcctac tttgaaacca aaggcccata 1860

cacagacata acttacaacc ccttttcaga cagaggagag ggtaacatgc tatggataga 1920

ctggctaact aaaaatgact cagtgtactc aaaaacaagt agcaagtgtc ttatagaaaa 1980

cctgcccctg tgggcctcag tatacggata taaagaatac tgcagcaagg taacaggaga 2040

cacaaacata gaacacaact gtagatgtgt tatcagaagc ccctacacag taccacaact 2100

gttagaccac aacaatccct tcagaggata cgtgccttat agcttcaact ttggaaatgg 2160

taaaatgcca ggcggtagca gcctagtgcc cattagaatg agagccaagt ggtaccccac 2220

tctgttccac caaaaagaag ttctagaagc catagcacag gcgggcccct tcgcatacca 2280

ctcagatatt aaaaaagtgt ccctgggcat aaagtacaga tttaagtggg tgtggggtgg 2340

caaccccgtg tcccaacagg ttgttagaaa cccctgcaag accacccaag gttcctcggg 2400

caatagagtg cctcgatcaa tacaagtcgt tgacccgcgg tacaacacgc cagaactcac 2460

catacacgcg tgggacttca gacatgggtt ctttggcaga aaagctatta agagaatgca 2520

agaacaacca atacctcatg acacttttc agcagggttc aagcgcagtc gccgagatac 2580

agaagcactc caatgcagcc aagaagagca acaaaaagaa aacttacttt tcccagtcca 2640

gcagctcaag cgagtccccc cgtgggagac ctcgcaagag agccaaagcg aggaagaaaa 2700

ctcgcaaaaa caggagaccc tctcccagca actcagagac cagctgcaca agcagcggct 2760

catgggagag caactccgat cgctcctcta ccaaatgcag agggtccaac aaaatcaaca 2820

cataaaccct atgttattgc caaagggtct ggcattaact tctatttctc acaatgtaat 2880

atagatatgt ttggtgaccc caaaccctac aagccctcct ccaatgactg gaaggaggag 2940

tacgaggccg caaagtactg ggacagaccc cccagacgcg acctgaggag cacccccttc 3000

tacccctggg cccccacccc caaaccatac aatgtcaact ttgccctcaa ctacaaataa 3060

acggtggccg tgggagtttc acttgtcggt gtctacctct taaggtcact aagcactccg 3120

agcgtaagcg aggagtgcga cccttcacca agggcaactc cctcgaagtc cggcgctacg 3180

cgcttcgcgc tgcgccggac atctcggacc ccccctcgac ccgaatcgct tgcgcgattc 3240

ggacctgcgg cctcgggggg gtcgggggct ttactaaaca gactccgagg tgccattgga 3300

cactgagggg gtgaacagca acgaaagtga gtggggccag acttcgccat aaggccttta 3360

tcttcttgcc atttgtccgc gaccgggggt cgctcctagg cgcggacccc gtttcggggt 3420

ccttccgggt tcatcggcgc cgttccagtg acgtcacggg cgccatgtta agtggctgtc 3480

gccgaggatt gacgtcacag ttcaaaggtc atcctcggcg gtaaccgcaa acatggcggt 3540

caatctcttc cgggtcaaag gtcgtgcata cgtcataagt cacatgacag gggtccactt 3600

aaacacggaa gtaggccccg acatgtgact cgtcacgtgt gtacacgtca cggccgccat 3660

tttgttttac aaaatggccg acttccttcc tgtttttttaa aaaaaggcgc gaaaaaaccg 3720

tcggcggggg ccgcgcgctg cgcgcgcggg aggcaatgcc tccccccccc cgcgcgcatg 3780

cgcgcgggtc ccccccccctc cgggggggctc cgccccccgg cccccccc 3828

<210> 904
<211> 119
<212> PRT
<213> Torque teno virus

<400> 904
Met Trp Arg Pro Pro Thr Arg Asn Ala Phe Asn Ile Gln Arg Asp Trp
1 5 10 15

Phe Tyr Ser Cys Phe His Ser His Ala Ser Met Cys Gly Cys Ala Asp
20 25 30

Phe Ile Gly His Phe Asn His Ile Ala Ala Met Leu Gly Arg Pro Glu
35 40 45

Asp Gln Asn Pro Pro Pro Pro Pro Gly Ala Leu Arg Pro Leu Pro Ala
50 55 60

Leu Pro Ala Ser Ser Glu Ala Pro Gly Asp Arg Ala Pro Trp Pro Met
65 70 75 80

Gly Gly Gly Gly Gly Asp Gly Gly Ala Arg Gly Gly Gly Gly Asp Gly
85 90 95

Ala Ala Gly Asp Ala Val Gly Asp Pro Ala Asp Ala Asp Leu Val Ala

100 105 110

Ala Ile Asp Ala Ala Glu Gln
115

<210> 905
<211> 273
<212> PRT
<213> Torque teno virus

<400> 905
Met Trp Arg Pro Pro Thr Arg Asn Ala Phe Asn Ile Gln Arg Asp Trp
1 5 10 15

Phe Tyr Ser Cys Phe His Ser His Ala Ser Met Cys Gly Cys Ala Asp
20 25 30

Phe Ile Gly His Phe Asn His Ile Ala Ala Met Leu Gly Arg Pro Glu
35 40 45

Asp Gln Asn Pro Pro Pro Pro Pro Gly Ala Leu Arg Pro Leu Pro Ala
50 55 60

Leu Pro Ala Ser Ser Glu Ala Pro Gly Asp Arg Ala Pro Trp Pro Met
65 70 75 80

Gly Gly Gly Gly Gly Asp Gly Gly Ala Arg Gly Gly Gly Gly Asp Gly
85 90 95

Ala Ala Gly Asp Ala Val Gly Asp Pro Ala Asp Ala Asp Leu Val Ala
100 105 110

Ala Ile Asp Ala Ala Glu Gln Leu Leu Glu Thr Pro Ala Arg Pro Pro
115 120 125

Lys Val Pro Arg Ala Ile Glu Cys Leu Asp Gln Tyr Lys Ser Leu Thr
130 135 140

Arg Gly Thr Thr Arg Gln Asn Ser Pro Tyr Thr Arg Gly Thr Ser Asp
145 150 155 160

Met Gly Ser Leu Ala Glu Lys Leu Leu Arg Glu Cys Lys Asn Asn Gln

```
                165                 170                 175

Tyr Leu Met Thr Leu Phe Gln Gln Gly Ser Ser Ala Val Ala Glu Ile
            180                 185                 190

Gln Lys His Ser Asn Ala Ala Lys Lys Ser Asn Lys Lys Lys Thr Tyr
        195                 200                 205

Phe Ser Gln Ser Ser Ser Ser Ser Glu Ser Pro Arg Gly Arg Pro Arg
    210                 215                 220

Lys Arg Ala Lys Ala Arg Lys Lys Thr Arg Lys Asn Arg Arg Pro Ser
225                 230                 235                 240

Pro Ser Asn Ser Glu Thr Ser Cys Thr Ser Ser Gly Ser Trp Glu Ser
                245                 250                 255

Asn Ser Asp Arg Ser Ser Thr Lys Cys Arg Gly Ser Asn Lys Ile Asn
            260                 265                 270

Thr
```

<210> 906
<211> 286
<212> PRT
<213> Torque teno virus

<400> 906

```
Met Trp Arg Pro Pro Thr Arg Asn Ala Phe Asn Ile Gln Arg Asp Trp
1               5                   10                  15

Phe Tyr Ser Cys Phe His Ser His Ala Ser Met Cys Gly Cys Ala Asp
            20                  25                  30

Phe Ile Gly His Phe Asn His Ile Ala Ala Met Leu Gly Arg Pro Glu
        35                  40                  45

Asp Gln Asn Pro Pro Pro Pro Pro Gly Ala Leu Arg Pro Leu Pro Ala
    50                  55                  60

Leu Pro Ala Ser Ser Glu Ala Pro Gly Asp Arg Ala Pro Trp Pro Met
```

```
65                  70                  75                  80

Gly Gly Gly Gly Gly Asp Gly Gly Ala Arg Gly Gly Gly Gly Asp Gly
                85                  90                  95

Ala Ala Gly Asp Ala Val Gly Asp Pro Ala Asp Ala Asp Leu Val Ala
            100                 105                 110

Ala Ile Asp Ala Ala Glu Gln Val Gln Ala Gln Ser Pro Arg Tyr Arg
        115                 120                 125

Ser Thr Pro Met Gln Pro Arg Arg Ala Thr Lys Arg Lys Leu Thr Phe
    130                 135                 140

Pro Ser Pro Ala Ala Gln Ala Ser Pro Pro Val Gly Asp Leu Ala Arg
145                 150                 155                 160

Glu Pro Lys Arg Gly Arg Lys Leu Ala Lys Thr Gly Asp Pro Leu Pro
                165                 170                 175

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

Pro Ile Ala Pro Leu Pro Asn Ala Glu Gly Pro Thr Lys Ser Thr His
        195                 200                 205

Lys Pro Tyr Val Ile Ala Lys Gly Ser Gly Ile Asn Phe Tyr Phe Ser
    210                 215                 220

Gln Cys Asn Ile Asp Met Phe Gly Asp Pro Lys Pro Tyr Lys Pro Ser
225                 230                 235                 240

Ser Asn Asp Trp Lys Glu Glu Tyr Glu Ala Ala Lys Tyr Trp Asp Arg
                245                 250                 255

Pro Pro Arg Arg Asp Leu Arg Ser Thr Pro Phe Tyr Pro Trp Ala Pro
            260                 265                 270

Thr Pro Lys Pro Tyr Asn Val Asn Phe Ala Leu Asn Tyr Lys
        275                 280                 285
```

<210> 907
<211> 51
<212> PRT
<213> Torque teno virus

<400> 907
Met Arg Ser Ile Phe Asn Val Thr Gly Ser Thr Val Ala Phe Thr Pro
1   5   10   15

Thr Leu Leu Cys Ala Ala Val Leu Ile Leu Leu Val Ile Ser Ile Ile
            20   25   30

Ser Leu Leu Cys Ser Ala Val Arg Lys Thr Arg Thr Leu Leu Arg His
        35   40   45

Pro Gly Leu
    50

<210> 908
<211> 767
<212> PRT
<213> Torque teno virus

<400> 908
Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Pro Trp Trp Arg
1   5   10   15

Arg Arg Trp Arg Arg Trp Arg Arg Arg Arg Arg Pro Arg Arg Arg Arg
            20   25   30

Pro Arg Arg Arg Tyr Arg Arg Arg Arg Thr Val Arg Arg Arg Gly Arg
        35   40   45

Gly Arg Trp Thr Arg Ala His Arg Arg Trp Arg Arg Lys Gly Lys Arg
    50   55   60

Ser Arg Lys Lys Lys Ile Ile Ile Arg Gln Trp Gln Pro Asn Tyr Thr
65   70   75   80

Arg Arg Cys Asn Ile Val Gly Tyr Met Pro Leu Leu Ile Cys Gly Glu
                85   90   95

Asn Thr Val Ala Thr Asn Tyr Ala Thr His Ser Asp Asp Ser Tyr Tyr

```
            100                 105                 110

Pro Gly Pro Phe Gly Gly Gly Met Thr Thr Asp Lys Phe Thr Leu Arg
        115                 120                 125

Ile Leu Tyr Asp Glu Tyr Lys Arg Phe Met Asn Tyr Trp Thr Ser Ser
    130                 135                 140

Asn Glu Asp Leu Asp Leu Cys Arg Tyr Leu Gly Cys Thr Leu Tyr Val
145                 150                 155                 160

Phe Arg His Pro Glu Val Asp Phe Ile Ile Ile Ile Asn Thr Ser Pro
                165                 170                 175

Pro Phe Leu Asp Thr Glu Ile Thr Gly Pro Ser Ile His Pro Gly Met
            180                 185                 190

Met Ala Leu Asn Lys Arg Ser Arg Trp Ile Pro Ser Ile Lys Asn Arg
        195                 200                 205

Pro Gly Arg Lys His Tyr Ile Lys Ile Lys Val Gly Ala Pro Arg Met
    210                 215                 220

Phe Thr Asp Lys Trp Tyr Pro Gln Thr Asp Leu Cys Asp Met Thr Leu
225                 230                 235                 240

Leu Thr Ile Phe Ala Ser Ala Ala Asp Met Gln Tyr Pro Phe Gly Ser
                245                 250                 255

Pro Leu Thr Asp Thr Ile Val Val Ser Phe Gln Val Leu Gln Ser Met
            260                 265                 270

Tyr Asn Asp Cys Leu Ser Val Leu Pro Asp Asn Phe Ala Glu Thr Ser
        275                 280                 285

Gly Lys Gly Thr Gln Leu His Glu Asn Ile Ile Gln His Leu Pro Tyr
    290                 295                 300

Tyr Asn Thr Thr Gln Thr Gln Ala Gln Phe Lys Arg Phe Ile Glu Asn
305                 310                 315                 320
```

```
Met Asn Ala Thr Asn Gly Asp Asn Ile Trp Ala Ser Tyr Ile Asn Thr
                325                 330                 335

Thr Lys Phe Ser Ser Ala Asn Thr Pro Lys Asn Asp Thr Gly Ile Gly
            340                 345                 350

Gly Pro Tyr Thr Thr Tyr Ser Asp Ser Trp Tyr Lys Gly Thr Val Tyr
        355                 360                 365

Asn Asp Lys Ile Lys Thr Ile Pro Ile Lys Ala Ser Lys Leu Tyr Tyr
    370                 375                 380

Glu Gln Thr Lys Asn Leu Ile Gly Ile Thr Phe Thr Gly Ser Thr His
385                 390                 395                 400

Arg Leu His Tyr Cys Gly Gly Leu Tyr Ser Ser Val Trp Leu Ser Ala
                405                 410                 415

Gly Arg Ser Tyr Phe Glu Thr Lys Gly Pro Tyr Thr Asp Ile Thr Tyr
            420                 425                 430

Asn Pro Phe Ser Asp Arg Gly Glu Gly Asn Met Leu Trp Ile Asp Trp
        435                 440                 445

Leu Thr Lys Asn Asp Ser Val Tyr Ser Lys Thr Ser Ser Lys Cys Leu
    450                 455                 460

Ile Glu Asn Leu Pro Leu Trp Ala Ser Val Tyr Gly Tyr Lys Glu Tyr
465                 470                 475                 480

Cys Ser Lys Val Thr Gly Asp Thr Asn Ile Glu His Asn Cys Arg Cys
                485                 490                 495

Val Ile Arg Ser Pro Tyr Thr Val Pro Gln Leu Leu Asp His Asn Asn
            500                 505                 510

Pro Phe Arg Gly Tyr Val Pro Tyr Ser Phe Asn Phe Gly Asn Gly Lys
        515                 520                 525

Met Pro Gly Gly Ser Ser Leu Val Pro Ile Arg Met Arg Ala Lys Trp
```

```
    530                 535                 540

Tyr Pro Thr Leu Phe His Gln Lys Glu Val Leu Glu Ala Ile Ala Gln
545                 550                 555                 560

Ala Gly Pro Phe Ala Tyr His Ser Asp Ile Lys Lys Val Ser Leu Gly
                565                 570                 575

Ile Lys Tyr Arg Phe Lys Trp Val Trp Gly Gly Asn Pro Val Ser Gln
            580                 585                 590

Gln Val Val Arg Asn Pro Cys Lys Thr Thr Gln Gly Ser Ser Gly Asn
        595                 600                 605

Arg Val Pro Arg Ser Ile Gln Val Val Asp Pro Arg Tyr Asn Thr Pro
    610                 615                 620

Glu Leu Thr Ile His Ala Trp Asp Phe Arg His Gly Phe Phe Gly Arg
625                 630                 635                 640

Lys Ala Ile Lys Arg Met Gln Glu Gln Pro Ile Pro His Asp Thr Phe
                645                 650                 655

Ser Ala Gly Phe Lys Arg Ser Arg Arg Asp Thr Glu Ala Leu Gln Cys
            660                 665                 670

Ser Gln Glu Glu Gln Gln Lys Glu Asn Leu Leu Phe Pro Val Gln Gln
        675                 680                 685

Leu Lys Arg Val Pro Pro Trp Glu Thr Ser Gln Glu Ser Gln Ser Glu
    690                 695                 700

Glu Glu Asn Ser Gln Lys Gln Glu Thr Leu Ser Gln Gln Leu Arg Asp
705                 710                 715                 720

Gln Leu His Lys Gln Arg Leu Met Gly Glu Gln Leu Arg Ser Leu Leu
                725                 730                 735

Tyr Gln Met Gln Arg Val Gln Gln Asn Gln His Ile Asn Pro Met Leu
            740                 745                 750
```

```
Leu Pro Lys Gly Leu Ala Leu Thr Ser Ile Ser His Asn Val Ile
        755                 760                 765


<210> 909
<211> 216
<212> PRT
<213> Torque teno virus

<400> 909
Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Pro Trp Trp Arg
1               5                   10                  15


Arg Arg Trp Arg Arg Trp Arg Arg Arg Arg Arg Pro Arg Arg Arg Arg
            20                  25                  30


Pro Arg Arg Arg Tyr Arg Arg Arg Arg Thr Val Val Arg Asn Pro Cys
        35                  40                  45


Lys Thr Thr Gln Gly Ser Ser Gly Asn Arg Val Pro Arg Ser Ile Gln
    50                  55                  60


Val Val Asp Pro Arg Tyr Asn Thr Pro Glu Leu Thr Ile His Ala Trp
65                  70                  75                  80


Asp Phe Arg His Gly Phe Phe Gly Arg Lys Ala Ile Lys Arg Met Gln
                85                  90                  95


Glu Gln Pro Ile Pro His Asp Thr Phe Ser Ala Gly Phe Lys Arg Ser
            100                 105                 110


Arg Arg Asp Thr Glu Ala Leu Gln Cys Ser Gln Glu Glu Gln Gln Lys
        115                 120                 125


Glu Asn Leu Leu Phe Pro Val Gln Gln Leu Lys Arg Val Pro Pro Trp
    130                 135                 140


Glu Thr Ser Gln Glu Ser Gln Ser Glu Glu Glu Asn Ser Gln Lys Gln
145                 150                 155                 160


Glu Thr Leu Ser Gln Gln Leu Arg Asp Gln Leu His Lys Gln Arg Leu
                165                 170                 175
```

Met Gly Glu Gln Leu Arg Ser Leu Leu Tyr Gln Met Gln Arg Val Gln
            180   185   190

Gln Asn Gln His Ile Asn Pro Met Leu Leu Pro Lys Gly Leu Ala Leu
        195   200   205

Thr Ser Ile Ser His Asn Val Ile
    210   215

<210> 910
<211> 131
<212> PRT
<213> Torque teno virus

<400> 910
Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Pro Trp Trp Arg
1   5   10   15

Arg Arg Trp Arg Arg Trp Arg Arg Arg Arg Arg Pro Arg Arg Arg Arg
            20   25   30

Pro Arg Arg Arg Tyr Arg Arg Arg Arg Thr Gly Ser Ser Ala Val Ala
        35   40   45

Glu Ile Gln Lys His Ser Asn Ala Ala Lys Lys Ser Asn Lys Lys Lys
    50   55   60

Thr Tyr Phe Ser Gln Ser Ser Ser Ser Ser Glu Ser Pro Arg Gly Arg
65   70   75   80

Pro Arg Lys Arg Ala Lys Ala Arg Lys Lys Thr Arg Lys Asn Arg Arg
                85   90   95

Pro Ser Pro Ser Asn Ser Glu Thr Ser Cys Thr Ser Ser Gly Ser Trp
            100   105   110

Glu Ser Asn Ser Asp Arg Ser Ser Thr Lys Cys Arg Gly Ser Asn Lys
        115   120   125

Ile Asn Thr
    130

<210> 911
<211> 3815
<212> DNA
<213> Torque teno virus

<400> 911
aagatcgtca ctaaccacgt gactcctctc gcccaatcag tgtctacgtc gtccatttcc 60

tgggcatggt ctacatcctg atataaagcg atgcacttcc gaatggctga gttttccacg 120

cccgtccgcg gcgagatcgc gacggaggag cgatcgagcg tcccgagggc gggtgccgga 180

ggtgagttta cacaccgcag tcaaggggca attcgggctc gggactggcc gggctatggg 240

caaggctctt aaagcgtacg tcccccgcta tgtttctcgg cagggtgtgg aggaaacaga 300

aaaggaaagt gcttctgctg gctgtgcgag ctacacagaa aacatcttcc atgagtatct 360

ggcgtccccc ccttgggaat gtctcctaca gggagagaaa ttggcttcag gccgtcgaaa 420

catcccacag ttctttttgt ggctgtggtg attttattct tcatcttact aatttggctg 480

cacgctttgc tctccagggg cccccgccag agggtggtcc acctcggccg aggccgccgc 540

tcctgagagc gctgccggcc cccgaggtcc gcagggagac gcgcacagag aaccggggcg 600

cctccggtga gccatggcct ggcgatggtg gtggcagaga cgatggcgcc gccgccggtg 660

gccccgcaga cggtggagac gcctacgacg ccggagacct agacgacctg ttcgccgccg 720

tcgaagaaga acaacagtaa ggaggcggag gtggaggggc agacgtgggc gacgcacata 780

cacccgacgc gcggtcagac gcagacgcag acccagaaag agacttgtac tgactcagtg 840

gagcccccag acagtcagaa actgctcaat aaggggcata gtgcccatgg taatatgcgg 900

acacacaaaa gcaggtagaa actatgctat tcatagcgag gacttcacca cacagataca 960

acccttcggg ggcagtttca gcacgaccac ctggtcccta aaagtgctgt gggacgagca 1020

ccagaaattc cagaacagat ggtcctaccc aaacacacaa ctagacctgg ccagatacag 1080

aggggtcacc ttctggttct acagagacca gaaaacagac tatatagtac agtggagtag 1140

gaatcccccc tttaaactca ataaatacag cagtgccatg taccacccgg gcatgatgat 1200

gcaggccaaa aggaaactag ttgtacctag tttccagacc agacccaaag gcaagaagag 1260

atacagagtc acaataaaac cccctaacat gtttgctgac aagtggtaca ctcaagagga 1320

cctgtgtccg gtacctcttg tgcaaattgt ggtttctgcg gcgagcctgc tacatccgtt 1380

ctgcccacca caaacgaaca acccttgcat caccttccag gttttgaaag acatatatga 1440

tgaatgcata ggagttaacg aaactatgaa agataagtat aagaaattac aaacaacact 1500

atacaccact tgcacatact atcaaacaac acaagtactg gcacagctat ctcctgcctt 1560

tcaacctgct atgaaaccta ctactacaca atcagcagct acagcgacaa cactaggaaa 1620

ctatgtacca gagttaaagt acaacaatgg ctcttttcac acaggacaaa acgcagtatt 1680

cggcatgtgc tcatacaaac caacagacag cataatgaca aaagctaatg gctggttttg 1740

gcaaaaccta atggtagaca acaacctaca tagttcttat ggcaaggcaa cattagaatg 1800

catggagtat cacacaggca tatacagctc tatatttcta agtccacaaa gatctttaga 1860

attcccagca gcataccaag acgttacata caaccctaac tgtgatagag cagttggaaa 1920

cgtagtttgg tttcagtaca gcactaaaat ggatacaaat tttgatgaaa caaaatgtaa 1980

atgtgtcctt aaaaacattc cactgtgggc ggccttcaat ggctactcag actttataat 2040

gcaagaactc agcataagta cagaaatcca caactttggc atagtgtgct ttcagtgccc 2100

gtacactttt ccccccctgtt tcaataaaaa caaaccccta aaggggtacg tgttctatga 2160

caccaccttt ggtaatggaa aaatgccaga cggatcgggg cacgtaccca tctactggca 2220

gcagagatgg tggatcagac tagccttcca ggtccaggtc atgcatgact ttgtactaac 2280

aggccccttt agctacaaag atgacctagc aaacaccaca ctcacagcca gatacaaatt 2340

taaattcaaa tggggcggca atatcatccc tgaacagatt atcaagaacc cgtgtcacag 2400

agagcagtcc ctcgcttcct atcccgatag acaacgtcgc gacctacaag ttgttgaccc 2460

atcaaccatg ggcccgatct acaccttcca cacatgggac tggcgacggg ggctttttgg 2520

tgcagatgct atccagagag tgtcacaaaa accgggagat gctctccgct ttacaaaccc 2580

tttcaagaga cccagatatc ttcccccgac agacagagaa gactaccgac aagaagaaga 2640

cttcgcttta caggaaaaaa gacggcgcac atccacagaa gaagcccagg acgaggagag 2700

ccccccggaa agcgcgccgc tcctacagca gcagcagcag cagcggcagc tctcagtcca 2760

cctcgcggag cagcagcgac tcggagtcca actccgatac atcctccaag aagtcctcaa 2820

aacgcaagcg ggtctccacc taaacccccct attattaggc ccgccacaaa caaggtctat 2880

ctctttgagc cctccaaagg cctactcccc atagtaggaa aagaggcctg ggaggacgag 2940

tactgcacct gcaagtactg ggatcgccct cccagaacca accacctaga catccccact 3000

tatccctgga tgcccacaaa cttcaaagtc agcttcaaac ttggatttaa accctaaata 3060

aaaatacaag gccgtacact gttcacttgt cggtgtctac ctctataagt cactaagcac 3120

tccgagcgca gcgaggagtg cgaccctcag cggtgggtgc aacgccctcg gcggccgcgc 3180

gctacgcctt cggctgcgcg cggcacctcg gacccccgct cgtgctgaca cgctcgcgcg 3240

tgtcagacca cttcgggctc gcgggggtcg ggaattttgc taaacagact ccgagttgct 3300

cttggacact gtagctgtga atcagtaacg aaagtgagtg gggccagact tcgccataag 3360

gcctttatct tcttgccatt ggtccgtctc gggggtcgcc ataggcttcg ggctcggttt 3420

taggccttcc ggactaccaa aatggcggat tccgtgacgt catggccgcc attttaagta 3480

aggcggaaca ggctgtcacc ccgtgtcaaa gttcaggggt cagccttccg ctttacacaa 3540

aatggaggtc aatatcttcc gggtcaaagg tcgctaccgc gtcataagtc acgtggggaa 3600

ggctgctgtg aatccggaag tagctgaccc acgtgacttg tcacgtgact agcacgtcac 3660

ggcagccatt ttgaatcaca aaatggccga cttccttcct ctttttttaaa aataacggcc 3720

cggcggcggc gcgcgcgctt cgcgccgctc cgccccccgc gcgcatgcgc gggacccccc 3780

cccgcggggg gctccgcccc ccggtccccc ccccg 3815

<210> 912
<211> 129
<212> PRT
<213> Torque teno virus

<400> 912
Met Ser Ile Trp Arg Pro Pro Leu Gly Asn Val Ser Tyr Arg Glu Arg
1 5 10 15

Asn Trp Leu Gln Ala Val Glu Thr Ser His Ser Ser Phe Cys Gly Cys
20 25 30

Gly Asp Phe Ile Leu His Leu Thr Asn Leu Ala Ala Arg Phe Ala Leu
35 40 45

Gln Gly Pro Pro Pro Glu Gly Gly Pro Pro Arg Pro Arg Pro Pro Leu
50 55 60

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

Asn Arg Gly Ala Ser Gly Glu Pro Trp Pro Gly Asp Gly Gly Gly Arg
85 90 95

Asp Asp Gly Ala Ala Ala Gly Gly Pro Ala Asp Gly Gly Asp Ala Tyr
100 105 110

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

Gln

<210> 913
<211> 283
<212> PRT
<213> Torque teno virus

<400> 913
Met Ser Ile Trp Arg Pro Pro Leu Gly Asn Val Ser Tyr Arg Glu Arg
1 5 10 15

Asn Trp Leu Gln Ala Val Glu Thr Ser His Ser Ser Phe Cys Gly Cys
20 25 30

Gly Asp Phe Ile Leu His Leu Thr Asn Leu Ala Ala Arg Phe Ala Leu
35 40 45

Gln Gly Pro Pro Pro Glu Gly Gly Pro Pro Arg Pro Arg Pro Pro Leu
50 55 60

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

Asn Arg Gly Ala Ser Gly Glu Pro Trp Pro Gly Asp Gly Gly Gly Arg
85 90 95

Asp Asp Gly Ala Ala Ala Gly Gly Pro Ala Asp Gly Gly Asp Ala Tyr
100 105 110

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

```
Gln Leu Ser Arg Thr Arg Val Thr Glu Ser Ser Pro Ser Leu Pro Ile
    130                 135                 140

Pro Ile Asp Asn Val Ala Thr Tyr Lys Leu Leu Thr His Gln Pro Trp
145                 150                 155                 160

Ala Arg Ser Thr Pro Ser Thr His Gly Thr Gly Asp Gly Gly Phe Leu
                165                 170                 175

Val Gln Met Leu Ser Arg Glu Cys His Lys Asn Arg Glu Met Leu Ser
            180                 185                 190

Ala Leu Gln Thr Leu Ser Arg Asp Pro Asp Ile Phe Pro Arg Gln Thr
        195                 200                 205

Glu Lys Thr Thr Asp Lys Lys Lys Thr Ser Leu Tyr Arg Lys Lys Asp
    210                 215                 220

Gly Ala His Pro Gln Lys Lys Pro Arg Thr Arg Arg Ala Pro Arg Lys
225                 230                 235                 240

Ala Arg Arg Ser Tyr Ser Ser Ser Ser Ser Ser Gly Ser Ser Gln Ser
                245                 250                 255

Thr Ser Arg Ser Ser Ser Asp Ser Glu Ser Asn Ser Asp Thr Ser Ser
            260                 265                 270

Lys Lys Ser Ser Lys Arg Lys Arg Val Ser Thr
        275                 280
```

<210> 914
<211> 305
<212> PRT
<213> Torque teno virus

```
<400> 914
Met Ser Ile Trp Arg Pro Pro Leu Gly Asn Val Ser Tyr Arg Glu Arg
1               5                   10                  15

Asn Trp Leu Gln Ala Val Glu Thr Ser His Ser Ser Phe Cys Gly Cys
            20                  25                  30
```

```
Gly Asp Phe Ile Leu His Leu Thr Asn Leu Ala Ala Arg Phe Ala Leu
        35                  40                  45

Gln Gly Pro Pro Pro Glu Gly Gly Pro Pro Arg Pro Arg Pro Pro Leu
    50                  55                  60

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

Asn Arg Gly Ala Ser Gly Glu Pro Trp Pro Gly Asp Gly Gly Gly Arg
                85                  90                  95

Asp Asp Gly Ala Ala Ala Gly Gly Pro Ala Asp Gly Gly Asp Ala Tyr
            100                 105                 110

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

Gln Cys Tyr Pro Glu Ser Val Thr Lys Thr Gly Arg Cys Ser Pro Leu
    130                 135                 140

Tyr Lys Pro Phe Gln Glu Thr Gln Ile Ser Ser Pro Asp Arg Gln Arg
145                 150                 155                 160

Arg Leu Pro Thr Arg Arg Arg Leu Arg Phe Thr Gly Lys Lys Thr Ala
                165                 170                 175

His Ile His Arg Arg Ser Pro Gly Arg Gly Glu Pro Pro Gly Lys Arg
            180                 185                 190

Ala Ala Pro Thr Ala Ala Ala Ala Ala Ala Ala Ala Leu Ser Pro Pro
        195                 200                 205

Arg Gly Ala Ala Ala Thr Arg Ser Pro Thr Pro Ile His Pro Pro Arg
    210                 215                 220

Ser Pro Gln Asn Ala Ser Gly Ser Pro Pro Lys Pro Pro Ile Ile Arg
225                 230                 235                 240

Pro Ala Thr Asn Lys Val Tyr Leu Phe Glu Pro Ser Lys Gly Leu Leu
                245                 250                 255
```

Pro Ile Val Gly Lys Glu Ala Trp Glu Asp Glu Tyr Cys Thr Cys Lys
260 265 270

Tyr Trp Asp Arg Pro Pro Arg Thr Asn His Leu Asp Ile Pro Thr Tyr
275 280 285

Pro Trp Met Pro Thr Asn Phe Lys Val Ser Phe Lys Leu Gly Phe Lys
290 295 300

Pro
305

<210> 915
<211> 55
<212> PRT
<213> Torque teno virus

<400> 915
Met Ser Pro Thr Gly Arg Glu Ile Gly Phe Arg Pro Ser Lys His Pro
1 5 10 15

Thr Val Leu Phe Val Ala Val Val Ile Leu Phe Phe Ile Leu Leu Ile
20 25 30

Trp Leu His Ala Leu Leu Ser Arg Gly Pro Arg Gln Arg Val Val His
35 40 45

Leu Gly Arg Gly Arg Arg Ser
50 55

<210> 916
<211> 766
<212> PRT
<213> Torque teno virus

<400> 916
Met Ala Trp Arg Trp Trp Trp Gln Arg Arg Trp Arg Arg Arg Arg Trp
1 5 10 15

Pro Arg Arg Arg Trp Arg Arg Leu Arg Arg Arg Arg Pro Arg Arg Pro
20 25 30

```
Val Arg Arg Arg Arg Arg Arg Thr Thr Val Arg Arg Arg Arg Trp Arg
        35                  40                  45

Gly Arg Arg Gly Arg Arg Thr Tyr Thr Arg Arg Ala Val Arg Arg Arg
    50                  55                  60

Arg Arg Pro Arg Lys Arg Leu Val Leu Thr Gln Trp Ser Pro Gln Thr
65                  70                  75                  80

Val Arg Asn Cys Ser Ile Arg Gly Ile Val Pro Met Val Ile Cys Gly
                85                  90                  95

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

Thr Gln Ile Gln Pro Phe Gly Gly Ser Phe Ser Thr Thr Thr Trp Ser
        115                 120                 125

Leu Lys Val Leu Trp Asp Glu His Gln Lys Phe Gln Asn Arg Trp Ser
    130                 135                 140

Tyr Pro Asn Thr Gln Leu Asp Leu Ala Arg Tyr Arg Gly Val Thr Phe
145                 150                 155                 160

Trp Phe Tyr Arg Asp Gln Lys Thr Asp Tyr Ile Val Gln Trp Ser Arg
                165                 170                 175

Asn Pro Pro Phe Lys Leu Asn Lys Tyr Ser Ser Ala Met Tyr His Pro
            180                 185                 190

Gly Met Met Met Gln Ala Lys Arg Lys Leu Val Val Pro Ser Phe Gln
        195                 200                 205

Thr Arg Pro Lys Gly Lys Lys Arg Tyr Arg Val Thr Ile Lys Pro Pro
    210                 215                 220

Asn Met Phe Ala Asp Lys Trp Tyr Thr Gln Glu Asp Leu Cys Pro Val
225                 230                 235                 240

Pro Leu Val Gln Ile Val Val Ser Ala Ala Ser Leu Leu His Pro Phe
                245                 250                 255
```

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

Asp Ile Tyr Asp Glu Cys Ile Gly Val Asn Glu Thr Met Lys Asp Lys
        275                 280                 285

Tyr Lys Lys Leu Gln Thr Thr Leu Tyr Thr Thr Cys Thr Tyr Tyr Gln
    290                 295                 300

Thr Thr Gln Val Leu Ala Gln Leu Ser Pro Ala Phe Gln Pro Ala Met
305                 310                 315                 320

Lys Pro Thr Thr Thr Gln Ser Ala Ala Thr Ala Thr Thr Leu Gly Asn
                325                 330                 335

Tyr Val Pro Glu Leu Lys Tyr Asn Asn Gly Ser Phe His Thr Gly Gln
            340                 345                 350

Asn Ala Val Phe Gly Met Cys Ser Tyr Lys Pro Thr Asp Ser Ile Met
        355                 360                 365

Thr Lys Ala Asn Gly Trp Phe Trp Gln Asn Leu Met Val Asp Asn Asn
    370                 375                 380

Leu His Ser Ser Tyr Gly Lys Ala Thr Leu Glu Cys Met Glu Tyr His
385                 390                 395                 400

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

Phe Pro Ala Ala Tyr Gln Asp Val Thr Tyr Asn Pro Asn Cys Asp Arg
            420                 425                 430

Ala Val Gly Asn Val Val Trp Phe Gln Tyr Ser Thr Lys Met Asp Thr
        435                 440                 445

Asn Phe Asp Glu Thr Lys Cys Lys Cys Val Leu Lys Asn Ile Pro Leu
    450                 455                 460
```

```
Trp Ala Ala Phe Asn Gly Tyr Ser Asp Phe Ile Met Gln Glu Leu Ser
465                 470                 475                 480

Ile Ser Thr Glu Ile His Asn Phe Gly Ile Val Cys Phe Gln Cys Pro
                485                 490                 495

Tyr Thr Phe Pro Pro Cys Phe Asn Lys Asn Lys Pro Leu Lys Gly Tyr
            500                 505                 510

Val Phe Tyr Asp Thr Thr Phe Gly Asn Gly Lys Met Pro Asp Gly Ser
        515                 520                 525

Gly His Val Pro Ile Tyr Trp Gln Gln Arg Trp Trp Ile Arg Leu Ala
    530                 535                 540

Phe Gln Val Gln Val Met His Asp Phe Val Leu Thr Gly Pro Phe Ser
545                 550                 555                 560

Tyr Lys Asp Asp Leu Ala Asn Thr Thr Leu Thr Ala Arg Tyr Lys Phe
                565                 570                 575

Lys Phe Lys Trp Gly Gly Asn Ile Ile Pro Glu Gln Ile Ile Lys Asn
            580                 585                 590

Pro Cys His Arg Glu Gln Ser Leu Ala Ser Tyr Pro Asp Arg Gln Arg
        595                 600                 605

Arg Asp Leu Gln Val Val Asp Pro Ser Thr Met Gly Pro Ile Tyr Thr
    610                 615                 620

Phe His Thr Trp Asp Trp Arg Arg Gly Leu Phe Gly Ala Asp Ala Ile
625                 630                 635                 640

Gln Arg Val Ser Gln Lys Pro Gly Asp Ala Leu Arg Phe Thr Asn Pro
                645                 650                 655

Phe Lys Arg Pro Arg Tyr Leu Pro Pro Thr Asp Arg Glu Asp Tyr Arg
            660                 665                 670

Gln Glu Glu Asp Phe Ala Leu Gln Glu Lys Arg Arg Arg Thr Ser Thr
        675                 680                 685
```

Glu Glu Ala Gln Asp Glu Glu Ser Pro Pro Glu Ser Ala Pro Leu Leu
690 695 700

Gln Gln Gln Gln Gln Gln Arg Gln Leu Ser Val His Leu Ala Glu Gln
705 710 715 720

Gln Arg Leu Gly Val Gln Leu Arg Tyr Ile Leu Gln Glu Val Leu Lys
725 730 735

Thr Gln Ala Gly Leu His Leu Asn Pro Leu Leu Leu Gly Pro Pro Gln
740 745 750

Thr Arg Ser Ile Ser Leu Ser Pro Pro Lys Ala Tyr Ser Pro
755 760 765

<210> 917
<211> 219
<212> PRT
<213> Torque teno virus

<400> 917
Met Ala Trp Arg Trp Trp Trp Gln Arg Arg Trp Arg Arg Arg Arg Trp
1 5 10 15

Pro Arg Arg Arg Trp Arg Arg Leu Arg Arg Arg Arg Pro Arg Arg Pro
20 25 30

Val Arg Arg Arg Arg Arg Arg Thr Thr Ile Ile Lys Asn Pro Cys His
35 40 45

Arg Glu Gln Ser Leu Ala Ser Tyr Pro Asp Arg Gln Arg Arg Asp Leu
50 55 60

Gln Val Val Asp Pro Ser Thr Met Gly Pro Ile Tyr Thr Phe His Thr
65 70 75 80

Trp Asp Trp Arg Arg Gly Leu Phe Gly Ala Asp Ala Ile Gln Arg Val
85 90 95

Ser Gln Lys Pro Gly Asp Ala Leu Arg Phe Thr Asn Pro Phe Lys Arg
100 105 110

```
Pro Arg Tyr Leu Pro Pro Thr Asp Arg Glu Asp Tyr Arg Gln Glu Glu
        115                 120                 125

Asp Phe Ala Leu Gln Glu Lys Arg Arg Arg Thr Ser Thr Glu Glu Ala
    130                 135                 140

Gln Asp Glu Glu Ser Pro Pro Glu Ser Ala Pro Leu Leu Gln Gln Gln
145                 150                 155                 160

Gln Gln Gln Arg Gln Leu Ser Val His Leu Ala Glu Gln Gln Arg Leu
                165                 170                 175

Gly Val Gln Leu Arg Tyr Ile Leu Gln Glu Val Leu Lys Thr Gln Ala
            180                 185                 190

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

Ile Ser Leu Ser Pro Pro Lys Ala Tyr Ser Pro
    210                 215


<210> 918
<211> 146
<212> PRT
<213> Torque teno virus

<400> 918
Met Ala Trp Arg Trp Trp Trp Gln Arg Arg Trp Arg Arg Arg Arg Trp
1               5                   10                  15

Pro Arg Arg Arg Trp Arg Arg Leu Arg Arg Arg Arg Pro Arg Arg Pro
            20                  25                  30

Val Arg Arg Arg Arg Arg Arg Thr Thr Met Leu Ser Arg Glu Cys His
        35                  40                  45

Lys Asn Arg Glu Met Leu Ser Ala Leu Gln Thr Leu Ser Arg Asp Pro
    50                  55                  60

Asp Ile Phe Pro Arg Gln Thr Glu Lys Thr Thr Asp Lys Lys Lys Thr
65                  70                  75                  80
```

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

Thr Arg Arg Ala Pro Arg Lys Ala Arg Arg Ser Tyr Ser Ser Ser Ser
100 105 110

Ser Ser Gly Ser Ser Gln Ser Thr Ser Arg Ser Ser Ser Asp Ser Glu
115 120 125

Ser Asn Ser Asp Thr Ser Ser Lys Lys Ser Ser Lys Arg Lys Arg Val
130 135 140

Ser Thr
145

<210> 919
<211> 677
<212> PRT
<213> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno midi virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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> Torque teno virus

<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
<211> 767
<212> PRT
<213> Torque teno virus

<400> 937

```
Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Pro Trp Trp Arg
1               5                   10                  15


Arg Arg Trp Arg Arg Trp Arg Arg Arg Arg Arg Pro Arg Arg Arg Arg
            20                  25                  30


Pro Arg Arg Arg Tyr Arg Arg Arg Arg Thr Val Arg Arg Arg Gly Arg
        35                  40                  45


Gly Arg Trp Thr Arg Ala His Arg Arg Trp Arg Arg Lys Gly Lys Arg
    50                  55                  60


Ser Arg Lys Lys Lys Ile Ile Ile Arg Gln Trp Gln Pro Asn Tyr Thr
65                  70                  75                  80


Arg Arg Cys Asn Ile Val Gly Tyr Met Pro Leu Leu Ile Cys Gly Glu
                85                  90                  95


Asn Thr Val Ala Thr Asn Tyr Ala Thr His Ser Asp Asp Ser Tyr Tyr
            100                 105                 110


Pro Gly Pro Phe Gly Gly Gly Met Thr Thr Asp Lys Phe Thr Leu Arg
        115                 120                 125


Ile Leu Tyr Asp Glu Tyr Lys Arg Phe Met Asn Tyr Trp Thr Ser Ser
    130                 135                 140


Asn Glu Asp Leu Asp Leu Cys Arg Tyr Leu Gly Cys Thr Leu Tyr Val
145                 150                 155                 160
```

```
Phe Arg His Pro Glu Val Asp Phe Ile Ile Ile Ile Asn Thr Ser Pro
                165                 170                 175

Pro Phe Leu Asp Thr Glu Ile Thr Gly Pro Ser Ile His Pro Gly Met
            180                 185                 190

Met Ala Leu Asn Lys Arg Ser Arg Trp Ile Pro Ser Ile Lys Asn Arg
        195                 200                 205

Pro Gly Arg Lys His Tyr Ile Lys Ile Lys Val Gly Ala Pro Arg Met
    210                 215                 220

Phe Thr Asp Lys Trp Tyr Pro Gln Thr Asp Leu Cys Asp Met Thr Leu
225                 230                 235                 240

Leu Thr Ile Phe Ala Ser Ala Ala Asp Met Gln Tyr Pro Phe Gly Ser
                245                 250                 255

Pro Leu Thr Asp Thr Ile Val Val Ser Phe Gln Val Leu Gln Ser Met
            260                 265                 270

Tyr Asn Asp Cys Leu Ser Val Leu Pro Asp Asn Phe Ala Glu Thr Ser
        275                 280                 285

Gly Lys Gly Thr Gln Leu His Glu Asn Ile Ile Gln His Leu Pro Tyr
    290                 295                 300

Tyr Asn Thr Thr Gln Thr Gln Ala Gln Phe Lys Arg Phe Ile Glu Asn
305                 310                 315                 320

Met Asn Ala Thr Asn Gly Asp Asn Ile Trp Ala Ser Tyr Ile Asn Thr
                325                 330                 335

Thr Lys Phe Ser Ser Ala Asn Thr Pro Lys Asn Asp Thr Gly Ile Gly
            340                 345                 350

Gly Pro Tyr Thr Thr Tyr Ser Asp Ser Trp Tyr Lys Gly Thr Val Tyr
        355                 360                 365
```

```
Asn Asp Lys Ile Lys Thr Ile Pro Ile Lys Ala Ser Lys Leu Tyr Tyr
    370                 375                 380

Glu Gln Thr Lys Asn Leu Ile Gly Ile Thr Phe Thr Gly Ser Thr His
385                 390                 395                 400

Arg Leu His Tyr Cys Gly Gly Leu Tyr Ser Ser Val Trp Leu Ser Ala
                405                 410                 415

Gly Arg Ser Tyr Phe Glu Thr Lys Gly Pro Tyr Thr Asp Ile Thr Tyr
            420                 425                 430

Asn Pro Phe Ser Asp Arg Gly Glu Gly Asn Met Leu Trp Ile Asp Trp
        435                 440                 445

Leu Thr Lys Asn Asp Ser Val Tyr Ser Lys Thr Ser Ser Lys Cys Leu
    450                 455                 460

Ile Glu Asn Leu Pro Leu Trp Ala Ser Val Tyr Gly Tyr Lys Glu Tyr
465                 470                 475                 480

Cys Ser Lys Val Thr Gly Asp Thr Asn Ile Glu His Asn Cys Arg Cys
                485                 490                 495

Val Ile Arg Ser Pro Tyr Thr Val Pro Gln Leu Leu Asp His Asn Asn
            500                 505                 510

Pro Phe Arg Gly Tyr Val Pro Tyr Ser Phe Asn Phe Gly Asn Gly Lys
        515                 520                 525

Met Pro Gly Gly Ser Ser Leu Val Pro Ile Arg Met Arg Ala Lys Trp
    530                 535                 540

Tyr Pro Thr Leu Phe His Gln Lys Glu Val Leu Glu Ala Ile Ala Gln
545                 550                 555                 560

Ala Gly Pro Phe Ala Tyr His Ser Asp Ile Lys Lys Val Ser Leu Gly
                565                 570                 575

Ile Lys Tyr Arg Phe Lys Trp Val Trp Gly Gly Asn Pro Val Ser Gln
            580                 585                 590
```

```
Gln Val Val Arg Asn Pro Cys Lys Thr Thr Gln Gly Ser Ser Gly Asn
        595                 600                 605

Arg Val Pro Arg Ser Ile Gln Val Val Asp Pro Arg Tyr Asn Thr Pro
    610                 615                 620

Glu Leu Thr Ile His Ala Trp Asp Phe Arg His Gly Phe Phe Gly Arg
625                 630                 635                 640

Lys Ala Ile Lys Arg Met Gln Glu Gln Pro Ile Pro His Asp Thr Phe
                645                 650                 655

Ser Ala Gly Phe Lys Arg Ser Arg Arg Asp Thr Glu Ala Leu Gln Cys
            660                 665                 670

Ser Gln Glu Glu Gln Gln Lys Glu Asn Leu Leu Phe Pro Val Gln Gln
        675                 680                 685

Leu Lys Arg Val Pro Pro Trp Glu Thr Ser Gln Glu Ser Gln Ser Glu
    690                 695                 700

Glu Glu Asn Ser Gln Lys Gln Glu Thr Leu Ser Gln Gln Leu Arg Asp
705                 710                 715                 720

Gln Leu His Lys Gln Arg Leu Met Gly Glu Gln Leu Arg Ser Leu Leu
                725                 730                 735

Tyr Gln Met Gln Arg Val Gln Gln Asn Gln His Ile Asn Pro Met Leu
            740                 745                 750

Leu Pro Lys Gly Leu Ala Leu Thr Ser Ile Ser His Asn Val Ile
        755                 760                 765
```

<210> 938
<211> 69
<212> PRT
<213> Torque teno virus

<400> 938

```
Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Pro Trp Trp Arg
1               5                   10                  15
```

```
Arg Arg Trp Arg Arg Trp Arg Arg Arg Arg Arg Pro Arg Arg Arg Arg
            20                  25                  30

Pro Arg Arg Arg Tyr Arg Arg Arg Arg Thr Val Arg Arg Arg Gly Arg
        35                  40                  45

Gly Arg Trp Thr Arg Ala His Arg Arg Trp Arg Arg Lys Gly Lys Arg
    50                  55                  60

Ser Arg Lys Lys Lys
65
```

<210> 939
<211> 200
<212> PRT
<213> Torque teno virus

<400> 939

```
Ile Ile Ile Arg Gln Trp Gln Pro Asn Tyr Thr Arg Arg Cys Asn Ile
1               5                   10                  15

Val Gly Tyr Met Pro Leu Leu Ile Cys Gly Glu Asn Thr Val Ala Thr
            20                  25                  30

Asn Tyr Ala Thr His Ser Asp Asp Ser Tyr Tyr Pro Gly Pro Phe Gly
        35                  40                  45

Gly Gly Met Thr Thr Asp Lys Phe Thr Leu Arg Ile Leu Tyr Asp Glu
    50                  55                  60

Tyr Lys Arg Phe Met Asn Tyr Trp Thr Ser Ser Asn Glu Asp Leu Asp
65                  70                  75                  80

Leu Cys Arg Tyr Leu Gly Cys Thr Leu Tyr Val Phe Arg His Pro Glu
                85                  90                  95

Val Asp Phe Ile Ile Ile Ile Asn Thr Ser Pro Pro Phe Leu Asp Thr
            100                 105                 110

Glu Ile Thr Gly Pro Ser Ile His Pro Gly Met Met Ala Leu Asn Lys
        115                 120                 125
```

Arg Ser Arg Trp Ile Pro Ser Ile Lys Asn Arg Pro Gly Arg Lys His
130 135 140

Tyr Ile Lys Ile Lys Val Gly Ala Pro Arg Met Phe Thr Asp Lys Trp
145 150 155 160

Tyr Pro Gln Thr Asp Leu Cys Asp Met Thr Leu Leu Thr Ile Phe Ala
165 170 175

Ser Ala Ala Asp Met Gln Tyr Pro Phe Gly Ser Pro Leu Thr Asp Thr
180 185 190

Ile Val Val Ser Phe Gln Val Leu
195 200

<210> 940
<211> 155
<212> PRT
<213> Torque teno virus

<400> 940
Gln Ser Met Tyr Asn Asp Cys Leu Ser Val Leu Pro Asp Asn Phe Ala
1 5 10 15

Glu Thr Ser Gly Lys Gly Thr Gln Leu His Glu Asn Ile Ile Gln His
20 25 30

Leu Pro Tyr Tyr Asn Thr Thr Gln Thr Gln Ala Gln Phe Lys Arg Phe
35 40 45

Ile Glu Asn Met Asn Ala Thr Asn Gly Asp Asn Ile Trp Ala Ser Tyr
50 55 60

Ile Asn Thr Thr Lys Phe Ser Ser Ala Asn Thr Pro Lys Asn Asp Thr
65 70 75 80

Gly Ile Gly Gly Pro Tyr Thr Thr Tyr Ser Asp Ser Trp Tyr Lys Gly
85 90 95

Thr Val Tyr Asn Asp Lys Ile Lys Thr Ile Pro Ile Lys Ala Ser Lys
100 105 110

```
Leu Tyr Tyr Glu Gln Thr Lys Asn Leu Ile Gly Ile Thr Phe Thr Gly
        115                 120                 125


Ser Thr His Arg Leu His Tyr Cys Gly Gly Leu Tyr Ser Ser Val Trp
    130                 135                 140


Leu Ser Ala Gly Arg Ser Tyr Phe Glu Thr Lys
145                 150                 155
```

<210> 941
<211> 160
<212> PRT
<213> Torque teno virus

<400> 941

```
Gly Pro Tyr Thr Asp Ile Thr Tyr Asn Pro Phe Ser Asp Arg Gly Glu
1               5                   10                  15


Gly Asn Met Leu Trp Ile Asp Trp Leu Thr Lys Asn Asp Ser Val Tyr
            20                  25                  30


Ser Lys Thr Ser Ser Lys Cys Leu Ile Glu Asn Leu Pro Leu Trp Ala
        35                  40                  45


Ser Val Tyr Gly Tyr Lys Glu Tyr Cys Ser Lys Val Thr Gly Asp Thr
    50                  55                  60


Asn Ile Glu His Asn Cys Arg Cys Val Ile Arg Ser Pro Tyr Thr Val
65                  70                  75                  80


Pro Gln Leu Leu Asp His Asn Asn Pro Phe Arg Gly Tyr Val Pro Tyr
                85                  90                  95


Ser Phe Asn Phe Gly Asn Gly Lys Met Pro Gly Gly Ser Ser Leu Val
            100                 105                 110


Pro Ile Arg Met Arg Ala Lys Trp Tyr Pro Thr Leu Phe His Gln Lys
        115                 120                 125


Glu Val Leu Glu Ala Ile Ala Gln Ala Gly Pro Phe Ala Tyr His Ser
    130                 135                 140
```

```
Asp Ile Lys Lys Val Ser Leu Gly Ile Lys Tyr Arg Phe Lys Trp Val
145                 150                 155                 160


<210> 942
<211> 183
<212> PRT
<213> Torque teno virus

<400> 942
Trp Gly Gly Asn Pro Val Ser Gln Gln Val Val Arg Asn Pro Cys Lys
1               5                   10                  15


Thr Thr Gln Gly Ser Ser Gly Asn Arg Val Pro Arg Ser Ile Gln Val
            20                  25                  30


Val Asp Pro Arg Tyr Asn Thr Pro Glu Leu Thr Ile His Ala Trp Asp
        35                  40                  45


Phe Arg His Gly Phe Phe Gly Arg Lys Ala Ile Lys Arg Met Gln Glu
    50                  55                  60


Gln Pro Ile Pro His Asp Thr Phe Ser Ala Gly Phe Lys Arg Ser Arg
65                  70                  75                  80


Arg Asp Thr Glu Ala Leu Gln Cys Ser Gln Glu Glu Gln Gln Lys Glu
                85                  90                  95


Asn Leu Leu Phe Pro Val Gln Gln Leu Lys Arg Val Pro Pro Trp Glu
            100                 105                 110


Thr Ser Gln Glu Ser Gln Ser Glu Glu Glu Asn Ser Gln Lys Gln Glu
        115                 120                 125


Thr Leu Ser Gln Gln Leu Arg Asp Gln Leu His Lys Gln Arg Leu Met
    130                 135                 140


Gly Glu Gln Leu Arg Ser Leu Leu Tyr Gln Met Gln Arg Val Gln Gln
145                 150                 155                 160


Asn Gln His Ile Asn Pro Met Leu Leu Pro Lys Gly Leu Ala Leu Thr
                165                 170                 175
```

Ser Ile Ser His Asn Val Ile
            180

<210> 943
<211> 766
<212> PRT
<213> Torque teno virus

<400> 943
Met Ala Trp Arg Trp Trp Trp Gln Arg Arg Trp Arg Arg Arg Arg Trp
1   5   10   15

Pro Arg Arg Arg Trp Arg Arg Leu Arg Arg Arg Arg Pro Arg Arg Pro
            20   25   30

Val Arg Arg Arg Arg Arg Arg Thr Thr Val Arg Arg Arg Arg Trp Arg
        35   40   45

Gly Arg Arg Gly Arg Arg Thr Tyr Thr Arg Arg Ala Val Arg Arg Arg
    50   55   60

Arg Arg Pro Arg Lys Arg Leu Val Leu Thr Gln Trp Ser Pro Gln Thr
65   70   75   80

Val Arg Asn Cys Ser Ile Arg Gly Ile Val Pro Met Val Ile Cys Gly
                85   90   95

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

Thr Gln Ile Gln Pro Phe Gly Gly Ser Phe Ser Thr Thr Thr Trp Ser
        115   120   125

Leu Lys Val Leu Trp Asp Glu His Gln Lys Phe Gln Asn Arg Trp Ser
    130   135   140

Tyr Pro Asn Thr Gln Leu Asp Leu Ala Arg Tyr Arg Gly Val Thr Phe
145   150   155   160

Trp Phe Tyr Arg Asp Gln Lys Thr Asp Tyr Ile Val Gln Trp Ser Arg
                165   170   175

```
Asn Pro Pro Phe Lys Leu Asn Lys Tyr Ser Ser Ala Met Tyr His Pro
            180                 185                 190

Gly Met Met Met Gln Ala Lys Arg Lys Leu Val Val Pro Ser Phe Gln
        195                 200                 205

Thr Arg Pro Lys Gly Lys Lys Arg Tyr Arg Val Thr Ile Lys Pro Pro
    210                 215                 220

Asn Met Phe Ala Asp Lys Trp Tyr Thr Gln Glu Asp Leu Cys Pro Val
225                 230                 235                 240

Pro Leu Val Gln Ile Val Val Ser Ala Ala Ser Leu Leu His Pro Phe
                245                 250                 255

Cys Pro Pro Gln Thr Asn Asn Pro Cys Ile Thr Phe Gln Val Leu Lys
            260                 265                 270

Asp Ile Tyr Asp Glu Cys Ile Gly Val Asn Glu Thr Met Lys Asp Lys
        275                 280                 285

Tyr Lys Lys Leu Gln Thr Thr Leu Tyr Thr Thr Cys Thr Tyr Tyr Gln
    290                 295                 300

Thr Thr Gln Val Leu Ala Gln Leu Ser Pro Ala Phe Gln Pro Ala Met
305                 310                 315                 320

Lys Pro Thr Thr Thr Gln Ser Ala Ala Thr Ala Thr Thr Leu Gly Asn
                325                 330                 335

Tyr Val Pro Glu Leu Lys Tyr Asn Asn Gly Ser Phe His Thr Gly Gln
            340                 345                 350

Asn Ala Val Phe Gly Met Cys Ser Tyr Lys Pro Thr Asp Ser Ile Met
        355                 360                 365

Thr Lys Ala Asn Gly Trp Phe Trp Gln Asn Leu Met Val Asp Asn Asn
    370                 375                 380
```

```
Leu His Ser Ser Tyr Gly Lys Ala Thr Leu Glu Cys Met Glu Tyr His
385                 390                 395                 400

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

Phe Pro Ala Ala Tyr Gln Asp Val Thr Tyr Asn Pro Asn Cys Asp Arg
            420                 425                 430

Ala Val Gly Asn Val Val Trp Phe Gln Tyr Ser Thr Lys Met Asp Thr
        435                 440                 445

Asn Phe Asp Glu Thr Lys Cys Lys Cys Val Leu Lys Asn Ile Pro Leu
    450                 455                 460

Trp Ala Ala Phe Asn Gly Tyr Ser Asp Phe Ile Met Gln Glu Leu Ser
465                 470                 475                 480

Ile Ser Thr Glu Ile His Asn Phe Gly Ile Val Cys Phe Gln Cys Pro
                485                 490                 495

Tyr Thr Phe Pro Pro Cys Phe Asn Lys Asn Lys Pro Leu Lys Gly Tyr
            500                 505                 510

Val Phe Tyr Asp Thr Thr Phe Gly Asn Gly Lys Met Pro Asp Gly Ser
        515                 520                 525

Gly His Val Pro Ile Tyr Trp Gln Gln Arg Trp Trp Ile Arg Leu Ala
    530                 535                 540

Phe Gln Val Gln Val Met His Asp Phe Val Leu Thr Gly Pro Phe Ser
545                 550                 555                 560

Tyr Lys Asp Asp Leu Ala Asn Thr Thr Leu Thr Ala Arg Tyr Lys Phe
                565                 570                 575

Lys Phe Lys Trp Gly Gly Asn Ile Ile Pro Glu Gln Ile Ile Lys Asn
            580                 585                 590

Pro Cys His Arg Glu Gln Ser Leu Ala Ser Tyr Pro Asp Arg Gln Arg
        595                 600                 605
```

```
Arg Asp Leu Gln Val Val Asp Pro Ser Thr Met Gly Pro Ile Tyr Thr
    610                 615                 620

Phe His Thr Trp Asp Trp Arg Arg Gly Leu Phe Gly Ala Asp Ala Ile
625                 630                 635                 640

Gln Arg Val Ser Gln Lys Pro Gly Asp Ala Leu Arg Phe Thr Asn Pro
                645                 650                 655

Phe Lys Arg Pro Arg Tyr Leu Pro Pro Thr Asp Arg Glu Asp Tyr Arg
            660                 665                 670

Gln Glu Glu Asp Phe Ala Leu Gln Glu Lys Arg Arg Arg Thr Ser Thr
        675                 680                 685

Glu Glu Ala Gln Asp Glu Glu Ser Pro Pro Glu Ser Ala Pro Leu Leu
    690                 695                 700

Gln Gln Gln Gln Gln Gln Arg Gln Leu Ser Val His Leu Ala Glu Gln
705                 710                 715                 720

Gln Arg Leu Gly Val Gln Leu Arg Tyr Ile Leu Gln Glu Val Leu Lys
                725                 730                 735

Thr Gln Ala Gly Leu His Leu Asn Pro Leu Leu Leu Gly Pro Pro Gln
            740                 745                 750

Thr Arg Ser Ile Ser Leu Ser Pro Pro Lys Ala Tyr Ser Pro
        755                 760                 765
```

<210> 944
<211> 70
<212> PRT
<213> Torque teno virus

```
<400> 944
Met Ala Trp Arg Trp Trp Trp Gln Arg Arg Trp Arg Arg Arg Arg Trp
1               5                   10                  15

Pro Arg Arg Arg Trp Arg Arg Leu Arg Arg Arg Arg Pro Arg Arg Pro
            20                  25                  30
```

```
Val Arg Arg Arg Arg Arg Arg Thr Thr Val Arg Arg Arg Arg Trp Arg
        35                  40                  45

Gly Arg Arg Gly Arg Arg Thr Tyr Thr Arg Arg Ala Val Arg Arg Arg
    50                  55                  60

Arg Arg Pro Arg Lys Arg
65                  70
```

<210> 945
<211> 201
<212> PRT
<213> Torque teno virus

<400> 945

```
Leu Val Leu Thr Gln Trp Ser Pro Gln Thr Val Arg Asn Cys Ser Ile
1               5                   10                  15

Arg Gly Ile Val Pro Met Val Ile Cys Gly His Thr Lys Ala Gly Arg
            20                  25                  30

Asn Tyr Ala Ile His Ser Glu Asp Phe Thr Thr Gln Ile Gln Pro Phe
        35                  40                  45

Gly Gly Ser Phe Ser Thr Thr Thr Trp Ser Leu Lys Val Leu Trp Asp
    50                  55                  60

Glu His Gln Lys Phe Gln Asn Arg Trp Ser Tyr Pro Asn Thr Gln Leu
65                  70                  75                  80

Asp Leu Ala Arg Tyr Arg Gly Val Thr Phe Trp Phe Tyr Arg Asp Gln
                85                  90                  95

Lys Thr Asp Tyr Ile Val Gln Trp Ser Arg Asn Pro Pro Phe Lys Leu
            100                 105                 110

Asn Lys Tyr Ser Ser Ala Met Tyr His Pro Gly Met Met Met Gln Ala
        115                 120                 125

Lys Arg Lys Leu Val Val Pro Ser Phe Gln Thr Arg Pro Lys Gly Lys
    130                 135                 140
```

```
Lys Arg Tyr Arg Val Thr Ile Lys Pro Pro Asn Met Phe Ala Asp Lys
145                 150                 155                 160


Trp Tyr Thr Gln Glu Asp Leu Cys Pro Val Pro Leu Val Gln Ile Val
                165                 170                 175


Val Ser Ala Ala Ser Leu Leu His Pro Phe Cys Pro Pro Gln Thr Asn
            180                 185                 190


Asn Pro Cys Ile Thr Phe Gln Val Leu
        195                 200


<210> 946
<211> 147
<212> PRT
<213> Torque teno virus

<400> 946
Lys Asp Ile Tyr Asp Glu Cys Ile Gly Val Asn Glu Thr Met Lys Asp
1               5                   10                  15


Lys Tyr Lys Lys Leu Gln Thr Thr Leu Tyr Thr Thr Cys Thr Tyr Tyr
            20                  25                  30


Gln Thr Thr Gln Val Leu Ala Gln Leu Ser Pro Ala Phe Gln Pro Ala
        35                  40                  45


Met Lys Pro Thr Thr Thr Gln Ser Ala Ala Thr Ala Thr Thr Leu Gly
    50                  55                  60


Asn Tyr Val Pro Glu Leu Lys Tyr Asn Asn Gly Ser Phe His Thr Gly
65                  70                  75                  80


Gln Asn Ala Val Phe Gly Met Cys Ser Tyr Lys Pro Thr Asp Ser Ile
                85                  90                  95


Met Thr Lys Ala Asn Gly Trp Phe Trp Gln Asn Leu Met Val Asp Asn
            100                 105                 110


Asn Leu His Ser Ser Tyr Gly Lys Ala Thr Leu Glu Cys Met Glu Tyr
        115                 120                 125
```

His Thr Gly Ile Tyr Ser Ser Ile Phe Leu Ser Pro Gln Arg Ser Leu
130 135 140

Glu Phe Pro
145

<210> 947
<211> 161
<212> PRT
<213> Torque teno virus

<400> 947
Ala Ala Tyr Gln Asp Val Thr Tyr Asn Pro Asn Cys Asp Arg Ala Val
1 5 10 15

Gly Asn Val Val Trp Phe Gln Tyr Ser Thr Lys Met Asp Thr Asn Phe
20 25 30

Asp Glu Thr Lys Cys Lys Cys Val Leu Lys Asn Ile Pro Leu Trp Ala
35 40 45

Ala Phe Asn Gly Tyr Ser Asp Phe Ile Met Gln Glu Leu Ser Ile Ser
50 55 60

Thr Glu Ile His Asn Phe Gly Ile Val Cys Phe Gln Cys Pro Tyr Thr
65 70 75 80

Phe Pro Pro Cys Phe Asn Lys Asn Lys Pro Leu Lys Gly Tyr Val Phe
85 90 95

Tyr Asp Thr Thr Phe Gly Asn Gly Lys Met Pro Asp Gly Ser Gly His
100 105 110

Val Pro Ile Tyr Trp Gln Gln Arg Trp Trp Ile Arg Leu Ala Phe Gln
115 120 125

Val Gln Val Met His Asp Phe Val Leu Thr Gly Pro Phe Ser Tyr Lys
130 135 140

Asp Asp Leu Ala Asn Thr Thr Leu Thr Ala Arg Tyr Lys Phe Lys Phe
145 150 155 160

Lys

<210> 948
<211> 187
<212> PRT
<213> Torque teno virus

<400> 948
Trp Gly Gly Asn Ile Ile Pro Glu Gln Ile Ile Lys Asn Pro Cys His
1   5   10   15

Arg Glu Gln Ser Leu Ala Ser Tyr Pro Asp Arg Gln Arg Arg Asp Leu
            20   25   30

Gln Val Val Asp Pro Ser Thr Met Gly Pro Ile Tyr Thr Phe His Thr
        35   40   45

Trp Asp Trp Arg Arg Gly Leu Phe Gly Ala Asp Ala Ile Gln Arg Val
    50   55   60

Ser Gln Lys Pro Gly Asp Ala Leu Arg Phe Thr Asn Pro Phe Lys Arg
65   70   75   80

Pro Arg Tyr Leu Pro Pro Thr Asp Arg Glu Asp Tyr Arg Gln Glu Glu
                85   90   95

Asp Phe Ala Leu Gln Glu Lys Arg Arg Arg Thr Ser Thr Glu Glu Ala
            100   105   110

Gln Asp Glu Glu Ser Pro Pro Glu Ser Ala Pro Leu Leu Gln Gln Gln
        115   120   125

Gln Gln Gln Arg Gln Leu Ser Val His Leu Ala Glu Gln Gln Arg Leu
    130   135   140

Gly Val Gln Leu Arg Tyr Ile Leu Gln Glu Val Leu Lys Thr Gln Ala
145   150   155   160

Gly Leu His Leu Asn Pro Leu Leu Leu Gly Pro Pro Gln Thr Arg Ser
                165   170   175

Ile Ser Leu Ser Pro Pro Lys Ala Tyr Ser Pro
180 185

<210> 949
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<220>
<221> VARIANT
<222> (1)..(1)
<223> /replace="Phe"

<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

<220>
<221> SITE
<222> (1)..(21)
<223> /note="Variant residues given in the sequence have no preference with respect to those in the annotations for variant positions"

<400> 949
Trp 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
<211> 203
<212> PRT
<213> Torque teno virus

<400> 950
Lys Ile Ile Ile Arg Gln Trp Gln Pro Asn Tyr Thr Arg Arg Cys Asn
1 5 10 15

Ile Val Gly Tyr Met Pro Leu Leu Ile Cys Gly Glu Asn Thr Val Ala
20 25 30

Thr Asn Tyr Ala Thr His Ser Asp Asp Ser Tyr Tyr Pro Gly Pro Phe
35 40 45

Gly Gly Gly Met Thr Thr Asp Lys Phe Thr Leu Arg Ile Leu Tyr Asp
50 55 60

Glu Tyr Lys Arg Phe Met Asn Tyr Trp Thr Ser Ser Asn Glu Asp Leu
65 70 75 80

Asp Leu Cys Arg Tyr Leu Gly Cys Thr Leu Tyr Val Phe Arg His Pro
85 90 95

Glu Val Asp Phe Ile Ile Ile Ile Asn Thr Ser Pro Pro Phe Leu Asp
100 105 110

Thr Glu Ile Thr Gly Pro Ser Ile His Pro Gly Met Met Ala Leu Asn
115 120 125

Lys Arg Ser Arg Trp Ile Pro Ser Ile Lys Asn Arg Pro Gly Arg Lys
130 135 140

His Tyr Ile Lys Ile Lys Val Gly Ala Pro Arg Met Phe Thr Asp Lys
145 150 155 160

Trp Tyr Pro Gln Thr Asp Leu Cys Asp Met Thr Leu Leu Thr Ile Phe
165 170 175

Ala Ser Ala Ala Asp Met Gln Tyr Pro Phe Gly Ser Pro Leu Thr Asp
180 185 190

Thr Ile Val Val Ser Phe Gln Val Leu Gln Ser
195 200

<210> 951
<211> 203
<212> PRT
<213> Torque teno virus

<400> 951
Lys Ile Ile Ile Arg Gln Trp Gln Pro Asn Tyr Arg Arg Arg Cys Asn
1 5 10 15

Ile Val Gly Tyr Leu Pro Ile Leu Ile Cys Gly Gly Asn Thr Val Ser
20 25 30

Arg Asn Tyr Ala Thr His Ser Asp Asp Thr Asn Tyr Pro Gly Pro Phe
35 40 45

Gly Gly Gly Met Thr Thr Asp Lys Phe Ser Leu Arg Ile Leu Tyr Asp
50 55 60

Glu Tyr Lys Arg Phe Met Asn Tyr Trp Thr Ala Ser Asn Glu Asp Leu
65 70 75 80

Asp Leu Cys Arg Tyr Leu Gly Cys Thr Phe Tyr Phe Phe Arg His Pro
85 90 95

Glu Val Asp Phe Ile Ile Lys Ile Asn Thr Met Pro Pro Phe Leu Asp
100 105 110

Thr Thr Ile Thr Ala Pro Ser Ile His Pro Gly Leu Met Ala Leu Asp
115 120 125

Lys Arg Ala Arg Trp Ile Pro Ser Leu Lys Asn Arg Pro Gly Lys Lys
130 135 140

His Tyr Ile Lys Ile Arg Val Gly Ala Pro Lys Met Phe Thr Asp Lys
145 150 155 160

Trp Tyr Pro Gln Thr Asp Leu Cys Asp Met Thr Leu Leu Thr Ile Tyr
165 170 175

Ala Thr Ala Ala Asp Met Gln Tyr Pro Phe Gly Ser Pro Leu Thr Asp
180 185 190

Thr Val Val Val Asn Ser Gln Val Leu Gln Ser
195 200

<210> 952
<211> 203
<212> PRT
<213> Torque teno virus

<400> 952
Lys Ile Ile Ile Lys Gln Trp Gln Pro Asn Phe Ile Arg Arg Cys Tyr
1 5 10 15

Ile Ile Gly Cys Leu Pro Leu Val Phe Cys Gly Glu Asn Thr Thr Ala
20 25 30

Gln Asn Tyr Ala Thr His Ser Asp Asp Met Ile Ser Lys Gly Pro Tyr
35 40 45

Gly Gly Gly Met Thr Thr Thr Lys Phe Thr Leu Arg Ile Leu Tyr Asp
50 55 60

Glu Phe Thr Arg Phe Met Asn Phe Trp Thr Val Ser Asn Glu Asp Leu
65 70 75 80

Asp Leu Cys Arg Tyr Val Gly Cys Lys Leu Ile Phe Phe Lys His Pro
85 90 95

Thr Val Asp Phe Met Val Gln Ile Asn Thr Gln Pro Pro Phe Leu Asp
100 105 110

Thr Ser Leu Thr Ala Ala Ser Ile His Pro Gly Ile Met Met Leu Ser
115 120 125

Lys Arg Arg Ile Leu Ile Pro Ser Leu Lys Thr Arg Pro Ser Arg Lys
130 135 140

```
His Arg Val Val Val Arg Val Gly Ala Pro Arg Leu Phe Gln Asp Lys
145                 150                 155                 160


Trp Tyr Pro Gln Ser Asp Leu Cys Asp Thr Val Leu Leu Ser Ile Phe
                165                 170                 175


Ala Thr Ala Arg Asp Leu Gln Tyr Pro Phe Gly Ser Pro Leu Thr Asp
            180                 185                 190


Asn Pro Cys Val Asn Phe Gln Ile Leu Gly Pro
        195                 200
```

<210> 953
<211> 203
<212> PRT
<213> Torque teno virus

<400> 953

```
Lys Ile Ile Ile Lys Gln Trp Gln Pro Asn Phe Ile Arg Arg Cys Tyr
1               5                   10                  15


Ile Ile Gly Tyr Leu Pro Leu Ile Phe Cys Gly Glu Asn Thr Thr Ala
            20                  25                  30


Asn Asn Phe Ala Thr His Ser Asp Asp Met Ile Ala Lys Gly Pro Trp
        35                  40                  45


Gly Gly Gly Met Thr Thr Thr Lys Phe Thr Leu Arg Ile Leu Tyr Asp
    50                  55                  60


Glu Phe Thr Arg Phe Met Asn Phe Trp Thr Val Ser Asn Glu Asp Leu
65                  70                  75                  80


Asp Leu Cys Arg Tyr Val Ser Cys Lys Leu Ile Phe Phe Lys His Pro
                85                  90                  95


Thr Val Asp Phe Ile Val Arg Ile Asn Thr Glu Pro Pro Phe Leu Asp
            100                 105                 110


Thr Asn Leu Thr Ala Ala Gln Ile His Pro Gly Ile Met Met Leu Ser
        115                 120                 125
```

```
Lys Lys His Ile Leu Ile Pro Ser Leu Lys Thr Arg Pro Ser Arg Lys
    130                 135                 140

His Arg Val Val Val Arg Val Gly Pro Pro Arg Leu Phe Gln Asp Lys
145                 150                 155                 160

Trp Tyr Pro Gln Ser Asp Leu Cys Asp Thr Val Leu Leu Ser Val Phe
                165                 170                 175

Ala Thr Ala Cys Asp Leu Gln Tyr Pro Phe Gly Ser Pro Leu Thr Asp
            180                 185                 190

Asn Pro Cys Val Asn Phe Gln Ile Leu Gly His
        195                 200
```

<210> 954
<211> 203
<212> PRT
<213> Torque teno virus

<400> 954

```
Lys Leu Ile Ile Lys Gln Trp Gln Pro Asn Phe Ile Arg His Cys Tyr
1               5                   10                  15

Ile Ile Gly Tyr Met Pro Leu Ile Ile Cys Gly Glu Asn Thr Phe Ser
            20                  25                  30

His Asn Tyr Ala Thr His Ser Asp Asp Met Leu Ser Thr Gly Pro Tyr
        35                  40                  45

Gly Gly Gly Met Thr Thr Thr Lys Phe Thr Leu Arg Ile Leu Phe Asp
    50                  55                  60

Glu Tyr Gln Arg His Leu Asn Phe Trp Thr Val Ser Asn Gln Asp Leu
65                  70                  75                  80

Asp Leu Ala Arg Tyr Leu Gly Thr Lys Ile Ile Phe Phe Arg His Pro
                85                  90                  95

Thr Val Asp Phe Val Val Gln Ile His Thr Gln Pro Pro Phe Gln Asp
            100                 105                 110
```

Thr Glu Ile Thr Ala Pro Ser Ile His Pro Gly Met Leu Ile Leu Ser
115 120 125

Lys Lys His Ile Leu Ile Pro Ser Leu Lys Thr Arg Pro Ser Lys Lys
130 135 140

His Tyr Val Lys Val Arg Val Gly Ala Pro Arg Leu Phe Gln Asp Lys
145 150 155 160

Trp Tyr Pro Gln Ser Glu Leu Cys Asp Val Thr Leu Leu Val Ile Tyr
165 170 175

Ala Thr Ala Cys Asp Leu Gln Tyr Pro Phe Gly Ser Pro Gln Thr Asp
180 185 190

Asn Val Cys Val Asn Phe Gln Ile Leu Gly Gln
195 200

<210> 955
<211> 204
<212> PRT
<213> Torque teno virus

<400> 955
Thr Leu Ile Leu Arg Gln Trp Gln Pro Asp Cys Ile Arg His Cys Lys
1 5 10 15

Ile Thr Gly Trp Met Pro Leu Ile Ile Cys Gly Lys Gly Ser Thr Gln
20 25 30

Phe Asn Tyr Ile Thr His Ala Asp Asp Ile Thr Pro Arg Gly Ala Ser
35 40 45

Tyr Gly Gly Asn Phe Thr Asn Met Thr Phe Ser Leu Glu Ala Ile Tyr
50 55 60

Glu Gln Phe Leu Tyr His Arg Asn Arg Trp Ser Ala Ser Asn His Asp
65 70 75 80

Leu Glu Leu Cys Arg Tyr Lys Gly Thr Thr Leu Lys Leu Tyr Arg His
85 90 95

```
Pro Glu Val Asp Tyr Ile Val Thr Tyr Ser Arg Thr Gly Pro Phe Glu
            100                 105                 110

Ile Ser His Met Thr Tyr Leu Ser Thr His Pro Met Leu Met Leu Leu
        115                 120                 125

Asn Lys His His Ile Val Val Pro Ser Leu Lys Thr Lys Pro Arg Gly
    130                 135                 140

Arg Lys Ala Ile Lys Val Arg Ile Arg Pro Pro Lys Leu Met Asn Asn
145                 150                 155                 160

Lys Trp Tyr Phe Thr Arg Asp Phe Cys Asn Ile Gly Leu Phe Gln Leu
                165                 170                 175

Trp Ala Thr Gly Leu Glu Leu Arg Asn Pro Trp Leu Arg Met Ser Thr
            180                 185                 190

Leu Ser Pro Cys Ile Gly Phe Asn Val Leu Lys Asn
        195                 200
```

<210> 956
<211> 204
<212> PRT
<213> Torque teno virus

<400> 956

```
Thr Leu Val Leu Arg Gln Trp Gln Pro Asp Val Ile Arg His Cys Lys
1               5                   10                  15

Ile Thr Gly Arg Met Pro Leu Ile Ile Cys Gly Lys Gly Ser Thr Gln
            20                  25                  30

Phe Asn Tyr Ile Thr His Ala Asp Asp Ile Thr Pro Arg Gly Ala Ser
        35                  40                  45

Tyr Gly Gly Asn Phe Thr Asn Met Thr Phe Ser Leu Glu Ala Ile Tyr
    50                  55                  60

Glu Gln Phe Leu Tyr His Arg Asn Arg Trp Ser Ala Ser Asn His Asp
65                  70                  75                  80
```

```
Leu Glu Leu Cys Arg Tyr Lys Gly Thr Thr Leu Lys Leu Tyr Arg His
                85                  90                  95


Pro Asp Val Asp Tyr Ile Val Thr Tyr Ser Arg Thr Gly Pro Phe Glu
            100                 105                 110


Ile Ser His Met Thr Tyr Leu Ser Thr His Pro Leu Leu Met Leu Leu
        115                 120                 125


Asn Lys His His Ile Val Val Pro Ser Leu Lys Thr Lys Pro Arg Gly
    130                 135                 140


Arg Lys Ala Ile Lys Val Arg Ile Arg Pro Pro Lys Leu Met Asn Asn
145                 150                 155                 160


Lys Trp Tyr Phe Thr Arg Asp Phe Cys Asn Ile Gly Leu Phe Gln Leu
                165                 170                 175


Trp Ala Thr Gly Leu Glu Leu Arg Asn Pro Trp Leu Arg Met Ser Thr
            180                 185                 190


Leu Ser Pro Cys Ile Gly Phe Asn Val Leu Lys Asn
        195                 200


<210> 957
<211> 203
<212> PRT
<213> Torque teno virus

<400> 957
Lys Leu Val Leu Thr Gln Trp Asn Pro Gln Thr Val Arg Lys Cys Val
1               5                   10                  15


Ile Arg Gly Phe Leu Pro Leu Phe Phe Cys Gly Gln Gly Ala Tyr His
            20                  25                  30


Arg Asn Phe Thr Asp His Tyr Asp Asp Val Phe Pro Lys Gly Pro Ser
        35                  40                  45


Gly Gly Gly His Gly Ser Met Val Phe Asn Leu Ser Phe Leu Tyr Gln
    50                  55                  60
```

```
Glu Phe Lys Lys His His Asn Lys Trp Ser Arg Ser Asn Leu Asp Phe
65                  70                  75                  80


Asp Leu Val Arg Tyr Lys Gly Thr Val Ile Lys Leu Tyr Arg His Gln
                85                  90                  95


Asp Phe Asp Tyr Ile Val Trp Ile Ser Arg Thr Pro Pro Phe Gln Glu
            100                 105                 110


Ser Leu Leu Thr Val Met Thr His Gln Pro Ser Val Met Leu Gln Ala
        115                 120                 125


Lys Lys Cys Ile Ile Val Lys Ser Tyr Arg Thr His Pro Gly Gly Lys
    130                 135                 140


Pro Tyr Val Thr Ala Lys Val Arg Pro Pro Arg Leu Leu Thr Asp Lys
145                 150                 155                 160


Trp Tyr Phe Gln Ser Asp Phe Cys Asn Val Pro Leu Phe Ser Leu Gln
                165                 170                 175


Phe Ala Leu Ala Glu Leu Arg Phe Pro Ile Cys Ser Pro Gln Thr Asp
            180                 185                 190


Thr Asn Cys Ile Asn Phe Leu Val Leu Asp Asp
        195                 200


<210> 958
<211> 203
<212> PRT
<213> Torque teno virus

<400> 958
Lys Leu Ile Ile Lys Leu Trp Gln Pro Ala Val Ile Lys Arg Cys Arg
1               5                   10                  15


Ile Lys Gly Tyr Ile Pro Leu Ile Ile Ser Gly Asn Gly Thr Phe Ala
            20                  25                  30


Thr Asn Phe Thr Ser His Ile Asn Asp Arg Ile Met Lys Gly Pro Phe
        35                  40                  45
```

```
Gly Gly Gly His Ser Thr Met Arg Phe Ser Leu Tyr Ile Leu Phe Glu
    50                  55                  60

Glu His Leu Arg His Met Asn Phe Trp Thr Arg Ser Asn Asp Asn Leu
65                  70                  75                  80

Glu Leu Thr Arg Tyr Leu Gly Ala Ser Val Lys Ile Tyr Arg His Pro
                85                  90                  95

Asp Gln Asp Phe Ile Val Ile Tyr Asn Arg Arg Thr Pro Leu Gly Gly
            100                 105                 110

Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly Asn Ala Ile Leu Ala
        115                 120                 125

Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr Arg Pro Lys Gly Arg
    130                 135                 140

Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr Leu Phe Thr Asp Lys
145                 150                 155                 160

Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr Leu Phe Asn Ile Met
                165                 170                 175

Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys Ser Pro Gln Thr Asp
            180                 185                 190

Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser
        195                 200
```

<210> 959
<211> 204
<212> PRT
<213> Torque teno virus

<400> 959

```
Arg Leu Val Leu Thr Gln Trp Ser Pro Gln Thr Val Arg Asn Cys Ser
1               5                   10                  15

Ile Arg Gly Ile Val Pro Met Val Ile Cys Gly His Thr Lys Ala Gly
            20                  25                  30
```

```
Arg Asn Tyr Ala Ile His Ser Glu Asp Phe Thr Thr Gln Ile Gln Pro
        35                  40                  45

Phe Gly Gly Ser Phe Ser Thr Thr Thr Trp Ser Leu Lys Val Leu Trp
    50                  55                  60

Asp Glu His Gln Lys Phe Gln Asn Arg Trp Ser Tyr Pro Asn Thr Gln
65                  70                  75                  80

Leu Asp Leu Ala Arg Tyr Arg Gly Val Thr Phe Trp Phe Tyr Arg Asp
                85                  90                  95

Gln Lys Thr Asp Tyr Ile Val Gln Trp Ser Arg Asn Pro Pro Phe Lys
            100                 105                 110

Leu Asn Lys Tyr Ser Ser Ala Met Tyr His Pro Gly Met Met Met Gln
        115                 120                 125

Ala Lys Arg Lys Leu Val Val Pro Ser Phe Gln Thr Arg Pro Lys Gly
    130                 135                 140

Lys Lys Arg Tyr Arg Val Thr Ile Lys Pro Pro Asn Met Phe Ala Asp
145                 150                 155                 160

Lys Trp Tyr Thr Gln Glu Asp Leu Cys Pro Val Pro Leu Val Gln Ile
                165                 170                 175

Val Val Ser Ala Ala Ser Leu Leu His Pro Phe Cys Pro Pro Gln Thr
            180                 185                 190

Asn Asn Pro Cys Ile Thr Phe Gln Val Leu Lys Asp
        195                 200
```

<210> 960
<211> 205
<212> PRT
<213> Torque teno virus

<400> 960

```
Lys Leu Val Leu Thr Gln Trp His Pro Asn Thr Met Arg Arg Cys Leu
1               5                   10                  15
```

```
Ile Lys Gly Ile Val Pro Leu Val Ile Cys Gly His Thr Arg Trp Asn
            20                  25                  30

Tyr Asn Tyr Ala Leu His Ser Lys Asp Tyr Thr Glu Glu Gly Arg Tyr
        35                  40                  45

Pro His Gly Gly Ala Leu Ser Thr Thr Thr Trp Ser Leu Lys Val Leu
    50                  55                  60

Tyr Asp Glu His Leu Lys His His Asp Phe Trp Gly Tyr Pro Asn Asn
65                  70                  75                  80

Gln Leu Asp Leu Ala Arg Tyr Lys Gly Ala Lys Phe Thr Phe Tyr Arg
                85                  90                  95

His Lys Lys Thr Asp Phe Ile Ile Phe Phe Asn Arg Lys Pro Pro Phe
            100                 105                 110

Lys Leu Asn Lys Tyr Ser Cys Ala Ser Tyr His Pro Gly Met Leu Met
        115                 120                 125

Gln Gln Arg His Lys Ile Leu Leu Pro Ser Tyr Glu Thr Lys Pro Lys
    130                 135                 140

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

Asp Lys Trp Tyr Thr Gln Gln Asp Leu Cys Asp Val Asn Leu Leu Gln
                165                 170                 175

Leu Val Val Thr Ala Ala Asp Phe Arg His Pro Leu Cys Ser Pro Gln
            180                 185                 190

Thr Asn Thr Pro Thr Thr Thr Phe Gln Val Leu Lys Asp
        195                 200                 205
```

<210> 961
<211> 204
<212> PRT
<213> Torque teno virus

<400> 961

```
Arg Leu Ile Leu Arg Gln Trp Gln Pro Ala Thr Arg Arg Arg Cys Thr
1               5                   10                  15

Ile Thr Gly Tyr Leu Pro Ile Val Phe Cys Gly His Thr Arg Gly Asn
            20                  25                  30

Lys Asn Tyr Ala Leu His Ser Asp Asp Tyr Thr Pro Gln Gly Gln Pro
        35                  40                  45

Phe Gly Gly Ala Leu Ser Thr Thr Ser Phe Ser Leu Lys Val Leu Phe
    50                  55                  60

Asp Gln His Gln Arg Gly Leu Asn Lys Trp Ser Phe Pro Asn Asp Gln
65                  70                  75                  80

Leu Asp Leu Ala Arg Tyr Arg Gly Cys Lys Phe Ile Phe Tyr Arg Thr
                85                  90                  95

Lys Gln Thr Asp Trp Val Gly Gln Tyr Asp Ile Ser Glu Pro Tyr Lys
            100                 105                 110

Leu Asp Lys Tyr Ser Cys Pro Asn Tyr His Pro Gly Asn Met Ile Lys
        115                 120                 125

Ala Lys His Lys Phe Leu Ile Pro Ser Tyr Asp Thr Asn Pro Arg Gly
    130                 135                 140

Arg Gln Lys Ile Ile Val Lys Ile Pro Pro Pro Asp Leu Phe Val Asp
145                 150                 155                 160

Lys Trp Tyr Thr Gln Glu Asp Leu Cys Ser Val Asn Leu Val Ser Leu
                165                 170                 175

Ala Val Ser Ala Ala Ser Phe Leu His Pro Phe Gly Ser Pro Gln Thr
            180                 185                 190

Asp Asn Pro Cys Tyr Thr Phe Gln Val Leu Lys Glu
        195                 200
```

<210> 962
<211> 204

<212> PRT
<213> Torque teno virus

<400> 962
Lys Leu Ile Leu Thr Gln Trp Asn Pro Ala Ile Val Lys Arg Cys Asn
1   5   10   15

Ile Lys Gly Gly Leu Pro Ile Ile Ile Cys Gly Glu Pro Arg Ala Ala
            20   25   30

Phe Asn Tyr Gly Tyr His Met Glu Asp Tyr Thr Pro Gln Pro Phe Pro
        35   40   45

Phe Gly Gly Gly Met Ser Thr Val Thr Phe Ser Leu Lys Ala Leu Tyr
    50   55   60

Asp Gln Tyr Leu Lys His Gln Asn Arg Trp Thr Phe Ser Asn Asp Gln
65   70   75   80

Leu Asp Leu Ala Arg Tyr Arg Gly Cys Lys Leu Arg Phe Tyr Arg Ser
                85   90   95

Pro Val Cys Asp Phe Ile Val His Tyr Asn Leu Ile Pro Pro Leu Lys
            100   105   110

Met Asn Gln Phe Thr Ser Pro Asn Thr His Pro Gly Leu Leu Met Leu
        115   120   125

Ser Lys His Lys Ile Ile Ile Pro Ser Phe Gln Thr Arg Pro Gly Gly
    130   135   140

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

Lys Trp Tyr Thr Gln Gln Asp Leu Cys Lys Val Pro Leu Val Ser Ile
                165   170   175

Thr Ala Thr Ala Ala Asp Leu Arg Tyr Pro Phe Cys Ser Pro Gln Thr
            180   185   190

Asn Asn Pro Cys Thr Thr Phe Gln Val Leu Arg Lys
        195   200

<210> 963
<211> 203
<212> PRT
<213> Torque teno virus

<400> 963
Lys Ile Val Leu Thr Gln Trp Asn Pro Gln Thr Thr Arg Lys Cys Ile
1   5   10   15

Ile Arg Gly Met Met Pro Val Leu Trp Ala Gly Met Gly Thr Gly Gly
            20   25   30

Arg Asn Tyr Ala Val Arg Ser Asp Asp Tyr Val Val Asn Lys Gly Phe
        35   40   45

Gly Gly Ser Phe Ala Thr Glu Thr Phe Ser Leu Lys Val Leu Tyr Asp
    50   55   60

Gln Phe Gln Arg Gly Phe Asn Arg Trp Ser His Thr Asn Glu Asp Leu
65   70   75   80

Asp Leu Ala Arg Tyr Arg Gly Cys Arg Trp Thr Phe Tyr Arg His Lys
                85   90   95

Asp Thr Asp Phe Ile Val Tyr Phe Thr Asn Asn Pro Pro Met Lys Thr
            100   105   110

Asn Gln Phe Ser Ala Pro Leu Thr Thr Pro Gly Met Leu Met Arg Ser
        115   120   125

Lys Tyr Lys Val Leu Ile Pro Ser Phe Gln Thr Arg Pro Lys Gly Arg
    130   135   140

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

Trp Tyr Thr Gln Gln Asp Leu Cys Ser Val Pro Leu Val Gln Leu Asn
                165   170   175

Val Thr Ala Ala Asp Phe Thr His Pro Phe Gly Ser Pro Leu Thr Glu
            180   185   190

Thr Pro Cys Val Glu Phe Gln Val Leu Gly Asp
195 200

<210> 964
<211> 204
<212> PRT
<213> Torque teno mini virus

<400> 964
Lys Leu Asn Ile Lys Glu Trp Gln Pro Ile Thr Ile Arg Lys Thr Lys
1 5 10 15

Val Lys Gly Leu Tyr Pro Ala Phe Leu Cys Asn Asp Gln Arg Ile Gly
20 25 30

Asn Asn Ala Ile Gln Tyr Leu Asp Ser Ile Ala Pro His His Phe Pro
35 40 45

Gly Gly Gly Gly Phe Gly Ile Ile Gln Phe Thr Leu Gln Gly Leu Tyr
50 55 60

Glu Gln Phe Ile Lys Ala Thr Asn Trp Trp Thr Gln Thr Asn Cys Ser
65 70 75 80

Leu Pro Leu Ile Arg Tyr Asn Phe Cys Lys Leu Lys Phe Tyr Arg Thr
85 90 95

Ala Asn Val Asp Tyr Val Val Lys Ile Ile Arg Cys Tyr Pro Leu Lys
100 105 110

Ala Thr His Asp Leu Tyr Met Gln Thr Gln Pro Ser Ile Leu Met Arg
115 120 125

Asp Lys His Ser Ile Leu Ile Pro Cys Met Lys Asn Gly Arg Asn Arg
130 135 140

Lys Thr Tyr Lys Thr Ile Lys Val Arg Pro Pro Thr Gln Met Thr Asn
145 150 155 160

Gly Trp Phe Phe Gln Lys Asp Leu Cys Asn Phe Pro Leu Leu Val Ile
165 170 175

```
Leu Ala Thr Ala Ala Ser Phe Asp Arg Tyr Tyr Thr Asn Ser Lys Ala
            180                 185                 190


Lys Ser Thr Thr Ile Gly Phe Ile Ser Leu Asn Thr
        195                 200


<210> 965
<211> 204
<212> PRT
<213> Anelloviridae sp.

<400> 965
Lys Ile Arg Ile Ser Glu Trp Gln Pro Thr Lys Ile Ile Lys Ser Lys
1               5                   10                  15


Val Lys Gly Ile Tyr Pro Cys Phe Leu Ser Asn His Lys Arg Leu Ser
            20                  25                  30


Asn Asn Phe Val Gln Trp Ile Asp Ser Thr Thr Ala His Leu Met Pro
        35                  40                  45


Gly Gly Gly Gly Phe Gly Ile Ile Gln Phe Thr Leu Asn Gly Leu Tyr
    50                  55                  60


Glu Leu Phe Cys Lys Ala Gln Asn Trp Trp Thr Lys Ser Asn Cys Asn
65                  70                  75                  80


Leu Pro Leu Val Arg Phe Cys Gly Thr Thr Leu Lys Phe Trp Ala Ala
                85                  90                  95


Glu Asn Tyr Asp Tyr Val Val His Ile Gln Arg Cys Tyr Pro Met Cys
            100                 105                 110


Cys Thr Asp Leu Met Tyr Met Ser Cys Gln Pro Phe Ile Met Met Met
        115                 120                 125


Thr Lys Asn Thr Ile Phe Val Pro Cys Thr Lys Asn Lys Pro Arg Ala
    130                 135                 140


Lys Arg Tyr Lys Lys Ile Phe Val Lys Pro Pro Ala Gln Met Thr Thr
145                 150                 155                 160
```

```
Gln Trp Tyr Phe Gln Ser Gln Leu Ala Lys Thr Gly Leu Ile Ile Ile
                165                 170                 175


Arg Thr Ala Ala Cys Ser Leu Asp Arg Ile Tyr Thr Ser Ser Thr Ala
            180                 185                 190


Ser Ser Thr Thr Ile Gly Leu Val Ser Leu His Thr
        195                 200
```

<210> 966
<211> 204
<212> PRT
<213> Torque teno mini virus

<400> 966

```
Thr Ile Arg Leu Arg Gln Trp Asn Pro Arg Thr Thr Arg Leu Cys Lys
1               5                   10                  15


Ile Lys Gly His Ile Pro Leu Ile Ile Cys Gly Arg Asp Arg Gln Ile
            20                  25                  30


Phe Asn Trp Met Gln Tyr Tyr Asp Ser Ile Gly Pro Val Glu Gln Ser
        35                  40                  45


Gly Gly Gly Gly Trp Ser Ser Ile Val Phe Ser Leu Gly Ala Leu Tyr
    50                  55                  60


Gln Gln Phe Lys Arg Leu Met Asn Trp Trp Thr Lys Asp Asn Asp Gly
65                  70                  75                  80


Leu Pro Leu Val Gln Tyr Arg Gly Cys Lys Phe Lys Phe Tyr Lys Ser
                85                  90                  95


Trp Asp Cys Asp Tyr Ile Val Thr Ala Gln Thr Cys Pro Pro Met Thr
            100                 105                 110


Asp Thr Glu Tyr Lys His Leu Asp Ser His Pro Tyr Arg Gln Leu Met
        115                 120                 125


Asn Lys Arg Pro Ile Ile Val Pro Asn Leu Val Arg Lys Pro Ser Lys
    130                 135                 140
```

```
Lys Thr Tyr Lys Ile Lys Arg Phe Pro Pro Pro Ser Leu Leu Gln Asn
145                 150                 155                 160


Lys Trp Tyr Phe Gln Gln Asp Phe Cys Ala Thr Asn Leu Leu Met Leu
                165                 170                 175


Thr Thr Ser Ala Ala Ser Phe Asp Gln Phe Trp Leu Pro Asn Asp Glu
            180                 185                 190


Ile Ser Gln Val Ile Thr Phe His Ser Leu Asn Thr
        195                 200


<210> 967
<211> 204
<212> PRT
<213> Torque teno mini virus

<400> 967
Lys Ile Ile Leu Lys Gln Phe Gln Pro Gln Thr Ile Lys Arg Val Lys
1               5                   10                  15


Ile Lys Arg Leu Glu Cys Leu Phe His Cys Asn Ala Lys Lys Ile Tyr
            20                  25                  30


Phe Asn Leu Gln Met Tyr Glu Thr Ser Thr Val Pro Pro His Leu Pro
        35                  40                  45


Gly Gly Gly Gly Trp Ser Cys Lys Val Phe Thr Leu Asn Ala Leu Tyr
    50                  55                  60


Asp Ala Phe Lys His Cys Arg Asn Val Trp Thr Gly Ser Asn His Asn
65                  70                  75                  80


Leu Pro Leu Val Arg Tyr Met Gly Cys Lys Phe Thr Leu Tyr Gln Ser
                85                  90                  95


Glu Thr Gln Asp Tyr Val Phe Lys Tyr Gln Asn His Tyr Pro Met Val
            100                 105                 110


Ser Thr Ile Glu Ser Tyr Asn Ala Cys Gln Pro Ser Met Leu Met Met
        115                 120                 125
```

Asp Asn Arg Thr Lys Lys Ile Pro Ser Lys Lys Thr Lys His Lys Arg
130 135 140

Lys Pro Tyr Thr Ile Val Lys Ile Arg Pro Pro Ala Gln Met Gln Asn
145 150 155 160

Lys Trp Tyr Phe Thr His Asp Ile Ala Asn Thr Pro Leu Leu Leu Ile
165 170 175

Tyr Ser Ala Ala Cys Ser Phe Asp Asn Tyr Tyr Ile Ser Thr Asp Ser
180 185 190

Glu Ser Thr Asn Ile Ser Ile Pro Leu Leu Arg Thr
195 200

<210> 968
<211> 204
<212> PRT
<213> Torque teno mini virus

<400> 968
Thr Ile Thr Val Lys Gln Phe Gln Pro Pro Asn Val Arg Lys Cys Lys
1 5 10 15

Ile Lys Gly Leu Met Asn Leu Ile Tyr Phe Asn Gln Lys Arg Leu Ile
20 25 30

Phe Asn Ser Val Met Tyr Glu Asn Ser Leu Val Pro Glu Glu His Pro
35 40 45

Gly Gly Gly Gly Phe Ser Val Ile Lys Leu Ser Leu Glu Thr Met Tyr
50 55 60

Asp Ser His Leu Arg Cys His Asn Trp Trp Thr Thr Ser Asn Glu Asp
65 70 75 80

Leu Pro Leu Val Arg Tyr Ser Gly Met Thr Val Lys Leu Tyr Gln Ser
85 90 95

Lys Tyr Thr Asp Tyr Val Phe Lys Tyr Gln Asn Tyr Leu Pro Gly Thr
100 105 110

Ser Thr Tyr Leu Thr Tyr Pro Ala Cys Gln Pro Ser Met Met Met Met
115 120 125

Ala Lys Asn Ser Val Ile Val Pro Ser Leu Glu Thr Lys Arg Arg Arg
130 135 140

Lys Pro Tyr Lys Lys Ile His Ile Lys Pro Thr Ser Gln Leu Gln Thr
145 150 155 160

Lys Trp Tyr Phe Gln Thr Asp Leu Asn Lys Thr Pro Leu Ala Ile Phe
165 170 175

Tyr Thr Ala Ala Cys Ser Leu Thr Ser Tyr Tyr Leu Ser Pro Asp Trp
180 185 190

Glu Ser Asn Asn Ile Ser Ile Leu His Leu Asn Thr
195 200

<210> 969
<211> 204
<212> PRT
<213> TTV-like mini virus

<400> 969
Thr Ile Pro Leu Lys Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr
1 5 10 15

Ile Lys Gly Gln Asp Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly
20 25 30

Met Asn Ser Thr Met Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro
35 40 45

Gly Gly Gly Ser Phe Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr
50 55 60

Asp Ile His Lys Leu Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp
65 70 75 80

Leu Pro Leu Val Arg Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser
85 90 95

```
Thr Phe Thr Asp Tyr Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn
            100                 105                 110

Ser Asn Lys Leu Thr Tyr Pro Asn Thr His Pro Leu Met Met Met Met
        115                 120                 125

Ser Lys Tyr Lys His Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys
    130                 135                 140

Lys Pro Tyr Thr Lys Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn
145                 150                 155                 160

Lys Trp Tyr Phe Ala Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile
                165                 170                 175

His Cys Thr Ala Cys Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys
            180                 185                 190

Leu Ser Asn Asn Val Thr Leu Trp Ser Leu Asn Thr
        195                 200
```

<210> 970
<211> 204
<212> PRT
<213> Torque teno mini virus

<400> 970

```
Lys Leu Thr Leu Lys Glu Trp Gln Pro Lys Thr Ile Arg Lys Ser Cys
1               5                   10                  15

Ile Lys Gly Leu His Cys Leu Phe Ile Val Thr Glu Asp Thr Ile Ser
            20                  25                  30

Arg Asn Tyr Arg Met Tyr Glu His Ser Phe Thr Gly Glu His Trp Pro
        35                  40                  45

Ser Gly Gly Gly Phe Ser Val Thr Lys Tyr Thr Leu Asp Gly Leu Tyr
    50                  55                  60

Glu Gln His Gln Leu Asp Arg Asn Trp Trp Thr Lys Pro Asn Thr Asn
65                  70                  75                  80
```

```
Met Pro Leu Val Arg Tyr Leu Gly Cys Lys Ile Thr Phe Tyr Gln Ser
                85                  90                  95


Trp Glu Val Asp Tyr Val Cys Asn Ile Asn Leu Thr Trp Pro Met Val
            100                 105                 110


Ala Thr Asn Leu Leu Tyr Leu Ser Cys His Pro Asn Phe Met Met Met
        115                 120                 125


Asn His Lys Ala Ile Phe Val Pro Ser Lys Ile Thr Lys Arg Leu Arg
    130                 135                 140


Arg Gly Lys Lys Thr Val Arg Leu Arg Pro Pro His Gln Met Ile Asn
145                 150                 155                 160


Lys Trp Tyr Phe Ala Lys Glu Leu Ala Lys Thr Gly Leu Val Met Leu
                165                 170                 175


Thr Ala Ala Ala Ala Ser Phe Asp His Tyr Tyr Ile Gly Ser Asp Lys
            180                 185                 190


Leu Ser Asn Asn Cys Thr Phe Thr Ser Leu Asn Pro
        195                 200


<210> 971
<211> 204
<212> PRT
<213> Torque teno midi virus

<400> 971
Lys Ile Pro Ile Tyr Gln Trp Gln Pro Asp Ser Ile Arg Lys Cys Lys
1               5                   10                  15


Ile Lys Gly Val Gly Thr Leu Val Leu Gly Ala His Gly Lys Gln Phe
            20                  25                  30


Val Cys Tyr Thr Asp Val Glu Thr Arg Ala Pro Pro Pro Lys Ala Pro
        35                  40                  45


Gly Gly Gly Gly Phe Gly Cys Lys Gln Phe Ser Leu Gln Tyr Leu Tyr
    50                  55                  60
```

```
Glu Glu Tyr Arg Phe Arg Asn Asn Ile Trp Thr His Thr Asn Ile Asn
65                  70                  75                  80


Leu Asp Leu Val Arg Tyr Leu Arg Ala Ala Phe Thr Phe Tyr Arg His
                85                  90                  95


Pro Asp Ile Asp Phe Ile Ile Asn Tyr Asp Arg Gln Pro Pro Phe Tyr
            100                 105                 110


Leu Asp Lys Phe Thr Tyr Pro Leu Cys His Pro Gln Asn Leu Leu Leu
        115                 120                 125


Gly Lys His Lys Ile Ile Leu Leu Ser Lys Ala Ser Lys Pro Asn Gly
    130                 135                 140


Lys Val Lys Lys Arg Ile Ile Ile Lys Pro Pro Lys Gln Met Ile Thr
145                 150                 155                 160


Lys Trp Phe Phe Gln Glu Gln Phe Thr Thr Gln Pro Leu Leu Ser Leu
                165                 170                 175


Arg Ala Ala Ala Ala Ser Phe Gln Tyr Pro His Ile Gly Cys Cys Met
            180                 185                 190


Pro Asn Arg Val Val Thr Phe Ser Ala Leu Asn Pro
        195                 200


<210> 972
<211> 204
<212> PRT
<213> Torque teno midi virus

<400> 972
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 Asn His
        195                 200
```

<210> 973
<211> 204
<212> PRT
<213> Torque teno midi virus

<400> 973

```
Thr Leu Ile Val Arg Gln Trp Gln Pro Asp Ser Ile Val Leu Cys Lys
1               5                   10                  15

Ile Lys Gly Tyr Asp Ser Ile Ile Trp Gly Ala Glu Gly Thr Gln Phe
            20                  25                  30
```

```
Gln Cys Ser Thr His Glu Met Tyr Glu Tyr Thr Arg Gln Lys Tyr Pro
        35                  40                  45

Gly Gly Gly Gly Phe Gly Val Gln Leu Tyr Ser Leu Glu Tyr Leu Tyr
    50                  55                  60

Asp Gln Trp Lys Leu Arg Asn Asn Ile Trp Thr Lys Thr Asn Gln Leu
65                  70                  75                  80

Lys Asp Leu Cys Arg Tyr Leu Lys Cys Val Met Thr Phe Tyr Arg His
                85                  90                  95

Gln His Ile Asp Phe Val Ile Val Tyr Glu Arg Gln Pro Pro Phe Glu
            100                 105                 110

Ile Asp Lys Leu Thr Tyr Met Lys Tyr His Pro Tyr Met Leu Leu Gln
        115                 120                 125

Arg Lys His Lys Ile Ile Leu Pro Ser Gln Thr Thr Asn Pro Arg Gly
    130                 135                 140

Lys Leu Lys Lys Lys Lys Thr Ile Lys Pro Pro Lys Gln Met Leu Ser
145                 150                 155                 160

Lys Trp Phe Phe Gln Gln Gln Phe Ala Lys Tyr Asp Leu Leu Leu Ile
                165                 170                 175

Ala Ala Ala Ala Cys Ser Leu Arg Tyr Pro Arg Ile Gly Cys Cys Asn
            180                 185                 190

Glu Asn Arg Met Ile Thr Leu Tyr Cys Leu Asn Thr
        195                 200
```

<210> 974
<211> 204
<212> PRT
<213> Torque teno midi virus

<400> 974

```
Lys Ile Thr Ile Lys Gln Trp Gln Pro Asp Ser Val Lys Lys Cys Lys
1               5                   10                  15
```

```
Ile Lys Gly Tyr Ser Thr Leu Val Met Gly Ala Gln Gly Lys Gln Tyr
            20                  25                  30

Asn Cys Tyr Thr Asn Gln Ala Ser Asp Tyr Val Gln Pro Lys Ala Pro
        35                  40                  45

Gln Gly Gly Gly Phe Gly Cys Glu Val Phe Asn Leu Lys Trp Leu Tyr
    50                  55                  60

Gln Glu Tyr Thr Ala His Arg Asn Ile Trp Thr Lys Thr Asn Glu Tyr
65                  70                  75                  80

Thr Asp Leu Cys Arg Tyr Thr Gly Ala Gln Ile Ile Leu Tyr Arg His
                85                  90                  95

Pro Asp Val Asp Phe Ile Val Ser Trp Asp Asn Gln Pro Pro Phe Leu
            100                 105                 110

Leu Asn Lys Tyr Thr Tyr Pro Glu Leu Gln Pro Gln Asn Leu Leu Leu
        115                 120                 125

Ala Arg Arg Lys Arg Ile Ile Leu Ser Gln Lys Ser Asn Pro Lys Gly
    130                 135                 140

Lys Leu Arg Ile Lys Leu Arg Ile Pro Pro Pro Lys Gln Met Ile Thr
145                 150                 155                 160

Lys Trp Phe Phe Gln Arg Asp Phe Cys Asp Val Asn Leu Phe Lys Leu
                165                 170                 175

Cys Ala Ser Ala Ala Ser Phe Arg Tyr Pro Gly Ile Ser His Gly Ala
            180                 185                 190

Gln Ser Thr Ile Phe Ser Ala Tyr Ala Leu Asn Thr
        195                 200
```

<210> 975
<211> 204
<212> PRT
<213> Torque teno midi virus

<400> 975
Phe Leu Lys Leu Val Gln Trp Gln Pro Asp Arg Ile Arg Lys Cys Thr
1   5   10   15

Ile Lys Gly Arg Gly Thr Leu Val Met Gly Ala His Gly Arg Gln Met
            20   25   30

Ile Cys Phe Thr Asn Val Lys Asp Arg Leu Thr Pro Ser Arg Ala Pro
        35   40   45

Ala Gly Gly Gly Phe Gly Val Glu Arg Ile Thr Leu Ala His Leu Tyr
    50   55   60

Glu Glu Trp Lys Ala Arg Asn Asn Val Trp Thr Ala Thr Asn Glu Asn
65   70   75   80

Thr Asp Leu Gly Arg Tyr Thr Gly Cys Lys Ile Ser Phe Tyr Arg His
                85   90   95

Leu Asp Thr Asp Phe Ile Ile Lys Tyr Ser Arg Asn Pro Pro Phe Asn
            100   105   110

Leu Glu Lys Tyr Ser Tyr Met Tyr Met Gln Pro Gln Glu Leu Leu Leu
        115   120   125

Gly Lys His Lys Arg Ile Leu Leu Ser Lys Lys Thr Asn Pro Lys Gly
    130   135   140

Lys Leu Lys Thr Thr Leu Lys Ile Gly Pro Pro Lys Leu Leu Thr Asn
145   150   155   160

Lys Trp Leu Leu Gln Ser Glu Leu Ala Lys Gln Asp Leu Val Glu Ile
                165   170   175

Ser Ala Ala Ala Ala Ser Phe Thr Tyr Pro Thr Ile Gly Cys Cys Asn
            180   185   190

Glu Asn Arg Ile Leu Asn Leu Tyr Ser Ile Asn Thr
        195   200

<210> 976
<211> 24
<212> PRT
<213> Torque teno virus

<400> 976
Pro Tyr Thr Asp Ile Thr Tyr Asn Pro Phe Ser Asp Arg Gly Glu Gly
1   5   10   15

Asn Met Leu Trp Ile Asp Trp Leu
            20

<210> 977
<211> 24
<212> PRT
<213> Torque teno virus

<400> 977
Ala Tyr Thr Asp Ile Lys Tyr Asn Pro Phe Thr Asp Arg Gly Glu Gly
1   5   10   15

Asn Met Leu Trp Ile Asp Trp Leu
            20

<210> 978
<211> 24
<212> PRT
<213> Torque teno virus

<400> 978
Ala Tyr Ser Asp Ile Ile Tyr Asn Pro Leu Thr Asp Lys Gly Thr Gly
1   5   10   15

Asn Ile Ile Trp Ile Asp Tyr Leu
            20

<210> 979
<211> 24
<212> PRT
<213> Torque teno virus

<400> 979
Ala Tyr Thr Asp Val Ile Tyr Asn Pro Phe Thr Asp Lys Gly Thr Gly
1   5   10   15

Asn Met Val Trp Ile Asp Tyr Leu
            20

<210> 980
<211> 24
<212> PRT
<213> Torque teno virus

<400> 980
Ala Tyr Gln Asp Ile Ile Tyr Asn Pro Phe Thr Asp Lys Gly Ile Gly
1   5   10   15

Asn Ile Val Trp Ile Asp Trp Leu
            20

<210> 981
<211> 24
<212> PRT
<213> Torque teno virus

<400> 981
Pro Tyr Thr His Val Arg Tyr Asn Pro Leu Val Asp Lys Gly Phe Gly
1   5   10   15

Asn Arg Ile Tyr Ile Gln Trp Cys
            20

<210> 982
<211> 24
<212> PRT
<213> Torque teno virus

<400> 982
Pro Tyr Thr His Val Arg Tyr Asn Pro Leu Val Asp Lys Gly Phe Gly
1   5   10   15

Asn Arg Ile Tyr Ile Gln Trp Cys
            20

<210> 983
<211> 24
<212> PRT
<213> Torque teno virus

<400> 983
Ala Tyr Thr Glu Val Ser Tyr Asn Pro Trp Thr Asp Glu Gly Thr Gly
1   5   10   15

Asn Val Val Cys Leu Gln Tyr Leu
            20

<210> 984
<211> 24
<212> PRT
<213> Torque teno virus

<400> 984
Leu Tyr Thr Glu Ile Ile Tyr Asn Pro Tyr Thr Asp Lys Gly Thr Gly
1   5   10   15

Asn Lys Val Trp Met Asp Pro Leu
            20

<210> 985
<211> 24
<212> PRT
<213> Torque teno virus

<400> 985
Ala Tyr Gln Asp Val Thr Tyr Asn Pro Asn Cys Asp Arg Ala Val Gly
1   5   10   15

Asn Val Val Trp Phe Gln Tyr Ser
            20

<210> 986
<211> 24
<212> PRT
<213> Torque teno virus

<400> 986
Ala Tyr Gln Asp Val Thr Tyr Asn Pro Asn Cys Asp Arg Gly Val Asn
1   5   10   15

Asn Arg Val Trp Val Gln Pro Leu
            20

<210> 987
<211> 24
<212> PRT
<213> Torque teno virus

<400> 987
Ala Tyr Gln Asp Val Thr Tyr Asn Pro Leu Met Asp Lys Gly Val Gly
1   5   10   15

Asn His Ile Trp Phe Gln Tyr Asn
            20

<210> 988
<211> 24
<212> PRT
<213> Torque teno virus

<400> 988
Ala Tyr Val Asp Thr Thr Tyr Asn Pro Leu Met Asp Lys Gly Lys Gly
1   5   10   15

Asn Lys Ile Trp Phe Gln Tyr Leu
            20

<210> 989
<211> 24
<212> PRT
<213> Torque teno virus

<400> 989
Ala Tyr Arg Asp Ile Val Tyr Asn Pro Phe Leu Asp Lys Ala Gln Gly
1   5   10   15

Asn Met Val Trp Phe Gln Tyr His
            20

<210> 990
<211> 21
<212> PRT
<213> Torque teno mini virus

<400> 990
Glu Cys Arg Tyr Asn Pro Phe Asn Asp Lys Ser Thr Gly Asn Asn Thr
1   5   10   15

Phe Phe Val Ser Asn
            20

<210> 991
<211> 21
<212> PRT
<213> Anelloviridae sp.

<400> 991

Glu Cys Arg Tyr Asn Pro Leu Ala Asp Lys Gly Lys Gly Asn Lys Ile
1   5   10   15

Tyr Leu Ile Ser Ile
            20

<210> 992
<211> 21
<212> PRT
<213> Torque teno mini virus

<400> 992
Thr Cys Arg Tyr Asn Pro Leu Lys Asp Thr Gly Glu Gly Asn Gln Val
1   5   10   15

Phe Phe Lys Ser Thr
            20

<210> 993
<211> 21
<212> PRT
<213> Torque teno mini virus

<400> 993
Thr Ile Arg Tyr Ser Pro Asn Arg Asp Thr Gly Glu Asn Asn Ile Cys
1   5   10   15

Tyr Phe Lys Pro Ile
            20

<210> 994
<211> 21
<212> PRT
<213> Torque teno mini virus

<400> 994
Thr Thr Arg Tyr Asn Pro Asn Arg Asp Glu Gly Ile Gln Asn Glu Met
1   5   10   15

Tyr Leu Val Asn Val
            20

<210> 995
<211> 21
<212> PRT
<213> TTV-like mini virus

<400> 995
Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp Thr Gln Leu
1   5   10   15

Tyr Leu Leu Ser Asn
            20

<210> 996
<211> 21
<212> PRT
<213> Torque teno mini virus

<400> 996
Asp Ile Arg Tyr Ala Pro Asp Arg Asp Thr Gly Val Gly Asn Lys Val
1   5   10   15

Tyr Leu Lys Ser Val
            20

<210> 997
<211> 20
<212> PRT
<213> Torque teno midi virus

<400> 997
Cys Arg Tyr Asn Pro Ala Ile Asp Thr Gly Lys Gly Asn Thr Ile Trp
1   5   10   15

Leu His Ser Asn
            20

<210> 998
<211> 20
<212> PRT
<213> Torque teno midi virus

<400> 998
Ala Ile Tyr Asn Pro Val Lys Asp Asn Gly Asp Gly Asn Met Ile Tyr
1   5   10   15

Leu Val Ser Thr
            20

<210> 999
<211> 20

<212> PRT
<213> Torque teno midi virus

<400> 999
Gly Arg Tyr Asn Pro Ala Ile Asp Asp Gly Lys Gly Asn Lys Ile Tyr
1   5   10   15

Leu Gln Thr Ile
            20

<210> 1000
<211> 20
<212> PRT
<213> Torque teno midi virus

<400> 1000
Leu Arg Tyr Asn Pro His Glu Asp Thr Gly His Gly Asn Glu Ile Tyr
1   5   10   15

Leu Thr Ser Thr
            20