

METHODS AND COMPOSITIONS FOR BAT IPSC PREPARATION AND USE**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims the benefit of and priority to Great Britain Patent Application No. GB 2115676.5, filed on November 1, 2021; U.S. Provisional Patent Application No. 63/360,472, filed on October 4, 2020; U.S. Provisional Patent Application No. 63/248,835, filed on September 27, 2021, the disclosure of each of which is hereby incorporated by reference in its entirety for all purposes.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0001] This invention was made with U.S. government support, Grant No. HR0011-19-2-0020, awarded by DARPA and Grant No. W81XWH-20-1-0270, awarded by Department of Defense (DoD), NIAID grant U19AI135972, and CRIPT (Center for Research on Influenza Pathogenesis and Response), a NIAID supported Center of Excellence for Influenza Research and Response grant CEIRR, contract #75N93019R00028. The U.S. government has certain rights to the invention.

BACKGROUND

[0002] Bats have evolved features unique amongst mammals, including flight, laryngeal echolocation, and an immune system that shows unusual tolerance for viruses that cause life-threatening diseases in humans (*e.g.*, SARS-CoVs, MERS-CoV, Ebola). Recent comparative genomic studies uncovered bat-specific changes to key immunity genes and exposed numerous integrated viral sequences, suggesting a particularly intimate and deep-rooted accord between bats and viruses. Still, what makes bats most distinctive is that they are home to the richest virosphere among mammals with some of the bat-related viruses causing significant outbreaks, including SARS, Ebola, and COVID-19. Remarkably, bats can be infected with viruses that are lethal to other mammals without causing any symptoms. Even more, the bat genome seems to act as a sponge for viral sequences. While endowed with a small genome, bats house a spacious number of ancient and contemporary viral insertions of retroviral and non-retroviral origin. Because some of the viral sequences are full length and even of non-bat origin, bats might supply an essential template for zoonotic viruses and act as super-spreaders. Nonetheless, how bats deal with viruses so well is poorly understood. It is clear that, although bats are a critically needed new model organism, limited access to animal

and cell models has hindered their study. Bat breeding colonies are notoriously challenging to establish, and bat primary cell lines typically have a limited lifespan in vitro. Therefore, induced pluripotent stem cells would offer a research tool for bat research.

SUMMARY

[0003] In one aspect, the disclosure provides a composition for an induced pluripotent bat stem cell (bat iPSC), wherein the cell is in a pluripotent state. In some embodiments the bat IPS cell is in a pluripotent state characterized by the expression of one or more factors for example of Klf4, Klf17, Essrb, Tfcp2l1, Tfe3, Dppa, Oct4, Sox2, Nanog, and Dusp6. In some embodiments, the iPSC cell is in a naïve pluripotent state. In some embodiments, the cell is characterized by the expression of one or more factors for example Otx2 or Zic2. In some embodiments the cell is a bat fibroblast or a bat embryonic fibroblast. In some embodiments the bat is a *Rhinolophus* bat or a *Rhinolophus ferrumequinum* bat, alternatively the bat is a *Myotis bat* or a *Myotis myotis bat*. In some embodiments, the IPS cell is capable of differentiating into embryonic bodies. In some embodiments, the embryonic bodies are capable of differentiating into three-dimensional structures comprising three germ layer markers.

[0004] In another aspect, the disclosure provides a method of producing induced pluripotent bat stem cells (bat iPSCs), the method comprising: (i) reprogramming isolated bat cells with Oct4, Sox2, cMyc, and Klf4 factors, (ii) culturing the reprogrammed cells on feeder cells in a medium comprising FGF, Leukemia inhibitory factor (Lif), SCF, and Forskolin until colonies appear; and (iii) splitting cells using a low concentration EDTA buffer; thereby producing iPSCs from bats. In some embodiments, the isolated bat cell is a fibroblast or an embryonic fibroblast. In some embodiments the cell is derived from a bat is a *Rhinolophus* bat or a *Rhinolophus ferrumequinum* bat, alternatively the bat is a *Myotis bat* or a *Myotis myotis bat*. In some embodiments, the Lif is at a concentration of 10^4 U/ml. In some embodiments, the FGF is at a concentration of 100 ng/ml. In some embodiments, the SCF is at a concentration of 100 ng/ml. In some embodiments, the Forskolin is at a concentration of 20 nM. In some embodiments, the feeder cell is a mouse CF1 mouse embryonic fibroblasts (MEF). In some embodiments, the method further comprises passaging the bat iPSCs every 5 days onto feeder cells. In some embodiments, the bat iPSC is further differentiated into embryonic bodies. In some embodiments, the embryonic bodies are further differentiated into three-dimensional structures comprising three germ layer markers.

[0005] In another aspect the disclosure provides a method of producing induced pluripotent bat stem cells (bat iPSCs), the method comprising: (i) reprogramming isolated bat cells with Oct4, Sox2, cMyc, and Klf4 factors; (ii) culturing the reprogrammed cells in feeder free medium comprising FGF, Leukemia inhibitory factor (Lif), SCF, and Forskolin until colonies appear; and (iii) splitting cells using a low concentration EDTA buffer thereby producing iPSCs from bats.

[0006] In another aspect the disclosure provides a composition for reprogramming a bat cell to produce pluripotent stem cells comprising a medium comprising FGF, Leukemia inhibitory factor (Lif), SCF, and Forskolin. In some embodiments, the Lif is at a concentration of 10^4 U/ml. In some embodiments, the FGF is at a concentration of 100 ng/ml. In some embodiments the SCF is at a concentration of 100 ng/ml. In some embodiments, the Forskolin is at a concentration of 20 nM.

[0007] In another aspect the disclosure provides a method of obtaining viral sequences from bat iPSCs, the method comprising obtaining bat iPSCs; identifying viral sequences residing in the bat iPSC genome or intracellular virus genome; and assembling the viral sequences; thereby obtaining viral sequences from the bat iPSCs. In some embodiments, the identifying comprises sequencing the bat genome or the genome of viral particles residing in the bat iPSCs, or of viral particles shed by the bat iPSCs. In some embodiments, identifying comprises sequencing the RNA of the bat genome or the genome of viral particles residing in the bat iPSCs, or of viral particles shed by the bat iPSCs. In some embodiments, the identifying the proteins and peptides produced by the viral genome by proteomics e.g., LC-MS. In some embodiments, the method comprises translating the sequence into a protein sequence and determining whether the translated sequence has a significant homology to a known protein sequence in a viral protein database. In some embodiments, the sequence is selected from SEQ ID NO: 1-349. In some embodiments, the virus is selected from the group of a SARS-CoV-2 virus, endogenous retrovirus (RfRV), and sindbis virus. In some embodiments, the virus is a coronavirus. In some embodiments, the sequence encodes a gag protein, a pol protein, or an env protein.

[0008] In another aspect the disclosure provides a method of obtaining viral sequences from virus particles shed by bat iPSCs or cells derived from bat iPSCs, the method comprising obtaining bat iPSCs or cells derived from bat iPSCs; culturing the bat iPSCs or cells derived from bat iPSCs under conditions that allows shedding of virus particles into the culture media; collecting the culture media; identifying viral sequences residing in the culture

media; and assembling the viral sequences, thereby obtaining viral sequences from virus particles shed by bat iPSCs or cells derived from bat iPSCs.

[0009] In another aspect the disclosure provides for the use of any one of the viral sequences described above for the development of a vaccine.

[0010] In another aspect the disclosure provides for a recombinant nucleic acid molecule, comprising a promoter, and a nucleic acid selected from SEQ ID NO: 1-349 encoding for a viral protein or fragment thereof. In some embodiments, a recombinant, replication deficient adenovirus, comprising nucleic acid described above is provided. In some embodiments, mRNA comprising the nucleic acid described above is provided.

[0011] In another aspect the disclosure provides for an expression vector comprising a promoter and a nucleic acid set forth in SEQ ID NO: 1-349 encoding for a viral protein or fragment thereof.

[0012] In another aspect the disclosure provides for an isolated protein or peptide comprising an amino acid sequence encoded in a nucleic acid set forth in SEQ ID NO: 1-349, wherein the peptide is no more than 100 amino acids in length, and an optional pharmaceutically acceptable carrier. In some embodiments, the protein or peptide is no more than 30 amino acids in length or 20 amino acids in length. In some embodiments, the protein or peptide is synthetic.

[0013] In another aspect the disclosure provides for a pharmaceutical composition comprising the adenovirus of described above, the mRNA described above, or the protein or peptide of any described above and a pharmaceutically acceptable carrier or excipient. In some embodiments, the pharmaceutical composition comprises a plurality of (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) proteins or peptides described above and a pharmaceutically acceptable carrier or excipient. In some embodiments, the pharmaceutical composition comprises a nucleic acid encoding the mRNA described above or the protein or peptide described above and a pharmaceutically acceptable carrier or excipient. In some embodiments, the pharmaceutical composition comprises one or more nucleic acids encoding a plurality of (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) mRNAs of described above or proteins or peptides of described above, and a pharmaceutically acceptable carrier or excipient. In some embodiments, the pharmaceutical composition further comprises a liposome, wherein the protein or peptide or the nucleic acid encoding the protein or peptide is disposed within the liposome. In some embodiments, the pharmaceutical composition further comprises a lipid nanoparticle, wherein the protein or peptide or the nucleic acid encoding the protein or

peptide is disposed within the lipid nanoparticle. In some embodiments, the pharmaceutical composition comprises an immunogenicity enhancing adjuvant.

[0014] In another aspect the disclosure provides for a vaccine that stimulates a T cell mediated immune response when administered to a subject, the vaccine comprising the pharmaceutical composition described above. In some embodiments, the vaccine is a priming vaccine and/or a booster vaccine.

[0015] In another aspect the disclosure provides for a recombinant cell comprising a nucleic acid or a portion of a nucleic acid set forth in SEQ ID NO: 1-349. In some embodiments, the recombinant cell comprises a protein or a portion of a protein encoded by a nucleic acid set forth in SEQ ID NO: 1-349.

[0016] In another aspect the disclosure provides for a composition comprising an inhibitor of a protein encoded by a nucleic acid selected from SEQ ID NO: 1-349.

[0017] For a fuller understanding of the nature and advantages of the present disclosure, reference should be had to the ensuing detailed description taken in conjunction with the accompanying figures. The present disclosure is capable of modification in various respects without departing from the present disclosure. Accordingly, the figures and description of these embodiments are not restrictive.

DESCRIPTION OF THE DRAWINGS

[0018] These and other features, aspects, and advantages of the present disclosure will become better understood with regard to the following description, and accompanying drawings, where:

[0019] **FIG. 1A-FIG. 1I** illustrate the derivation of pluripotent bat stem cells. **FIG. 1A**, illustrates the bat pluripotent stem cell derivation strategy. BEF, embryonic fibroblasts; OSMK, Oct4, Sox2, cMyc, Klf4; FB, fibroblast medium; PSC, pluripotent stem cell medium; PSC+, PSC with additives, **FIG. 1B**, shows exemplary morphologies of established BiPS cell colonies grown on mouse embryonic fibroblasts. **FIG. 1C**, Immunofluorescent detection of Oct4 in BiPS cells. **FIG. 1D**, MA plot of RNA-seq data illustrating the transcriptional differences between bat embryonic fibroblast (BEF) and pluripotent stem cells (BiPS). Selected genes with known functions in the establishment or maintenance of pluripotency are highlighted in dark filled circles. **FIG. 1E**, shows a Kmean cluster analysis of ATAC-seq signals obtained from BEF or BiPS cells. C, cluster. **FIG. 1F**, shows a density plot of RRBS results obtained from BEF and BiPS cells. PCC, Pearson correlation coefficient. **FIG. 1G**, shows scatter plots of histone 3 methylation status at K4 (activating chromatin modification)

or K27 (repressing chromatin modification) after ChIP-seq from BEF or BiPS cells as indicated. **FIG. 1H**, shows a scatter plot of H3K4me3 and H3K27me3 in BiPS cells illustrating the occurrence of bivalent chromatin sites in BiPS cells. **FIG. 1I**, shows RNA-seq, ATAC-seq and H3K4me3 or H3K27me3 ChIP-seq signals of selected genes with known roles in reprogramming that are activated (Nanog, Kit) or repressed (Thy1) in BiPS when compared to BEF cells.

[0020] FIG. 2A-FIG. 2M. illustrate the characterization of pluripotent stem cells generated from *Rhinolophus ferrumequinum* and *Myotis myotis* fibroblasts. **FIG. 2A**, shows exemplary microscopic images of human embryonic stem cells (H9)(lower panels) and bat pluripotent stem cells (upper panel) at indicated magnifications showing cytoplasmic vesicles. **FIG. 2B**, shows a karyotype analysis of BiPS cells at passage 17. Shown is a representative image after Giemsa staining of a metaphase spread with 56 chromosomes. **FIG. 2C**, shows PCR verification of reprogramming-associated virus clearing. Bat iPS cells (BiPS) at passage 92 were tested for Sendai virus clearance in comparison to the embryonic fibroblasts used as starting material (BEF), adult fibroblasts as negative control (NC), and freshly-transduced cells at passage 3 as a positive control (PC). bp, base pairs; SeV, Sendai virus; KOS, KLF4-OCT4- SOX2. **FIG. 2D**, shows a correlation scatter plot of methylation level at common CpG sites in duplicate samples of BEF or BiPS cells. BEF, bat embryonic fibroblast cells; BiPS, bat pluripotent stem cells; PCC, Pearson correlation coefficient. **FIG. 2E** Venn diagram illustrating the overlap of bivalent genes in bat iPSCs and human ES cells. **FIG. 2F**, Correlation plot of shrunken log2-fold changes in ATAC-seq signal with log2-fold expression changes. Shown are all values with $p < 0.05$. **FIG. 2G**, Correlation of log2-fold changes in H3K4 trimethylation (H3K4me3, left) or H3K27 trimethylation (H3K27me3, right) with log2-fold changes in gene expression. **FIG. 2H**, Correlation of log2-fold gene expression changes with the difference in the methylated fraction of promoters (left) or gene bodies (right) fractions. **FIG. 2I**, Characterization of *Myotis myotis* induced pluripotent stem cells. Microscopic images of *Myotis myotis* iPS cells after immunostaining to detect pluripotency marker Oct4. **FIG. 2J**, Microscopic images of *Myotis myotis* iPS cells that underwent differentiation and immunostaining to detect Pax6, Brachyury (T) and Afp as markers of ectoderm, mesoderm and endoderm, respectively. **FIG. 2K-FIG. 2M** illustrate the characterization of pluripotency markers in pluripotent stem cells generated from *Rhinolophus ferrumequinum* fibroblasts **FIG. 2K**, Sequencing tracks showing expression, ATAC-seq signal, Histone H3K27 trimethylation (H3K27me3) and Histone H3K4 trimethylation (H3K4me3) status of pluripotency markers Oct4, and Sox2 in bat embryonic

fibroblasts (BEF) or induced pluripotent stem cells (BiPS). **FIG. 2L**, Fraction of methylated sites in promoters of pluripotency genes that did show promoter methylation. **FIG. 2M**, Immunofluorescence images of bat pluripotent stem cells after staining of markers of naive (Tfe3 and Tfcp2l1) or primed pluripotency (Zic2 and Otx2).

[0021] **FIG. 3A-FIG. 3G** illustrate the differentiation potential of bat pluripotent stem cells. **FIG. 3A**, illustrates exemplary immunofluorescence microscopy images after staining with antibodies detecting the expression of lineage-specific markers Pax6, Afp or Brachyury (T) following specific directed differentiation into ectoderm, endoderm or mesoderm, respectively. **FIG. 3B** illustrates exemplary immunofluorescence images of embryonic bodies (EB) that formed after 3D-differentiation of BiPS cells and were stained with antibodies to detect markers specific to all three germ layers as in **FIG. 3A**. **FIG. 3C** shows RNA-seq signal of selected lineage-specific marker genes in BiPS cells that underwent monolayer differentiation as in (**FIG. 3A**) or embryonic body differentiation as in (**FIG. 3B**). EB, embryonic body differentiation; EC, human ectoderm differentiation protocol; EN, human endoderm differentiation protocol; M, human mesoderm differentiation protocol. **FIG. 3D**, illustrates exemplary microscopic images of Hematoxylin-Eosin-stained sections of tumor tissue after injection of BiPS cells into immunocompromised mice exhibiting ectodermal (left), mesodermal (middle) and endodermal (right) features. **FIG. 3E** shows exemplary images of floating blastoids that were obtained from BiPS cells after exposure to Bmp4 to capture their morphology by phase-contrast microscopy (left) and to detect Oct4 expression in inner-cell mass-like cell clusters by after immunofluorescence staining (middle, right). **FIG. 3F** illustrates Phase-contrast microscopy image of a typical blastocyst outgrowth-like cell cluster that formed after attachment of blastoids to the cell culture vessel surface during Bmp4-induced differentiation as in **FIG. 3E**. ICL, Inner cell mass-like; TLO, trophoblast-like outgrowth. **FIG. 3G** shows an expression profile of genes associated with tumor suppression. The data sets were from this study (bat), GSE53212 (mouse, GEO), PRJNA400257 (Naked mole-rat, BioProject), and GEOGSE175070 (human, GEO). ARF, ADP ribosylation factor; BEF, bat embryonic fibroblasts; BiPS, bat induced pluripotent stem cells; ERAS, ES cell-expressed Ras; FOXO6, Forkhead Box O6; H9, human ES cells; HAS, Hyaluron-synthase; MEFs, mouse embryonic fibroblasts; NMR, naked mole-rat.

[0022] **FIG. 4A-FIG. 4D**, illustrate the differentiation potential of bat pluripotent stem cells. **FIG. 4A**, Schematic of differentiation strategies. **FIG. 4B**, Representative image of embryoid bodies differentiated for 3 days. **FIG. 4C**, shows a MA plot depicting the log2

mean expression and log 2 fold expression changes of all genes in bat pluripotent stem cells (BiPS) after exposure to the noted differentiation conditions illustrated in **FIG. 4A**. EB, Embryoid body differentiation; EC, human ectoderm differentiation conditions; EN, human endoderm differentiation conditions; M, human mesoderm differentiation conditions. **FIG. 4D**, shows a heatmap depicting expression changes of genes known as markers for human ectoderm, mesoderm, or endoderm during the differentiation of BiPS under the conditions described in **FIG. 4A**.

[0023] FIG. 5A-5D. illustrate distinct characteristics of pluripotent bat stem cells. **FIG. 5A** shows principal component analysis of induced pluripotent bat stem cells (BiPS) in comparison to those derived from other species. h, human; m, mouse. PS, pluripotent stem cells, iPS, induced pluripotent stem cells, S, embryonic stem cells, EF, embryonic fibroblasts. **FIG. 5B** shows a plot of genes that contribute to the differences of pluripotent bat and mouse stem cells as part of principal component 1 (PC1). Highlighted in light blue is the “leading edge” comprised of the top 5% of PC1-contributing genes. **FIG. 5C** shows selected GO and **FIG. 5D** shows KEGG pathways identified to be significantly enriched among the top 5 % of PC1-contributing genes/leading edge genes defined in (**FIG. 5B**) were plotted by their odds ratio, with the color of each circle indicating the enrichment p-value and the size indicating the number of genes present in the respective category. ER, endoplasmic reticulum; PT, protein targeting; Pos, positive; Reg, regulation.

[0024] FIG. 6A illustrates the interaction of genes that are part of the KEGG Corona Virus Disease pathway. Nodes are colored based on the log2 fold change between BiPS and mouse iPS cells. Red indicates genes that are expressed at a higher level in BiPS, blue indicates those that are expressed at a lower level. Bold borders indicate proteins that were present in the top 5% of genes in PC1 (leading edge). **FIG. 6B** illustrates that the selection analyses of leading edge- genes by comparative genomics analyses of the *R. ferrumequinum* lineage identified eight genes showing significant evidence of positive selection. Additional lineages and the number of genes showings selection found in them, are highlighted in brackets.

[0025] FIG. 7A-7J illustrate viral tolerance of pluripotent bat stem cells. **FIG. 7A** shows the expression of indicated ERV elements in bat embryonic fibroblasts (BEF) and iPS cells (BiPS) as determined by extracting the overlap between RNA-seq reads mapped to the *R. ferrumequinum* genome and known mapped ERV elements. Shown are the elements with the

most evident differences. **FIG. 7B**, shows an exemplary electron microscopy image of cytoplasmic vesicles of BiPS cells containing virus-like structures. Bottom: higher magnification of viroid structures: Intracellular inclusions of virus-like particles (black arrows) with granular and electron-dense content (white arrowheads), typically surrounded by double membrane structures (white arrows), and some of them coated with protrusions (black arrowheads). **FIG. 7C**, Western blotting in human 293FT (kidney tumor cell line) and embryonic stem cells (H9), mouse 3T3 (fibroblasts) and embryonic stem cells (R1), and bat pluripotent stem cells (BiPS) with a HERV K capsid (Cap) specific antibody detecting human endogenous retroviruses. **FIG. 7D**, shows exemplary immunofluorescence images of BiPS cells detecting the HERVK Gag/Cap protein. **FIG. 7E**, shows Western blotting in human 293FT, H9, mouse 3T3 and R1, and BiPS with a pan coronavirus antibody known to be specific for the nucleocapsid; its reactivity includes but might not be limited to feline infectious peritonitis virus type 1 and 2, the canine coronavirus (CCV), pig coronavirus transmissible gastroenteritis virus (TGEV), and ferret coronavirus. **FIG. 7F**, illustrates exemplary immunofluorescence images of BiPS cells after detection of pan coronavirus antigen. **FIG. 7G**, shows exemplary immunofluorescence images of BiPS cells after detection of double stranded RNA characteristic RNA viruses.

[0026] **FIG. 8A-FIG. 8C** illustrate exemplary microscopic images of bat pluripotent stem cells. **FIG. 8A**, shows a 40x magnification of a bat pluripotent stem cell colony. **FIG. 8B** and **FIG. 8C** show an overview of transmission electron microscopy of bat pluripotent stem cells. Vi, vesicles containing viral-like structures; OV, other vesicle structures filled with homogenous content; Nu, Nucleus; A, autophagosome; M, mitochondria. **FIG. 8D** shows a higher magnification of the structures.

[0027] **FIG. 9A-9H** illustrate exemplary virome mining in BiPS cells. **FIG. 9A** flow diagram of the sequence mining for viral sequences in the bat genome. **FIG. 9B** shows the taxonomic distribution of virome reads as determined by the metagenomic classifier Kraken2. The distribution of the reads that were mapped according to the virus data base are shown in a phylogenetic tree. The green color coding represents the number of taxa observed, the red nodes denote particular taxa of interest. **FIG. 9B** shows the number of viral species as classified by Kraken through RNA-seq and iso-seq sequencing. **FIG. 9C** shows the number of individual viruses species and subspecies obtained from iso-seq (top panel) and RNA-seq (bottom panel). **FIG. 9D** shows RNA and Iso-seq sequencing tracks for a newly discovered

full-length retrovirus sequence, RFe-V-MD1, aligned to the *R. ferrumequinum* genome. The Iso-seq fragment represents a 6088 bp-long transcript. **FIG. 9E** shows genomic and sequence track for short integrated viral sequences for Columbid/Falconid herpesvirus and Sindbis virus. **FIG. 9F** illustrate the short viral insertion shown in **FIG. 9E** form stem-loop structures. **FIG. 9G** illustrates another example of a short viral integration showing homology to two human herpesvirus 4 isolates (HKD40 and HKNPC60), the human respiratory syncytial virus (Kilifi isolate), and a fragment of about 500 bp that was identified at the end of a SARS-CoV2 isolate in an infected patient (OU077605.1). **FIG. 9H** shows a genome track for a *Scotophilus* bat coronavirus 512 homologous sequence of the spike protein coding region. **FIG. 9I** ImageStream analysis after immunofluorescence staining of BiPS cells. A brightfield image, Crystal Violet nuclear staining (Nucleus), dsRNA staining (dsRNA) and an overlay is shown for each representative cell.

[0028] **FIG. 10A** shows exemplary results of long-read RNA sequencing (iso-seq). the sequencing reads were mapped against a virus database, using a metagenomic classification tool (Kraken) including viruses from several significant viral families, including Paramyxoviridae, Rhabdoviridae, Filoviridae, Bornaviridae, Flaviviridae, Coronaviridae, Picornaviridae, and Retroviridae. **FIG. 10B** shows the number of viral species as classified in BEFs and BiPS. **FIG. 10C** illustrates an exemplary assembly of full-length viruses, shorter viral insertions, and novel, more distant viruses based on the sequencing data from BiPS cells such as the shown full-length bat retrovirus (RFeRV). The top shows short nucleotide reads aligned to a full length sequence. The middle and lower part of the figure shows the position of a Gag, Pol, and Env protein in the genome.

[0029] **FIG. 11A-11D** illustrate exemplary protein and nucleotide sequences identified in the BiPS cells that are associated with viruses. **FIG. 11A** shows a protein sequence with homology to a hypothetical protein CoVHLJ_8- from Columbid alphaherpesvirus 1 and a nucleotide sequence that is similar to a Sindbis virus defective interfering particle di-2. **FIG. 11B** shows a protein or a protein fragment with homologies to an RNA-dependent DNA polymerase of the lymphocystis disease virus and of the erythrocytic necrosis virus. **FIG. 11C** illustrates the results of mapping of a region residing in the first intron of the XPA gene (a DNA damage and repair factor) on chromosome 12. A BLAST search with the fragment showed homology to two human herpesvirus 4 isolates (HKD40 and HKNPC60), the human respiratory syncytial virus (Kilifi isolate), and a fragment of about 500 bp that was identified

at the end of a SARS-CoV2 isolate in an infected patient. **FIG. 11D** shows a phylogenetic analysis of the genomic sequences mostly resembled the spike protein-encoding genomic portion of human coronavirus 229E and the human coronavirus OC43.

DETAILED DESCRIPTION

[0030] Various features and aspects of the disclosure are discussed in more detail below.

[0031] The disclosure is based, in part, upon the discovery that induced pluripotent bat stem cells can be produced and are stable in culture, readily differentiate into all three germ layers, and form complex embryoid bodies, including organoids. Bat iPSCs (BiPS) and their differentiated progeny can be used for example as an accessible and versatile tool required to advance bats as a new model system. Further, BiPS can provide the platform to further understand the role bats play as virus reservoirs and enable new insights into emerging viruses, such as SARS-CoV-2, and better prepare for future pandemics. BiPS can enable studies that directly impact every aspect of bats' particular biology, including this mammal's unique adaptations of flight, echolocation, extreme longevity, and unique immunity. Further, BiPS are also useful for example in understanding of bats' asymptomatic response to viral pathogens.

[0032] Accordingly, the disclosure provides BiPS, methods of producing and using BiPS, and compositions for reprogramming bat cells.

[0033] In another aspect, the disclosure is based in part on the discovery of viruses and viral nucleic acids and proteins in BiPS. The viruses, viral nucleic acids, viral proteins, viral nucleic acid sequences, and protein sequences are useful in the development of therapeutics and prophylactics for viral diseases, such as vaccines, antibodies, and small molecule antivirals.

[0034] Accordingly, the disclosure provides viral nucleic acid and protein sequences, expression constructs, vectors comprising the expression constructs, methods of making and using therapeutics and prophylactics against viral diseases such as vaccines, antibodies, and small molecule antivirals.

[0035] Unless otherwise defined herein, scientific and technical terms used in this application shall have the meanings that are commonly understood by those of ordinary skill in the art.

[0036] Generally, nomenclature used in connection with, and techniques of, pharmacology, cell and tissue culture, molecular biology, cell and cancer biology, neurobiology, neurochemistry, virology, immunology, microbiology, genetics and protein

and nucleic acid chemistry, described herein, are those well-known and commonly used in the art. In case of conflict, the present specification, including definitions, will control.

[0037] The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are within the skill of the art. Such techniques are explained fully in the literature, such as, Molecular Cloning: A Laboratory Manual, second edition (Sambrook *et al.*, 1989) Cold Spring Harbor Press; Oligonucleotide Synthesis (M.J. Gait, ed., 1984); Methods in Molecular Biology, Humana Press; Cell Biology: A Laboratory Notebook (J.E. Cellis, ed., 1998) Academic Press; Animal Cell Culture (R.I. Freshney, ed., 1987); Introduction to Cell and Tissue Culture (J. P. Mather and P.E. Roberts, 1998) Plenum Press; Cell and Tissue Culture: Laboratory Procedures (A. Doyle, J.B. Griffiths, and D.G. Newell, eds., 1993-1998) J. Wiley and Sons; Methods in Enzymology (Academic Press, Inc.); Gene Transfer Vectors for Mammalian Cells (J.M. Miller and M.P. Calos, eds., 1987); Current Protocols in Molecular Biology (F.M. Ausubel *et al.*, eds., 1987); PCR: The Polymerase Chain Reaction, (Mullis *et al.*, eds., 1994); Sambrook and Russell, Molecular Cloning: A Laboratory Manual, 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley & Sons, NY (2002); Harlow and Lane Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1998); Coligan *et al.*, Short Protocols in Protein Science, John Wiley & Sons, NY (2003); Short Protocols in Molecular Biology (Wiley and Sons, 1999).

[0038] In general, terms used in the claims and the specification are intended to be construed as having the plain meaning understood by a person of ordinary skill in the art. Certain terms are defined below to provide additional clarity. In case of conflict between the plain meaning and the provided definitions, the provided definitions are to be used.

[0039] Throughout this specification and embodiments, the word “comprise,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

[0040] It is understood that wherever embodiments are described herein with the language “comprising,” otherwise analogous embodiments described in terms of “consisting of” and/or “consisting essentially of” are also provided.

[0041] The term “including” is used to mean “including but not limited to.” “Including” and “including but not limited to” are used interchangeably.

[0042] Any example(s) following the term “*e.g.*” or “for example” is not meant to be exhaustive or limiting.

[0043] Unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

[0044] The articles “a” and “an” are used herein to refer to one or to more than one (*i.e.*, to at least one) of the grammatical object of the article. By way of example, “an element” means one element or more than one element. Reference to “about” a value or parameter herein includes (and describes) embodiments that are directed to that value or parameter *per se*. For example, description referring to “about X” includes description of “X.” Numeric ranges are inclusive of the numbers defining the range.

[0045] Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the disclosure are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements. Moreover, all ranges disclosed herein are to be understood to encompass any and all subranges subsumed therein. For example, a stated range of “1 to 10” should be considered to include any and all subranges between (and inclusive of) the minimum value of 1 and the maximum value of 10; that is, all subranges beginning with a minimum value of 1 or more, *e.g.*, 1 to 6.1, and ending with a maximum value of 10 or less, *e.g.*, 5.5 to 10.

[0046] Where aspects or embodiments of the disclosure are described in terms of a Markush group or other grouping of alternatives, the present disclosure encompasses not only the entire group listed as a whole, but each member of the group individually and all possible subgroups of the main group, but also the main group absent one or more of the group members. The present disclosure also envisages the explicit exclusion of one or more of any of the group members in an embodiment of the disclosure.

[0047] Exemplary methods and materials are described herein, although methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present disclosure. The materials, methods, and examples are illustrative only and not intended to be limiting.

I. Definitions

[0048] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the claimed

subject matter belongs. It is to be understood that the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of any subject matter claimed. The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[0049] The practice of some methods disclosed herein employ, unless otherwise indicated, techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA. See for example Sambrook and Green, *Molecular Cloning: A Laboratory Manual*, 4th Edition (2012); the series *Current Protocols in Molecular Biology* (F. M. Ausubel, et al. eds.); the series *Methods In Enzymology* (Academic Press, Inc.), *PCR 2: A Practical Approach* (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) *Antibodies, A Laboratory Manual*, and *Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications*, 6th Edition (R.I. Freshney, ed. (2010)) (which is entirely incorporated by reference herein).

[0050] As used herein, the singular forms “a”, “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. Furthermore, to the extent that the terms “including”, “includes”, “having”, “has”, “with”, or variants thereof are used in either the detailed description and/or the claims, such terms are intended to be inclusive in a manner similar to the term “comprising”.

[0051] The term “about” or “approximately” means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, i.e., the limitations of the measurement system. For example, “about” can mean within one or more than one standard deviation, per the practice in the art. Alternatively, “about” can mean a range of up to 20%, up to 15%, up to 10%, up to 5%, or up to 1% of a given value.

[0052] As used herein, “residue” refers to a position in a protein and its associated amino acid identity.

[0053] As used herein the term “antigen” is a substance that induces an immune response. An antigen can be a neoantigen.

[0054] As used herein the term “antigen-based vaccine” is a vaccine composition based on one or more antigens, e.g., a plurality of antigens. The vaccines can be nucleotide-based

(*e.g.*, virally based, RNA based, or DNA based), protein-based (*e.g.*, peptide based), or a combination thereof.

[0055] As used herein the term “coding region” is the portion(s) of a gene that encode protein.

[0056] As used herein the term “coding mutation” is a mutation occurring in a coding region.

[0057] As used herein the term “ORF” means open reading frame.

[0058] As used herein the term “epitope” is the specific portion of an antigen typically bound by an antibody or T cell receptor.

[0059] As used herein the term “immunogenic” is the ability to elicit an immune response, *e.g.*, via T cells, B cells, or both.

[0060] As used herein the term “HLA binding affinity” “MHC binding affinity” means affinity of binding between a specific antigen and a specific MHC allele.

[0061] As used herein the term “ELISPOT” means Enzyme-linked immunosorbent spot assay – which is a common method for monitoring immune responses in humans and animals.

[0062] The term “lipid” includes hydrophobic and/or amphiphilic molecules. Lipids can be cationic, anionic, or neutral. Lipids can be synthetic or naturally derived, and in some instances biodegradable. Lipids can include cholesterol, phospholipids, lipid conjugates including, but not limited to, polyethyleneglycol (PEG) conjugates (PEGylated lipids), waxes, oils, glycerides, fats, and fat-soluble vitamins. Lipids can also include dilinoleylmethyl- 4-dimethylaminobutyrate (MC3) and MC3-like molecules.

[0063] The term “lipid nanoparticle” or “LNP” includes vesicle like structures formed using a lipid containing membrane surrounding an aqueous interior, also referred to as liposomes. Lipid nanoparticles includes lipid-based compositions with a solid lipid core stabilized by a surfactant. The core lipids can be fatty acids, acylglycerols, waxes, and mixtures of these surfactants. Biological membrane lipids such as phospholipids, sphingomyelins, bile salts (sodium taurocholate), and sterols (cholesterol) can be utilized as

stabilizers. Lipid nanoparticles can be formed using defined ratios of different lipid molecules, including, but not limited to, defined ratios of one or more cationic, anionic, or neutral lipids. Lipid nanoparticles can encapsulate molecules within an outer-membrane shell and subsequently can be contacted with target cells to deliver the encapsulated molecules to the host cell cytosol. Lipid nanoparticles can be modified or functionalized with non-lipid molecules, including on their surface. Lipid nanoparticles can be single-layered (unilamellar) or multi-layered (multilamellar). Lipid nanoparticles can be complexed with nucleic acid. Unilamellar lipid nanoparticles can be complexed with nucleic acid, wherein the nucleic acid is in the aqueous interior. Multilamellar lipid nanoparticles can be complexed with nucleic acid, wherein the nucleic acid is in the aqueous interior or and/or can be sandwiched between the layers.

[0064] Unless specifically stated or otherwise apparent from context, as used herein the term “about” is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from context, all numerical values provided herein are modified by the term about.

[0065] As known in the art, “polynucleotide,” or “nucleic acid,” as used interchangeably herein, refer to chains of nucleotides of any length, and include DNA and RNA. The nucleotides can be deoxyribonucleotides, ribonucleotides, modified nucleotides or bases, and/or their analogs, or any substrate that can be incorporated into a chain by DNA or RNA polymerase. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and their analogs. If present, modification to the nucleotide structure may be imparted before or after assembly of the chain. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Other types of modifications include, for example, “caps”, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (*e.g.*, methylphosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example, proteins (*e.g.*, nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with

intercalators (*e.g.*, acridine, psoralen, etc.), those containing chelators (*e.g.*, metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide(s). Further, any of the hydroxyl groups ordinarily present in the sugars may be replaced, for example, by phosphonate groups, phosphate groups, protected by standard protecting groups, or activated to prepare additional linkages to additional nucleotides, or may be conjugated to solid supports. The 5' and 3' terminal OH can be phosphorylated or substituted with amines or organic capping group moieties of from 1 to 20 carbon atoms. Other hydroxyls may also be derivatized to standard protecting groups. Polynucleotides can also contain analogous forms of ribose or deoxyribose sugars that are generally known in the art, including, for example, 2'-O-methyl-, 2'-O-allyl, 2'-fluoro- or 2'-azido-ribose, carbocyclic sugar analogs, alpha- or beta-anomeric sugars, epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and abasic nucleoside analogs such as methyl riboside. One or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein phosphate is replaced by P(O)S ("thioate"), P(S)S ("dithioate"), (O)NRi ("amidate"), P(O)R, P(O)OR', CO or CH₂ ("formacetal"), in which each R or R' is independently H or substituted or unsubstituted alkyl (1-20 C) optionally containing an ether (-O-) linkage, aryl, alkenyl, cycloalkyl, cycloalkenyl or araldyl. Not all linkages in a polynucleotide need be identical. The preceding description applies to all polynucleotides referred to herein, including RNA and DNA.

[0066] The terms "polypeptide," "oligopeptide," "peptide" and "protein" are used interchangeably herein to refer to chains of amino acids of any length. The chain may be linear or branched, it may comprise modified amino acids, and/or may be interrupted by non-amino acids. The terms also encompass an amino acid chain that has been modified naturally or by intervention; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labeling component. Also included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), as well as other modifications known in the art. It is understood that the polypeptides can occur as single chains or associated chains.

[0067] The term "expression", as used herein, generally refers to the process by which a nucleic acid sequence or a polynucleotide is transcribed from a DNA template (such as into mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is

subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as “gene product.” If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

[0068] As used herein, “operably linked”, “operable linkage”, “operatively linked”, or grammatical equivalents thereof generally refer to juxtaposition of genetic elements, e.g., a promoter, an enhancer, a polyadenylation sequence, etc., wherein the elements are in a relationship permitting them to operate in the expected manner. For instance, a regulatory element, which may comprise promoter and/or enhancer sequences, is operatively linked to a coding region if the regulatory element helps initiate transcription of the coding sequence. There may be intervening residues between the regulatory element and coding region so long as this functional relationship is maintained.

[0069] A “vector” as used herein, generally refers to a macromolecule or association of macromolecules that comprises or associates with a polynucleotide and which may be used to mediate delivery of the polynucleotide to a cell. Examples of vectors include plasmids, viral vectors, liposomes, and other gene delivery vehicles. The vector generally comprises genetic elements, e.g., regulatory elements, operatively linked to a gene to facilitate expression of the gene in a target.

[0070] As used herein, “an expression cassette” and “a nucleic acid cassette” are used interchangeably generally to refer to a combination of nucleic acid sequences or elements that are expressed together or are operably linked for expression. In some cases, an expression cassette refers to the combination of regulatory elements and a gene or genes to which they are operably linked for expression.

[0071] As used herein, the term percent “identity,” in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (e.g., BLASTP and BLASTN or other algorithms available to persons of skill) or by visual inspection. Depending on the application, the percent “identity” can exist over a region of the sequence being compared, e.g., over a

functional domain, or, alternatively, exist over the full length of the two sequences to be compared.

[0072] The term “sequence similarity,” in all its grammatical forms, refers to the degree of identity or correspondence between nucleic acid or amino acid sequences that may or may not share a common evolutionary origin.

[0073] “Percent (%) sequence identity” or “percent (%) identical to” with respect to a reference polypeptide (or nucleotide) sequence is defined as the percentage of amino acid residues (or nucleic acids) in a candidate sequence that are identical with the amino acid residues (or nucleic acids) in the reference polypeptide (nucleotide) sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for aligning sequences, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information.

[0074] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters. Alternatively, sequence similarity or dissimilarity can be established by the combined presence or absence of particular nucleotides, or, for translated sequences, amino acids at selected sequence positions (e.g., sequence motifs).

[0075] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444

(1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally Ausubel et al., *infra*).

[0076] “Homologous,” in all its grammatical forms and spelling variations, refers to the relationship between two proteins that possess a “common evolutionary origin,” including proteins from superfamilies in the same species of organism, as well as homologous proteins from different species of organism. Such proteins (and their encoding nucleic acids) have sequence homology, as reflected by their sequence similarity, whether in terms of percent identity or by the presence of specific residues or motifs and conserved positions.

[0077] However, in common usage and in the instant application, the term “homologous,” when modified with an adverb such as “highly,” may refer to sequence similarity and may or may not relate to a common evolutionary origin.

[0078] The term “transgene” refers to a polynucleotide that is introduced into a cell and is capable of being transcribed into RNA and optionally, translated and/or expressed under appropriate conditions. In aspects, it confers a desired property to a cell into which it was introduced, or otherwise leads to a desired therapeutic or diagnostic outcome. In another aspect, it may be transcribed into a molecule that mediates RNA interference, such as miRNA, siRNA, or shRNA.

[0079] As used herein, “isolated molecule” (where the molecule is, for example, a polypeptide, a polynucleotide, or fragment thereof) is a molecule that by virtue of its origin or source of derivation (1) is not associated with one or more naturally associated components that accompany it in its native state, (2) is substantially free of one or more other molecules from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature.

[0080] The term “subject” encompasses a cell, tissue, or organism, human or non-human, whether *in vivo*, *ex vivo*, or *in vitro*, male or female. The term subject is inclusive of mammals including humans.

[0081] The term “mammal” encompasses both humans and non-humans and includes but is not limited to humans, non-human primates, canines, felines, murines, bovines, equines, pteropines, and porcines.

[0082] As used herein, a “vector,” refers to a recombinant plasmid or virus that comprises a nucleic acid to be delivered into a host cell, either *in vitro* or *in vivo*. A “recombinant viral vector” refers to a recombinant polynucleotide vector comprising one or more heterologous sequences (*i.e.* a nucleic acid sequence not of viral origin). In the case of recombinant AAV

vectors, the recombinant nucleic acid is flanked by at least one inverted terminal repeat sequence (ITR). In some embodiments, the recombinant nucleic acid is flanked by two ITRs.

[0083] The phrase “pharmaceutical composition” refers to a mixture containing a specified amount of a therapeutic, *e.g.*, a therapeutically effective amount, of a therapeutic compound in a pharmaceutically acceptable carrier to be administered to a mammal, *e.g.*, a human, in order to treat a disease.

[0084] The phrase “pharmaceutically acceptable carrier” means buffers, carriers, and excipients suitable for use in contact with the tissues of human beings and animals without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit/risk ratio.

[0085] Each embodiment described herein may be used individually or in combination with any other embodiment described herein.

II. Bat Pluripotent Stem Cells (BiPS)

[0086] The disclosure is based, in part, upon the discovery that bat induced pluripotent stem cells (iPSC) (BiPS) can be produced and are stable in culture, proliferate, readily differentiate into all three germ layers, and form complex embryoid bodies, including organoids.

[0087] Accordingly, compositions and methods of making and using the BiPS are provided herein.

BiPS of the Disclosure

[0088] In some embodiments, BiPS are provided. In some embodiments the pluripotent state of the BiPS is characterized by the expression of one or more factors selected from the group of Klf4, Klf17, Essrb, Tfcp2l1, Tfe3, Dppa, Oct4, Sox2, Nanog, and Dusp6. In some embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9, or all 10 factors are expressed in the BiPS. Pluripotent stem cells can be classified into at least naïve and primed stem cell states based on the growth characteristics in vitro and their potential rise to all somatic lineages and the germ line in chimeras. In some embodiments, the BiPS are in a naïve pluripotent state. In some embodiments, the BiPS are further characterized by the expression of one or more factors for example Otx2 or Zic2.

[0089] Bats are divided in two groups: fruit-eating megabats, and the echolocating microbats. Megabats are further divided into Yinpterochiroptera that include the Pteropodidae, or megabat family, as well as the family of Rhinolophoidea, and

Yangochiroptera. Rhinolophoidea can be further divided into Hipposideridae, Craseonycteridae, Megadermatidae, Rhinopomatidae and Rhinolophidae. In some embodiments, the BiPS can be derived from isolated source bat cells from embryonic, young, or adult bats. In some embodiments, the bat is a *Rhinolophus* bat. In some embodiments the bat is a wild horseshoe bat (*Rhinolophus ferrumequinum*). In some embodiments, the bat is a *Myotis* bat or a *Myotis myotis* bat. In some embodiments, embryonic fibroblasts (BEF) cells can be isolated from the bat. In some embodiments, adult fibroblasts cells can be isolated from the bat.

[0090] A BiPS of the disclosure may be isolated, substantially isolated, purified or substantially purified. The iPSC is isolated or purified if it is completely free of any other components, such as culture medium, other cells of the disclosure or other cell types. The iPSC is substantially isolated if it is mixed with carriers or diluents, such as culture medium, which will not interfere with its intended use. Alternatively, the iPSC of the disclosure may be present in a growth matrix or immobilized on a surface as discussed below.

[0091] In some embodiments, the BiPS are further differentiated into embryonic bodies. In some embodiments, the BiPS can be further differentiated into endoderm (Afp+), ectoderm (Tbxt+), and mesoderm (Pax6+). The embryonic bodies derived from the BiPS can be further differentiated into three-dimensional structures comprising the three germ layer markers.

[0092] Techniques for producing and culturing iPSCs are well known to a person skilled in the art. Suitable conditions are discussed below.

Method of Producing an BiPS of the Disclosure

[0093] The one aspect, the disclosure also provides a method of producing a population of BiPS, comprising culturing source bat cells under conditions which reprogram the source bat cells to produce the BiPS. Any of the source bat cells discussed above may be used.

[0094] Induced pluripotent stem cells (iPSCs) are a type of pluripotent stem cell that can be generated (reprogrammed) from a non-pluripotent cell of a multicellular organism, such as a somatic cell. iPSCs are characterized in that they propagate indefinitely and can differentiate into the three germ layers endoderm, mesoderm and ectoderm, form embryonic bodies, develop into teratomas in vivo, and can form fully differentiated tissues including but not limited to neurons, cardiomyocytes, hepatocytes, and immune cells. Typically, iPSCs express a group of markers for stem cells on the surface of the cell such as SSEA-4, TRA-1-

60, and CD30, though expressed markers and timing of expression for the markers can vary (for example as described in Pomeroy et al., *Stem Cells Transl Med.* (2016) 5(7): 870–882). Recently, two protocols to produce bat reprogrammed stem cells were published (Mo et al., *Theriogenology* (2014)15;82(2):283-93, Aurine et al., *BioRxiv* (2019)). However, neither of the protocols provides for BiPS that are able to differentiate into the three germ layers or form embryonic bodies or teratomas in vivo. Thus, lack of access to robust cell models has hindered further understanding of bat asymptomatic response to viral pathogens.

[0095] To establish bats as new model study species, initially the Yamanaka reprogramming protocol based on four reprogramming factors (Oct4, Sox2, Klf4, and cMyc) (Takahashi K. et al., *Cell* (2006) 25;126(4):663-76, and . Hochedlinger K. et al., *Cold Spring Harb Perspect Biol.* (2015) 7(12): a019448), that is highly effective in mice, humans, and other mammalian species (e.g., dog, pig, marmoset) was tried to produce induced pluripotent stem cells (iPSCs) from a wild horseshoe bat (*Rhinolophus ferrumequinum*). However, the protocol failed to produce BiPS that were stable in culture, and that proliferated. Though the protocols failed, the Yamanaka factors triggered the formation of rudimentary stem cell-like colonies even though they ceased to expand.

[0096] Here, methods of making BiPS are provided that overcome these problems.

[0097] The method preferably comprises culturing the source bat cells with a Sendai virus system, a retroviral system, a lentiviral system, microRNA or other reprogramming factors which is/are capable of reprogramming the source bat cells to produce the BiPS. In some embodiments, the method of making bat iPSCs comprises (i) reprogramming isolated bat cells with Oct4, Sox2, cMyc, and Klf4 factors; (ii) culturing the reprogrammed cells in a medium comprising FGF, Leukemia inhibitory factor (Lif), SCF, and Forskolin until colonies appear; and (iii) splitting cells using a low concentration EDTA buffer.

[0098] In some embodiments, the reprogramming factors can be delivered to the bat cells with viruses such as a Sendai virus, retrovirus, AAV, nonviral vector systems, physical delivery, mechanical and chemical methods, or with mRNA delivery. In some embodiments, the reprogramming factors comprise Oct4, Sox2, cMyc, and Klf4 factors. In some embodiments, the reprogramming factors comprise additional factors.

[0099] In some embodiments, the method comprises culturing the cells in a feeder free medium. In some embodiments, the cells can be cultured on feeder cells, such as CF1 mouse embryonic fibroblasts.

[0100] In some embodiments, the feeder cell free or the feeder cell culture medium comprises FGF, Leukemia inhibitory factor (Lif), SCF, and Forskolin. In some embodiments, the Lif is at a concentration of 10^4 U/ml. In some embodiments, the FGF is at a concentration of 100 ng/ml. In some embodiments, the SCF is at a concentration of 100 ng/ml. In some embodiments, the Forskolin is at a concentration of 20 nM. In some embodiments, the Lif is at a concentration of 10^4 U/ml, the FGF is at a concentration of 100 ng/ml, the SCF is at a concentration of 100 ng/ml and the Forskolin is at a concentration of 20 nM. In some embodiments, the Lif is at a concentration of 10^4 to 10^5 U/ml. In some embodiments, the FGF is at a concentration of 100 ng/ml. In some embodiments, the SCF is at a concentration of 10-100 ng/ml. In some embodiments, the Forskolin is at a concentration of 5-20 nM. In some embodiments, the Lif is at a concentration of 10^4 to 10^5 U/ml, the FGF is at a concentration of 4-100 ng/ml, the SCF is at a concentration of 10-100 ng/ml and the Forskolin is at a concentration of 5-20 nM. In some embodiments, the concentration of Lif is 40%, 30%, 20%, 10%, or 5% more or less than 10^4 U/ml. In some embodiments, the concentration of FGF is 40%, 30%, 20%, 10%, or 5% more or less than 100 ng/ml. In some embodiments, the concentration of SCF is 40%, 30%, 20%, 10%, or 5% more or less than 100 ng/ml. In some embodiments, the concentration of Forskolin is 40%, 30%, 20%, 10%, or 5% more or less than 20 nM. In some embodiments, the concentration of Lif is about 10^4 U/ml. In some embodiments, the concentration of FGF is about 100 ng/ml. In some embodiments, the concentration of SCF is about 100 ng/ml. In some embodiments, the concentration of Forskolin is about 20 nM.

[0101] In some embodiments, the BiPS are passaged, i.e. moved into fresh media. In some embodiments the BiPS are passaged every 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 days. In some embodiments, the BiPS are passaged every 5 days. In some embodiments, the BiPS are passaged when they are 50%, 60%, 70%, 80%, 90%, or 100% confluent. In some embodiments, the BiPS are passaged before they are confluent. In some embodiments, the feeder cells are freshly changed every passage. In some embodiments, the feeder cells are irradiated. In some embodiments, the BiPS are passaged using a low concentration EDTA buffer. In some embodiments, the BiPS are passaged using a low concentration EDTA buffer with a EDTA concentration less than 0.48 mM EDTA. In some embodiments the BiPS can be passaged indefinitely. In some embodiments the BiPS can be passaged at least to passage 78.

[0102] In some embodiments, the BiPS are further differentiated into embryonic bodies. In some embodiments, the BiPS can be further differentiated into endoderm (Afp+),

ectoderm (Tbxt+), and mesoderm (Pax6+). The embryonic bodies can be further differentiated into three-dimensional structures comprising the three germ layer markers.

[0103] In some embodiments, a medium is provided that is conducive to producing and maintaining BiPS comprising FGF, Leukemia inhibitory factor (Lif), SCF, and Forskolin. In some embodiments, the medium comprises FGF at a concentration of 20nM, Leukemia inhibitory factor (Lif) at a concentration of 10^4 U/ml, SCF at a concentration of 100 ng/ml, and Forskolin at a concentration of 100 ng/ml. In some embodiments, the Lif is at a concentration of 10^4 U/ml, the FGF is at a concentration of 100 ng/ml, the SCF is at a concentration of 100 ng/ml and the Forskolin is at a concentration of 20 nM. In some embodiments, the Lif is at a concentration of 10^4 to 10^5 U/ml. In some embodiments, the FGF is at a concentration of 100 ng/ml. In some embodiments, the SCF is at a concentration of 10-100 ng/ml. In some embodiments, the Forskolin is at a concentration of 5-20 nM. In some embodiments, the Lif is at a concentration of 10^4 to 10^5 U/ml, the FGF is at a concentration of 4-100 ng/ml, the SCF is at a concentration of 10-100 ng/ml and the Forskolin is at a concentration of 5-20 nM. In some embodiments, the medium comprises FGF at a concentration of 40%, 30%, 20%, 10%, or 5% more or less than 20nM, Leukemia inhibitory factor (Lif) at a concentration of 40%, 30%, 20%, 10%, or 5% more or less than 10^4 U/ml, SCF at a concentration of 40%, 30%, 20%, 10%, or 5% more or less than 100 ng/ml, and Forskolin at a concentration of 40%, 30%, 20%, 10%, or 5% more or less than 100 ng/ml.

[0104] An important method for reprogramming is the use of messenger RNA specific for the reprogramming factors since this does not involve any genetic modification of the cells and the risk of tumorigenesis. Another method is to produce from the reprogramming genes, recombinant proteins modified to permit their penetration of the plasma and nuclear membranes. Other reprogramming factors include, but are not limited to, small compounds synthesized through medicinal chemistry.

[0105] The method preferably further comprises isolating clonal lines of BiPS of the disclosure. For instance, the method preferably further comprises isolating clonal lines of BiPS of the disclosure by limiting dilution or the manual 'picking' of individual colonies.

[0106] Standard methods known in the art may be used to determine the detectable expression and level of expression of the various markers discussed above. Suitable methods

include, but are not limited to, immunocytochemistry, flow cytometry, western blotting and quantitative PCR.

III. Viruses and Viral Sequences

[0107] Provided herein are also methods and compositions for using the viruses and viral sequences identified herein from the bat pluripotent stem cells. In particular, viruses, viral families, and viral sequences are disclosed herein.

[0108] In some embodiments, the method of obtaining viral sequences from bat iPSCs, comprises obtaining bat iPSCs; identifying viral sequences residing in the bat iPSC genome or intracellular virus genome; and assembling the viral sequences. In some embodiments, the bat iPSCs (BiPS) are produced by the methods described above. In some embodiments, the nucleic acid sequences are obtained by sequencing RNA transcripts such as RNA seq, long read sequencing such as ss Iso-seq (PacBio), or sequencing the genomic DNA such as by DNA sequencing of samples derived from the BiPS. In some embodiments, amino acid sequences can be obtained by LC-MS or amino acid sequencing of samples derived from the BiPS. In some embodiments the samples can be derived directly from the BiPS or the medium BiPS were grown in. In some embodiments, the samples can be derived from differentiated cells derived from the BiPS.

[0109] In some embodiments, the obtained nucleic acid sequences are assembled into longer nucleic acid sequences. Short and long assembled sequences can be classified as potentially viral origin or non-viral origin for example as described in **Example 10**. The sequences can be further classified into virus clades by comparing with known sequences from virus nucleic acids in databases such as the NCBI Assembly database (www.ncbi.nlm.nih.gov/assembly) or Virus Pathogen Resource (www.viprbrc.org/brc/home.spg?decorator=vipr). Nucleic acid sequences can be also classified using metagenomic classifiers, such as Kraken2.

[0110] **TABLE 1** Exemplary virus families and viruses found in a taxonomic distribution of virome reads from BiPS as determined by the metagenomic classifier Kraken2.

TABLE 1

Virus Family	Virus
Retroviridae	ND
Picornavirales	Rotavirus

Coronaviridae	ND
Hantaviridae	ND
Herpesvirales	ND
Poxviridae	ND
Adenoviridae	ND
Papillomaviridae	ND
Myoviridae	ND
Flaviviridae	ND
Siphoviridae	ND
Baculoviridae	ND
Duplondaviria	ND
Riboviria	ND
Filoviridae	Ebola
Filoviridae	Cueva
Filoviridae	Dianlovirus
Mononegavirales	ND

ND, virus was not determined

[0111] More exemplary viral families, viruses and sequences identified from the BiPS are shown in **TABLE A**.

[0112] In some embodiments the nucleic acid sequences are derived from sequencing transcripts derived from the BiPS by Iso-seq. Exemplary Iso-Seq derived sequences are set forth in SEQ ID NO: 1-7. The sequences can be classified using Kraken 2. Exemplary Kraken 2 classification of Iso-Seq derived sequences and bat genome sequences are presented in **TABLE 2**. Exemplary full-length retrovirus sequence identified are RFe-V-MD1, RFe-V-MD2 RFe-V-MD3 RFe-V-MD4, and RFe-V-MD5, set forth in SEQ ID NO: 1-7. A detailed analysis of the sequence of RFe-V-MD1 is shown in **FIG. 9D**, showing the location of the Env, Pol, and Gag proteins in the genome. A detailed analysis of RFe-V-MD2 sequences is shown in **FIG. 9E**. The sequences comprise Columbidae/Falconidae herpesvirus and Sindbis virus sequences as shown. Detailed alignments of exemplary protein sequences are shown in **FIG. 11A**. A detailed analysis of RFe-V-MD3 sequences show similarities with HKHD40, HKNPC60, human respiratory syncytial virus and SARS-CoV2 (**FIG. 9G**). Detailed alignments of exemplary protein sequences of the SARS-CoV2 similar sequence with the

sequence of a SARS-CoV2 virus isolated from a patient is shown in **FIG. 11C**. A detailed analysis and comparison of RFe-V-MD4 sequences with Scotophilus bat coronavirus spike protein is shown in **FIG. 9H**.

[0113] In some embodiments, exemplary nucleic acid sequences and an alignment with known viruses such as Scotophilus bat coronavirus 512 are shown in **TABLE 3** and RaTG13 bat coronavirus are shown in **TABLE 4**.

[0114] **FIG. 11B** shows alignments of sequences identified to be similar to Lymphocystis disease virus and Erythrocytic necrosis virus.

[0115] Other viral sequences such as presented in **TABLE 3** and **TABLE 4**, or SEQ ID NO: 1-349 can be identified. Translated into amino acid sequences, and aligned with known viral sequences as described herein.

III. Antigens and T cell epitopes

[0116] Methods for identifying antigens (e.g., antigens derived from an infectious disease organism) include identifying antigens that are likely to be presented on a cell surface (e.g., presented by MHC on an infected cell or an immune cell, including professional antigen presenting cells such as dendritic cells), and/or are likely to be immunogenic. As an example, one such method may comprise the steps of: obtaining at least one of exome, transcriptome or whole genome nucleotide sequencing and/or expression data from an infected cell or an infectious disease organism (e.g., RFe-V-MD1, RFe-V-MD2 RFe-V-MD3 RFe-V-MD4, and RFe-V-MD5, Columbid/Falconid herpesvirus, and Sindbis virus), wherein the nucleotide sequencing data and/or expression data is used to obtain data representing peptide sequences of each of a set of antigens (e.g., antigens derived from the infectious disease organism); inputting the peptide sequence of each antigen into one or more presentation models to generate a set of numerical likelihoods that each of the antigens is presented by one or more MHC alleles on a cell surface, such as an infected cell of the subject, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and selecting a subset of the set of antigens based on the set of numerical likelihoods to generate a set of selected antigens. Antigens can include nucleotides or polypeptides. For example, an antigen can be an RNA sequence that encodes for a polypeptide sequence. Antigens useful in vaccines can therefore include nucleotide sequences or polypeptide sequences. Antigens can be selected that are predicted to be presented on the cell surface of a cell, such as an infected cell or an immune cell, including professional antigen presenting cells such as dendritic cells. Antigens can be selected that are predicted to be immunogenic. Exemplary antigens predicted

using the methods described herein to be presented on the cell surface by an MHC include predicted MHC class I epitopes and predicted MHC class II epitopes. Exemplary nucleic acid sequences or polypeptide sequences for antigen prediction are presented in SEQ ID NO: 1-349, **FIG. 9D-9H** and **FIG. 11A-11C**, **TABLE 3** and **TABLE 4**.

[0117] Protein sequences for the desired antigen are analyzed for potential HLA specific antigens by using for example the SYFPEITHI algorithm (Rammensee *et al.* (1999) *Immunogenetics* 50:213-219), and the artificial neural network (ANN) and stabilized matrix method (SMM) algorithms from IEDB (Peters *et al.* (2005) *PLoS Biol.* 3:e91). Peptides are selected based on a predicted binding value of either >21 for SYFPEITHY, <6000 for ANN, or <600 for SMM. Selected peptides are synthesized.

[0118] Binding assays can be performed using a fluorescence polarization (FP) assay as previously described (*e.g.*, Buchi *et al.* (2004) *Biochemistry* 43:14852-14863; Sette *et al.* (1994) *Mol. Immunol.* 31:813-822). To determine binding capacity of the peptides, percentage inhibition relative to controls can be determined in an FP competition assay with the placeholder peptide.

[0119] In some embodiments, the peptides bound to the pMHC multimers are from an unbiased library of peptides derived from the antigen. In some embodiments, the peptides are 9-mers. In some embodiments, the peptides bound to the pMHCI multimers are 9-mers which include an HLA-A2 binding motif with key amino acids at positions 2 and 9 which can include isoleucine (I), valine (V) or leucine (L).

[0120] In some embodiments, the library comprises all k-mer peptides produced by transcription and translation of any polynucleotide sequence of interest, for example, *in silico* production of the transcription and translation products of both the forward and reverse strands of a genome or metagenome in all six reading frames.

[0121] In some embodiments, a library of the disclosure comprises all k-mer peptides that can be derived from *in silico* translation of an exome of interest. In some embodiments, a library of the disclosure comprises all k-mer peptides that can be derived from *in silico* translation of a transcriptome of interest. In some embodiments, a library of the disclosure comprises all k-mer peptides that can be derived from a proteome of interest. In some embodiments, a library of the disclosure comprises all k-mer peptides that can be derived from *in silico* translation of an ORFeome of interest. In some embodiments, an algorithm can be used to select peptides in a peptide library. For example, an algorithm can be used to predict peptides most likely to fold or dock in an MHC/HLA binding pocket, and peptides above a certain threshold value can be selected for inclusion in the library.

[0122] In some embodiments, a library of the disclosure comprises all peptides that can be derived from *in silico* transcription and translation or translation of a group of genomes, proteomes, transcriptomes, ORFeomes, or any combination thereof. In some embodiments, the peptides are derived from *in silico* transcription and translation or translation of polynucleotide sequences from a group of samples, for example, clinical samples from a patient population, or a group of pathogen genomes.

[0001] One or more polypeptides encoded by an antigen nucleotide sequence can comprise at least one of: a binding affinity with MHC with an IC₅₀ value of less than 1000nM, for MHC Class I peptides a length of 8-15, 8, 9, 10, 11, 12, 13, 14, or 15 amino acids, presence of sequence motifs within or near the peptide promoting proteasome cleavage, and presence or sequence motifs promoting TAP transport. For MHC Class II peptides a length 6-30, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 amino acids, presence of sequence motifs within or near the peptide promoting cleavage by extracellular or lysosomal proteases (e.g., cathepsins) or HLA-DM catalyzed HLA binding.

[0002] One or more antigens can be presented on the surface of an infected cell (e.g., a RFe-V-MD1, RFe-V-MD2 RFe-V-MD3 RFe-V-MD4, and RFe-V-MD5, Columbid/Falconid herpesvirus, or Sindbis virus infected cell).

[0003] One or more antigens can be immunogenic in a subject having or suspected to have an infection (e.g., a RFe-V-MD1, RFe-V-MD2 RFe-V-MD3 RFe-V-MD4, and RFe-V-MD5, Columbid/Falconid herpesvirus, or Sindbis virus infection), e.g., capable of eliciting a T cell response or a B cell response in the subject. One or more antigens can be immunogenic in a subject at risk of an infection (e.g., a RFe-V-MD1, RFe-V-MD2 RFe-V-MD3 RFe-V-MD4, and RFe-V-MD5, Columbid/Falconid herpesvirus, or Sindbis virus infection), e.g., capable of eliciting a T cell response or a B cell response in the subject that provides immunological protection (*i.e.*, immunity) against the infection, e.g., such as stimulating the production of memory T cells, memory B cells, or antibodies specific to the infection.

[0004] One or more antigens can be capable of eliciting a B cell response, such as the production of antibodies that recognize the one or more antigens (e.g., antibodies that recognize a RFe-V-MD1, RFe-V-MD2 RFe-V-MD3 RFe-V-MD4, and RFe-V-MD5, Columbid/Falconid herpesvirus, and Sindbis virus antigen and/or virus). Antibodies can recognize linear polypeptide sequences or recognize secondary and tertiary structures. Accordingly, B cell antigens can include linear polypeptide sequences or polypeptides having

secondary and tertiary structures, including, but not limited to, full-length proteins, protein subunits, protein domains, or any polypeptide sequence known or predicted to have secondary and tertiary structures. In general, antigens capable of eliciting a B cell response to an infection are antigens found on the surface of an infectious disease organism (*e.g.*, RFe-V-MD1, RFe-V-MD2 RFe-V-MD3 RFe-V-MD4, and RFe-V-MD5, Columbid/Falconid herpesvirus, and Sindbis virus). Exemplary antigens capable of eliciting a B cell response include, but are not limited to, ORF1ab, spike (S), envelope (E), membrane (M), and nucleocapsid (N).

[0005] One or more antigens that induce an autoimmune response in a subject can be excluded from consideration in the context of vaccine generation for a subject.

[0006] The size of at least one antigenic peptide molecule (*e.g.*, an epitope sequence) can comprise, but is not limited to, about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 31, about 32, about 33, about 34, about 35, about 36, about 37, about 38, about 39, about 40, about 41, about 42, about 43, about 44, about 45, about 46, about 47, about 48, about 49, about 50, about 60, about 70, about 80, about 90, about 100, about 110, about 120 or greater amino molecule residues, and any range derivable therein. In specific embodiments the antigenic peptide molecules are equal to or less than 50 amino acids.

[0007] Antigenic peptides and polypeptides can be: for MHC Class I 15 residues or less in length and usually consist of between about 8 and about 11 residues, particularly 9 or 10 residues; for MHC Class II, 6-30 residues, inclusive.

[0123] In some embodiments, a recombinant cell is provided comprising a nucleic acid or polypeptide set forth in SEQ ID NO: 1-349. The recombinant cells can be used in therapeutic development, such as vaccines, small molecules and biologics. In some embodiments, a recombinant cell is provided comprising a nucleic acid or protein or part thereof set forth in **FIG. 9D-9H** and **FIG. 11A-11C, TABLE 3, and TABLE 4**. In some embodiments, the recombinant cell expresses a protein encoded by the nucleic acid or a portion thereof acid or a polypeptide set forth in SEQ ID NO: 1-349. In some embodiments, the recombinant cell expresses a protein encoded by the nucleic acid or a portion thereof acid set forth in **FIG. 9D-9H** and **FIG. 11A-11C, TABLE 3, and TABLE 4**. In some embodiments the recombinant cell is used to assay for suitable antigens. In some embodiments the recombinant cell is used to produce a selected antigen.

IV. Pharmaceutical Compositions

[0124] The present disclosure also features pharmaceutical compositions that contain a therapeutically effective amount of one or more T cell epitopes, nucleic acids coding for T cells epitopes or peptides. The composition can be formulated for use in a variety of drug delivery systems. One or more physiologically acceptable excipients or carriers can also be included in the composition for proper formulation.

[0125] In various embodiments, the pharmaceutical compound includes an acceptable pharmaceutically acceptable carrier. The carrier(s) should be “acceptable” in the sense of being compatible with the other ingredients of the formulations and not deleterious to the subject. Pharmaceutically acceptable carriers include buffers, solvents, dispersion media, coatings, isotonic and absorption delaying agents, and the like, that are compatible with pharmaceutical administration. In one embodiment the pharmaceutical composition is administered orally and includes an enteric coating suitable for regulating the site of absorption of the encapsulated substances within the digestive system or gut.

[0126] Pharmaceutical compositions containing a therapeutic, such as those disclosed herein, can be presented in a dosage unit form and can be prepared by any suitable method. A pharmaceutical composition should be formulated to be compatible with its intended route of administration. Useful formulations can be prepared by methods well known in the pharmaceutical art. For example, see *Remington's Pharmaceutical Sciences*, 18th ed. (Mack Publishing Company, 1990).

[0127] Pharmaceutical formulations, in some embodiments, are sterile. Sterilization can be accomplished, for example, by filtration through sterile filtration membranes. Where the composition is lyophilized, filter sterilization can be conducted prior to or following lyophilization and reconstitution.

Vaccines

[0128] Disclosed herein is an immunogenic composition, e.g., a vaccine composition, capable of raising a specific immune response, e.g., a tumor-specific immune response. Vaccine compositions typically comprise a plurality of viral antigens, e.g., selected using a method described herein. Vaccine compositions can also be referred to as vaccines.

[0129] The viral nucleic acids, proteins, antigens, and T cell epitopes can be used to design prophylactic or therapeutic vaccines comprising such composition (e.g., pharmaceutical compositions) for immunizing subjects at risk of contracting, or subjects

having already contacted, a virus set forth in **TABLE 1** or **TABLE A**. In certain embodiments, the vaccine is a subunit vaccine. In certain embodiments, the vaccine elicits a protective immune reaction against a plurality of viruses (*e.g.*, RFe-V-MD1, RFe-V-MD2 RFe-V-MD3 RFe-V-MD4, or RFe-V-MD5). In certain embodiments, the vaccine elicits a protective immune reaction against a virus set forth in **TABLE 1** or **TABLE A**.

[0130] In some embodiments, the vaccine comprises a recombinant nucleic acid molecule comprising one or more promoter and a nucleic acid encoding for a T cell epitope. In some embodiments the nucleic acid is set forth in SEQ ID NO: 1-349, **TABLE 3**, **TABLE 4**, or a functional portion thereof.

[0131] A vaccine composition of the disclosure can comprise a peptide composition(s) comprising the T cell epitope(s). Alternatively, a vaccine composition of the disclosure can comprise a nucleic acid composition, *e.g.*, an RNA composition or DNA composition, encoding the T cell epitope(s). For such nucleic acid vaccines, suitable regulatory sequences are included such that the peptide epitope is expressed from the nucleic acid (RNA or DNA) in cells of the subject being immunized. In some embodiments, the nucleic acids or the peptides are synthetic.

[0132] A vaccine can contain between 1 and 30 peptides, 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, or 30 different peptides, 6, 7, 8, 9, 10 11, 12, 13, or 14 different peptides, or 12, 13 or 14 different peptides. Peptides can include post-translational modifications. A vaccine can contain between 1 and 100 or more nucleotide sequences, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more different nucleotide sequences, 6, 7, 8, 9, 10 11, 12, 13, or 14 different nucleotide sequences, or 12, 13 or 14 different nucleotide sequences. A vaccine can contain between 1 and 30 viral antigen sequences, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more different viral antigen sequences, 6, 7, 8, 9, 10 11, 12, 13, or 14 different viral antigen sequences, or 12, 13 or 14 different viral antigen sequences.

[0133] In some embodiments, the pharmaceutical composition comprises a plurality of (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) proteins or peptides and a pharmaceutically acceptable

carrier or excipient. A pharmaceutical composition comprising a nucleic acid encoding the mRNA of claim 44 or the protein or peptide of any one of claims 46-48 and a pharmaceutically acceptable carrier or excipient.

[0134] In some embodiments, the pharmaceutical composition comprises one or more nucleic acids encoding a plurality of (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) mRNAs and a pharmaceutically acceptable carrier or excipient.

[0135] In one embodiment, antigens or T cell epitopes are for example ORF1ab, spike (S), envelope (E), membrane (M) and nucleocapsid (N), RNA polymerases, kinases, and viral proteases. Exemplary antigens are shown in **FIG. 9D-9H** and **FIG. 11A-11C**, exemplary nucleic acids encoding antigens or portions of antigens are set forth in **TABLE 3** and **TABLE 4**.

[0136] In certain embodiments, the two or more of the T cell peptides collectively recognize MHC molecules in at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% of the human population. In certain embodiments, the vaccine contains individualized components according to the personal need (*e.g.*, MHC variants) of the particular patient.

[0137] In one embodiment, different peptides and/or polypeptides or nucleotide sequences encoding them are selected so that the peptides and/or polypeptides capable of associating with different MHC molecules, such as different MHC class I molecule. In some aspects, one vaccine composition comprises coding sequence for peptides and/or polypeptides capable of associating with the most frequently occurring MHC class I molecules. Hence, vaccine compositions can comprise different fragments capable of associating with at least 2 preferred, at least 3 preferred, or at least 4 preferred MHC class I molecules.

[0138] The vaccine composition can be capable of raising a specific cytotoxic T-cell response and/or a specific helper T-cell response.

[0139] A vaccine composition of the disclosure can comprise one or more short (*e.g.*, 8-35 amino acids) peptides as the immunostimulatory agent. In certain embodiments, a cell surface antigen sequence is incorporated into a larger carrier polypeptide or protein, to create a chimeric carrier polypeptide or protein that comprises the T cell epitope(s). This chimeric carrier polypeptide or protein can then be incorporated into the vaccine composition.

[0140] Recombinant cells can be engineered to express proteins and peptides of the disclosure. Vectors can be designed for the expression of cell surface antigens (*e.g.* nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, cell surface antigens can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. The cell surface antigens can be purified from the recombinant cells and used in antibody development or further formulated into pharmaceutical compositions. Additionally or alternatively, the recombinant cells expressing the cell surface antigens can be used for producing antibodies or T cells specific to the cell surface antigens.

[0141] It is understood that a peptide can be expressed from a nucleic acid (*e.g.*, an mRNA) in a cell of the subject. Exemplary methods of producing peptides by translation *in vitro* or *in vivo* are described in U.S. Patent Application Publication No. 2012/0157513 and He *et al.*, *J. Ind. Microbiol. Biotechnol.* (2015) 42(4):647-53. The present disclosure provides a composition (*e.g.*, pharmaceutical composition) comprising one or more nucleic acids (*e.g.*, mRNAs) encoding one or more cell surface antigens or derived peptides. It is understood that a peptide can be expressed from a nucleic acid (*e.g.*, an mRNA) in a cell of the subject. Exemplary methods of producing peptides by translation *in vitro* or *in vivo* are described in U.S. Patent Application Publication No. 2012/0157513 and He *et al.*, *J. Ind. Microbiol. Biotechnol.* (2015) 42(4):647-53. The present disclosure provides a composition (*e.g.*, pharmaceutical composition) comprising one or more nucleic acids (*e.g.*, mRNAs) encoding one or more peptides disclosed herein, optionally further comprising a pharmaceutically acceptable carrier or excipient. In certain embodiments, the composition comprises nucleic acid sequences encoding two or more (*e.g.*, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, 11 or more, 12 or more, 13 or more, 14, or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, or 20 or more) of the peptides disclosed herein. In certain embodiments, the two or more peptides are derived from the same cell surface antigen. In certain embodiments, the two or more peptides are derived from at least two different cell surface antigens. In certain embodiments, the two or more peptides collectively are recognized by MHC molecules in at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% of the human population. In certain embodiments, the vaccine contains individualized

components according to the personal need (*e.g.*, MHC variants) of the particular patient. In certain embodiments, each of the nucleic acids further comprises one or more expression control sequences (*e.g.*, promoter, enhancer, translation initiation site, internal ribosomal entry site, and/or ribosomal skipping element) operably linked to one or more of the peptide coding sequences.

[0142] A vaccine composition can further comprise an adjuvant and/or a carrier. Examples of useful adjuvants and carriers are given herein below. A composition can be associated with a carrier such as *e.g.* a protein or an antigen-presenting cell such as *e.g.* a dendritic cell (DC) capable of presenting the peptide to a T-cell.

[0143] Adjuvants are any substance whose admixture into a vaccine composition increases or otherwise modifies the immune response to a viral antigen. Carriers can be scaffold structures, for example a polypeptide or a polysaccharide, to which a viral antigen, is capable of being associated. Optionally, adjuvants are conjugated covalently or non-covalently.

[0144] The ability of an adjuvant to increase an immune response to an antigen is typically manifested by a significant or substantial increase in an immune-mediated reaction, or reduction in disease symptoms. For example, an increase in humoral immunity is typically manifested by a significant increase in the titer of antibodies raised to the antigen, and an increase in T-cell activity is typically manifested in increased cell proliferation, or cellular cytotoxicity, or cytokine secretion. An adjuvant may also alter an immune response, for example, by changing a primarily humoral or Th response into a primarily cellular, or Th response.

[0145] Suitable adjuvants include, but are not limited to 1018 ISS, alum, aluminium salts, Amplivax, AS15, BCG, CP-870,893, CpG7909, CyaA, dSLIM, GM-CSF, IC30, IC31, Imiquimod, ImuFact IMP321, IS Patch, ISS, ISCOMATRIX, JuvImmune, LipoVac, MF59, monophosphoryl lipid A, Montanide IMS 1312, Montanide ISA 206, Montanide ISA 50V, Montanide ISA-51, OK-432, OM-174, OM-197-MP-EC, ONTAK, PepTel vector system, PLG microparticles, resiquimod, SRL172, Virosomes and other Virus-like particles, YF-17D, VEGF trap, R848, beta-glucan, Pam3Cys, Aquila's QS21 stimulon (Aquila Biotech, Worcester, Mass., USA) which is derived from saponin, mycobacterial extracts and synthetic bacterial cell wall mimics, and other proprietary adjuvants such as Ribi's Detox. Quil or Superfos. Adjuvants such as incomplete Freund's or GM-CSF are useful. Several immunological adjuvants (*e.g.*, MF59) specific for dendritic cells and their preparation have

been described previously (Dupuis M, et al., Cell Immunol. 1998; 186(1):18-27; Allison A C; Dev Biol Stand. 1998; 92:3-11). Also cytokines can be used. Several cytokines have been directly linked to influencing dendritic cell migration to lymphoid tissues (e.g., TNF-alpha), accelerating the maturation of dendritic cells into efficient antigen-presenting cells for T-lymphocytes (e.g., GM-CSF, IL-1 and IL-4) (U.S. Pat. No. 5,849,589, specifically incorporated herein by reference in its entirety) and acting as immunoadjuvants (e.g., IL-12) (Gabrilovich D I, et al., J Immunother Emphasis Tumor Immunol. 1996 (6):414-418).

[0146] CpG immunostimulatory oligonucleotides have also been reported to enhance the effects of adjuvants in a vaccine setting. Other TLR binding molecules such as RNA binding TLR 7, TLR 8 and/or TLR 9 may also be used.

[0147] Other examples of useful adjuvants include, but are not limited to, chemically modified CpGs (e.g. CpR, Idera), Poly(I:C)(e.g. polyi:CI2U), non-CpG bacterial DNA or RNA as well as immunoactive small molecules and antibodies such as cyclophosphamide, sunitinib, bevacizumab, celebrex, NCX-4016, sildenafil, tadalafil, vardenafil, sorafenib, XL-999, CP-547632, pazopanib, ZD2171, AZD2171, ipilimumab, tremelimumab, and SC58175, which may act therapeutically and/or as an adjuvant. The amounts and concentrations of adjuvants and additives can readily be determined by the skilled artisan without undue experimentation. Additional adjuvants include colony-stimulating factors, such as Granulocyte Macrophage Colony Stimulating Factor (GM-CSF, sargramostim).

[0148] A vaccine composition of the disclosure can comprise one or more short (e.g., 8-35 amino acids) peptides as the immunostimulatory agent. In certain embodiments, a T cell epitope sequence is incorporated into a larger carrier polypeptide or protein, to create a chimeric carrier polypeptide or protein that comprises the T cell epitope(s). This chimeric carrier polypeptide or protein can then be incorporated into the vaccine composition.

[0149] A vaccine composition can comprise more than one different adjuvant. Furthermore, a therapeutic composition can comprise any adjuvant substance including any of the above or combinations thereof. It is also contemplated that a vaccine and an adjuvant can be administered together or separately in any appropriate sequence.

[0150] A carrier (or excipient) can be present independently of an adjuvant. The function of a carrier can for example be to increase the molecular weight of in particular mutant to increase activity or immunogenicity, to confer stability, to increase the biological activity, or to increase serum half-life. Furthermore, a carrier can aid presenting peptides to T-cells. A carrier can be any suitable carrier known to the person skilled in the art, for example a protein

or an antigen presenting cell. A carrier protein could be but is not limited to keyhole limpet hemocyanin, serum proteins such as transferrin, bovine serum albumin, human serum albumin, thyroglobulin or ovalbumin, immunoglobulins, or hormones, such as insulin or palmitic acid. For immunization of humans, the carrier is generally a physiologically acceptable carrier acceptable to humans and safe. However, tetanus toxoid and/or diphtheria toxoid are suitable carriers. Alternatively, the carrier can be dextrans for example sepharose.

[0151] Cytotoxic T-cells (CTLs) recognize an antigen in the form of a peptide bound to an MHC molecule rather than the intact foreign antigen itself. The MHC molecule itself is located at the cell surface of an antigen presenting cell. Thus, an activation of CTLs is possible if a trimeric complex of peptide antigen, MHC molecule, and APC (antigen presenting cell) is present. Correspondingly, it may enhance the immune response if not only the peptide is used for activation of CTLs, but if additionally APCs with the respective MHC molecule are added. Therefore, in some embodiments a vaccine composition additionally contains at least one antigen presenting cell.

[0152] Viral antigens can also be included in viral vector-based vaccine platforms, such as vaccinia, fowlpox, self-replicating alphavirus, marabavirus, adenovirus (*See, e.g., Tatsis et al., Adenoviruses, Molecular Therapy* (2004) 10, 616—629), or lentivirus, including but not limited to second, third or hybrid second/third generation lentivirus and recombinant lentivirus of any generation designed to target specific cell types or receptors (*See, e.g., Hu et al., Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, Immunol Rev.* (2011) 239(1): 45-61, Sakuma et al., Lentiviral vectors: basic to translational, *Biochem J.* (2012) 443(3):603-18, Cooper et al., Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, *Nucl. Acids Res.* (2015) 43 (1): 682-690, Zufferey et al., Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, *J. Virol.* (1998) 72 (12): 9873-9880). Dependent on the packaging capacity of the above mentioned viral vector-based vaccine platforms, this approach can deliver one or more nucleotide sequences that encode one or more viral antigen peptides. The sequences may be flanked by non-mutated sequences, may be separated by linkers or may be preceded with one or more sequences targeting a subcellular compartment (*See, e.g., Gros et al., Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, Nat Med.* (2016) 22 (4):433-8, Stronen et al., Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, *Science.* (2016) 352 (6291):1337-41, Lu et al., Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions, *Clin Cancer Res.* (2014) 20(

13):3401-10). Upon introduction into a host, infected cells express the viral antigens, and thereby elicit a host immune (e.g., CTL) response against the peptide(s). Vaccinia vectors and methods useful in immunization protocols are described in, e.g., U.S. Pat. No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover et al. (Nature 351:456-460 (1991)). A wide variety of other vaccine vectors useful for therapeutic administration or immunization of viral antigens, e.g., Salmonella typhi vectors, and the like will be apparent to those skilled in the art from the description herein. In some embodiments, the viral vector is a adenovirus vector.

[0153] The compositions (e.g., pharmaceutical compositions) disclosed herein may be formulated for delivery into cells (e.g., APCs, such as dendritic cells, monocytes, macrophages, or artificial APCs). In certain embodiments, the composition comprises an agent that facilitate transfection *in vitro* or *in vivo*, such as a liposome or a nanoparticle (e.g., lipid nanoparticle). In certain embodiments, the liposome or nanoparticle further comprises a binding moiety (e.g., an antibody or an antigen-binding fragment thereof) for delivering the liposome or nanoparticle to a target T cell (e.g., a professional APC). Another delivery method employs virus particles (e.g., adenovirus, adeno-associated virus, vaccinia virus, fowlpox virus, self-replicating alphavirus, marabavirus, or lentivirus). In certain embodiments, the composition comprises a pharmaceutically acceptable carrier or excipient, such as a diluent, an isotonic solution, water, etc. Excipients also can be selected for enhancement of delivery of the composition.

[0154] Suitable routes of administration and dosages for vaccines are known in the art and can be determined by a person of medical skill. In certain embodiments, the vaccine is administered parenterally, e.g., by intramuscular, intradermal, subcutaneous, intravenous, topical, nasal, or local administration. In certain embodiments, the vaccine comprising peptide(s) is administered via skin scarification. In certain embodiments, the vaccine comprising peptide(s) is administered at a dosage of 0.1-10 mg, e.g., 0.1-0.5 mg, 0.5-1 mg, 1-3 mg, 1-5 mg, or 5-10 mg of total amount per human patient. In certain embodiments, the vaccine comprises a plurality of different peptides, wherein each peptide is provided at a dosage of 0.01-0.05 mg, 0.05-0.1, or 0.1-0.5 mg per human patient. Stimulation of an anti-virus T cell immune response in a subject by the vaccine can be monitored by methods established in the art, e.g., by isolating T cells from the subject and measuring reactivity of the T cells to the viral T cell epitope(s) contained within the vaccine (see, e.g.,

Immunohistochemistry, ELISPOT, binding assays such as Biacore and ELISA, and LC-MC techniques).

Small molecule drugs

[0155] Small molecule drug therapeutics generally refer to therapeutics of low molecular weight (e.g., below 1 kDa) that modulate cellular behavior to treat a disease. Such small molecule drugs bind one or more biological targets of a target cell, thereby causing a change in the activity or function of the biological target of the target cell. Given their size, small molecule drug therapeutics are able to penetrate cellular membranes, thereby enabling them to bind or affect biological targets located within cells.

[0156] In various embodiments, small molecule drug therapeutics are inhibitors that serve to inhibit a biologic target that is involved in a disease. For example, small molecule drug therapeutics may be kinase inhibitors, proteasome inhibitors, proteinase inhibitors, or protein inhibitors. Additionally, small molecule drug therapeutics can be chemotherapeutics that prevent cell replication such as alkylating agents, anti-microtubule agents, topoisomerase inhibitors, DNA intercalators, and the like.

[0157] More comprehensive lists of small molecule drug therapeutics are found in publicly available databases such as DrugBank, ChemSpider, ChEMBL, KEGG, and PubChem. In some embodiments, the small molecule is an inhibitor of a protein or portion thereof encoded by the nucleic acid sequence set forth in SEQ ID NO: 1-349. In some embodiments, the small molecule is an inhibitor of a protein or portion thereof set forth in **FIG. 9D-9H** and **FIG. 11A-11C**, or encoded by the nucleic acid sequence or a portion thereof set forth in **TABLE 3** and **TABLE 4**.

Biologics

[0158] Biologics generally refer to therapeutics that are manufactured from biologic sources (e.g., produced in cells). Biologics are larger than small molecule drugs and often times more complex in structure and molecular makeup. In various embodiments, biologics are synthesized through manufacturing methods that include 1) inserting a DNA sequence encoding for the biologic or a portion of the biologic into a living cell, 2) having the cell produce transcribe/translate the DNA sequence into a protein, 3) isolating the protein from the cells, where the protein serves as the biologic or a component of the biologic. Example of biologics include antibodies (e.g., monoclonal or polyclonal antibodies), cytokines, growth factors, enzymes, immunomodulators, recombinant proteins, vaccines, allergenics, blood

components, hormones, therapeutic cells (e.g., stem cells), tissues, carbohydrates, and nucleic acids.

V. Kits

[0159] In some embodiments, any of the BiPS or viral sequences disclosed herein is assembled into a pharmaceutical or diagnostic or research kit to facilitate their use in therapeutic, diagnostic or research applications. A kit may include one or more containers housing any of the vectors, nucleic acids, proteins, peptides, or viruses disclosed herein and instructions for use.

[0160] The kit may be designed to facilitate use of the methods described herein by researchers and can take many forms. Each of the compositions of the kit, where applicable, may be provided in liquid form (*e.g.*, in solution), or in solid form, (*e.g.*, a dry powder). In certain cases, some of the compositions may be constitutable or otherwise processable (*e.g.*, to an active form), for example, by the addition of a suitable solvent or other species (for example, water or a cell culture medium), which may or may not be provided with the kit. As used herein, "instructions" can define a component of instruction and/or promotion, and typically involve written instructions on or associated with packaging of the disclosure. Instructions also can include any oral or electronic instructions provided in any manner such that a user will clearly recognize that the instructions are to be associated with the kit, for example, audiovisual (*e.g.*, videotape, DVD, etc.), Internet, and/or web-based communications, etc. The written instructions may be in a form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which instructions can also reflect approval by the agency of manufacture, use or sale for animal administration.

[0161] Throughout the description, where compositions are described as having, including, or comprising specific components, or where processes and methods are described as having, including, or comprising specific steps, it is contemplated that, additionally, there are compositions of the present disclosure that consist essentially of, or consist of, the recited components, and that there are processes and methods according to the present disclosure that consist essentially of, or consist of, the recited processing steps.

[0162] In the application, where an element or component is said to be included in and/or selected from a list of recited elements or components, it should be understood that the element or component can be any one of the recited elements or components, or the element or component can be selected from a group consisting of two or more of the recited elements or components.

[0163] Further, it should be understood that elements and/or features of a composition or a method described herein can be combined in a variety of ways without departing from the spirit and scope of the present disclosure, whether explicit or implicit herein. For example, where

reference is made to a particular compound, that compound can be used in various embodiments of compositions of the present disclosure and/or in methods of the present disclosure, unless otherwise understood from the context. In other words, within this application, embodiments have been described and depicted in a way that enables a clear and concise application to be written and drawn, but it is intended and will be appreciated that embodiments may be variously combined or separated without parting from the present teachings and disclosure. For example, it will be appreciated that all features described and depicted herein can be applicable to all aspects of the disclosure described and depicted herein.

[0164] It should be understood that the expression “at least one of” includes individually each of the recited objects after the expression and the various combinations of two or more of the recited objects unless otherwise understood from the context and use. The expression “and/or” in connection with three or more recited objects should be understood to have the same meaning unless otherwise understood from the context.

[0165] The use of the term “include,” “includes,” “including,” “have,” “has,” “having,” “contain,” “contains,” or “containing,” including grammatical equivalents thereof, should be understood generally as open-ended and non-limiting, for example, not excluding additional unrecited elements or steps, unless otherwise specifically stated or understood from the context.

[0166] Where the use of the term “about” is before a quantitative value, the present disclosure also includes the specific quantitative value itself, unless specifically stated otherwise. As used herein, the term “about” refers to a $\pm 10\%$ variation from the nominal value unless otherwise indicated or inferred.

[0167] It should be understood that the order of steps or order for performing certain actions is immaterial so long as the embodiments remain operable. Moreover, two or more steps or actions may be conducted simultaneously.

[0168] The use of any and all examples, or exemplary language herein, for example, “such as” or “including,” is intended merely to illustrate better the embodiments and does not pose a limitation on the scope of the invention unless claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the present invention.

EXAMPLES

[0169] The following Examples are merely illustrative and are not intended to limit the scope or content of the invention in any way.

EXAMPLE 1 Isolation of bat embryonic fibroblasts

[0170] This example describes the isolation of embryonic fibroblasts from bats. An embryo (approximately developmental stage 20) acquired from a Spanish *Rhinolophus ferrumequinum* bat (wild horseshoe bat) was cut into several pieces while removing the head and as much as the inner organ tissue as possible. The pieces were then flushed with PBS and processed separately. The tissue was covered with 0.05% trypsin, minced with a scalpel, and incubated in a cell culture incubator at 37°C and 5% CO₂ for 45 minutes. The trypsin was deactivated with fibroblast medium consisting of DMEM (Life Technologies, CA), 10% fetal bovine serum (Sigma, MO), 0.1 mM MEM Non-essential amino acids (Life Technologies, CA), 2 mM GlutaMax supplement (Life Technologies, CA), and Penicillin-Streptomycin (10 U/ml and 10 µg/ml, respectively; Life Technologies, CA). The cells were broken up by pipetting up and down 20 times, collected by centrifugation, transferred to a gelatin-coated (Sigma-Aldrich, MO) T75 cell culture treated flasks (Corning, AZ) in 15 ml of fibroblast medium, and cultured at 37°C and 5% CO₂. After 3 days, when reaching ~80% confluency, the attached cells were washed with DPBS (Life Technologies, CA), treated with 0.05% trypsin-EDTA, (Life Technologies, CA) to obtain a single cell solution and either split at a ratio of 1:4 or used directly in a reprogramming experiment.

EXAMPLE 2 Isolation of bat fibroblasts from tail biopsies

[0171] This example describes the isolation of fibroblasts from tail biopsies from adult bats.

[0172] *M. myotis* bats were sampled in Morbihan, Brittany in North-West France in accordance with the permits and ethical guidelines issued by ‘Arrêté’ by the Préfet du Morbihan and the University College Dublin ethics committee. This population has been transponded and followed since 2010 as part of on-going mark-recapture studies by Bretagne Vivante and the Teeling laboratory (Huang et al., 2019). Once captured, all bats were placed in individual cloth bags before processing. A single 3 mm biopsy was taken from the outstretched uropatagium of each bat using a sterile biopsy punch and immediately submerged in a Cryotube with 2ml of DMEM cell culture medium supplemented with 20%

FBS, 1% NEA, and 1% Antibiotic-Antimycotic containing Streptomycin, Amphotericin B and Penicillin, maintaining as sterile conditions as possible. All bats were offered food and water and rapidly released after processing. Biopsies were then stored at 4°C and transported to the laboratory for processing within 6 days. Samples were further processed through a cell extraction methodology similar to a previously established protocol (Kacprzyk et al., 2021) with a few modifications. The samples were rinsed with DPBS and cut finely within a minimal amount of cell culture medium using sterile blades to result in six 0.5 mm pieces. These pieces were then transferred aseptically to a cryotube containing cell culture medium and incubated for 18 hours with collagenase type II at 37°C with 5% CO₂ to allow for digestion. The pieces were collected by centrifugation for 5 minutes at 300 rcf, resuspended in 2 ml of fresh cell culture medium and transferred to a 35 mm cell culture treated plate for initial P1 expansion. Cells were then fed every 2-3 days with cell culture medium as above but a reduced 0.2% concentration of antibiotic-antimycotic. For the first feeding a $\frac{1}{3}$ media change was performed to avoid sudden changes in antibiotic-antimycotic concentration from 1% to 0.2%. When the cells reached 70% confluency, they were transferred to a T25 flask in cell culture medium after treatment with 0.05% Trypsin and were fed every 2-3 days as necessary. At 85% confluency, the cells were trypsinized as before and 1×10^6 cells were frozen in 1 ml cell culture medium containing 10% DMSO.

EXAMPLE 3 Reprogramming and expansion of bat embryonic and adult fibroblasts into bat iPSCs

[0173] This example describes the reprogramming of bat embryonic fibroblasts for the generation of bat iPSCs. First, the original Yamanaka reprogramming protocol (Takahashi et al., *Cell* (2006) 126, 663-676) based on four reprogramming factors (Oct4, Sox2, Klf4, and cMyc) was tried, because it provides the most direct way to generate pluripotent stem cells in most species. Strikingly, the standard protocol that is highly effective in mice, humans and other mammalian species (domestic dog, (*Canis familiaris*), domestic pig, (*Sus scrofa*), common marmoset (*Callithrix jacchus*)) failed in bats. Even though the standard reprogramming protocol failed, it provided the crucial insight that the Yamanaka factors triggered the formation of rudimentary stem cell-like colonies even though the reprogrammed cells ceased to expand. Thus, the core pluripotency network might be conserved in bats. However, the signaling cascades that usually shield this network from differentiation cues are different. An exemplary bat pluripotent stem cell derivation strategy is illustrated in **FIG. 1A**.

[0174] Briefly, 150,000 embryonic *Rhinolophus ferrumequinum* fibroblasts at passage 2, adult *Myotis myotis* at passage 3, or CF1 mouse embryonic fibroblasts at passage 3 were resuspended in 1 ml of fibroblast medium and mixed with Sendai-virus particles containing the reprogramming factors Oct4, Sox2, cMyc, and Klf4 (CytoTune iPS 2.0, Life Technologies, CA) with a final multiplicity of infection (MOI) of 10, 10, 10, and 15, respectively. The cells were plated on one gelatin-coated well of a 6-well plate and cultured at 37°C with 5% CO₂. The medium was replaced every 24 hours. 6 days after transduction, the cells of each well were collected by treatment with 0.05% trypsin-EDTA, seeded at a density of 50,000 cells per 60 cm² on irradiated CF1 mouse embryonic fibroblasts (MEFs; ThermoFisher, MA) in fibroblast medium. After 24 hours, the medium was switched to 50% fibroblast medium and 50% pluripotent stem cell (PSC) medium consisting of DMEM/F-12 (Life Technologies, CA), 20% knockout serum replacement, 0.1 mM MEM Non-essential amino acids, 2 mM GlutaMax supplement, Penicillin-Streptomycin (10 U/ml and 10 µg/ml, respectively), 100 µM 2-mercaptoethanol, and 40 ng/ml FGF2. From then on, the medium was replaced every day with PSC medium until day 14 when the FGF concentration was increased to 100 ng/ml and the medium was supplemented with 10⁻⁴ U/ml Leukemia inhibitory factor (Lif), 100 ng/ml SCF (R&D Systems, MN) and 20 nM Forskolin. Colonies appeared 14 to 16 days after transduction, were picked on day 20 and expanded on irradiated MEFs with Gentle Cell dissociation Reagent (StemCell Technologies, MA). After that, cells were passaged approximately every 5 days, or when they were confluent, at a ratio of 1:6 to 1:12 onto irradiated MEFs. Cell and colony morphology were recorded with an EVOS digital inverted microscope (Invitrogen, MA).

[0175] Thus, specific ratios of reprogramming factors, and the addition of Lif, Scf, the Pka activator forskolin and Fgf2 to the culture medium allowed for the uninterrupted growth of bat pluripotent stem cells. Under these conditions, bat stem cell colonies typically appeared after 14-16 days of culture. These initial stem cell colonies were, however, not readily pickable and expandable using conventional EDTA- (Versene), collagenase- or trypsin-based methods that are normally used to passage pluripotent stem cells from other species. To split cells for further passaging and growth cells were lightly flushed off the feeder cell layer after gentle treatment with low concentrations of EDTA. Exemplary cell morphology of the reprogrammed bat iPSCs is shown in **FIG. 1B** and **FIG. 2A**. Bat pluripotent stem cell colonies appeared tight and homogeneous. The cells had a large, apparent nucleus with one or two prominent nucleoli. Their proliferation rate was similar to

human pluripotent cells despite a somewhat lower clonogenicity. The iPSC reprogramming protocol was further validated by developing iPS cells from an evolutionary distant bat species *Myotis myotis* (greater mouse-eared bat) non-lethally sampled in the wild, which exhibited similar attributes to the greater horseshoe bat iPS cells, suggesting that this unique pluripotent state evolved in the ancestral bat lineage. The iPSC cells derived from the *M. myotis* tail cell show that these fibroblasts were also readily reprogrammable using the new ‘batified’ Yamanaka protocol and yielded similar bat iPSCs that were Oct4 positive in immunostaining and differentiated into all three germ layers (**FIG. 2I-J**), suggesting that the protocol is applicable across the deepest basal divergencies in bats.

EXAMPLE 4 Characterization of the reprogrammed cells

[0176] This example illustrates the characterization of the reprogrammed cells. After reprogramming, cells were analyzed for karyotype, chromatin organization, and gene and RNA expression.

Karyotyping

[0177] This example illustrates the karyotyping of reprogrammed cells. Briefly, cells were treated with 100 ng/ml KaryMax Colcemid Solution in HBSS (Life Technologies, CA) for 16 hours, then treated with 0.05% trypsin-EDTA for 15 minutes and filtered through a 40 μ m cell strainer to remove clumps. Cells were collected by centrifugation, resuspended in 1 ml 0.075 M potassium chloride (Sigma-Aldrich, MO) and incubated for 20 minutes at room temperature. 0.5 ml fixative (1 part glacial acetic (Fisher Scientific, MA) mixed with 3 parts methanol (Sigma-Aldrich, MO) were added, cells were collected as before, resuspended in 4 ml fixative, and incubated for 20 minutes at room temperature. The fixation step was repeated, the cells collected as before and all but about 200 μ l of the fixative was removed. The cells were resuspended in the remaining fixative and dropped onto slides that were precooled at -20°C. The slides were airdried and the cells stained for 10 minutes with Giemsa Staining solution consisting of 1 part KaryoMax Giemsa solution (Life Technologies, CA) and 3 parts Gurr buffer (Invitrogen, MA). The slides were washed with water, dried, and mounted in Cytoseal 60 (Thermo Scientific, MA). High-resolution pictures of chromosome spreads were acquired with an AxioObserver microscope (Zeiss) using the 100x oil objective. Even after prolonged culture (over 50 passages), the cells retained a normal karyotype, with most cells containing 56 chromosomes (**FIG. 2B**).

RT-PCR

[0178] mRNA was extracted with the RNeasy Mini Kit (Qiagen). 500 ng of each sample were used to generate cDNA by reverse transcription using the SuperScript™ IV VILO™ Master Mix (Invitrogen). 2 µl of the cDNA were used to detect the presence of Sendai virus transcripts using GoTaq Green Polymerase (Promega), and the oligos as recommended in the CytoTune iPS 2.0 kit (Invitrogen). Gapdh was amplified as loading control using oligos with the following sequence: Z25-132:GAPDH_F1_GHB: TGGTGAAGGTCGGAGTGAAC and Z25-133:GAPDH_R1_GHB: GAAGGGGTCATTGATGGCGA). The PCR products were analyzed on a 2% agarose gel containing ethidium bromide.

Immunofluorescence staining

[0179] For immunofluorescence staining, cells were plated on µ-slides (Ibidi, Germany). After 4 days, cells were washed once with DPBS and fixed with Cytotfix/Cytoperm solution (Becton Dickinson, NJ) for 20 minutes at 4°C. Cells were rinsed with Perm/Wash buffer (Becton Dickinson, NJ) and then incubated overnight at 4°C in Perm/Wash buffer containing primary anti-Afp (R&D Systems, MN) anti-Pax6 (BioLegend, CA), J2 anti-dsRNA (Scicons, Hungary), anti-(gag/pol) HERVK (Austrial Biological) or FIPV3-70 anti-Pan Corona (Life Technologies, CA) or directly conjugated anti-Oct3/4-AF488 (Santa Cruz, CA) or anti-Brachyury (R&D Systems, MN) anti-Otx2 (R&D Systems), anti-Zic2 (Abcam), anti-Tfe3 (Sigma Aldrich) or anti-Tfcp2l1 (R&D Systems) in a 1:50 (anti-Oct3/4) or 1:100 dilution (all others). Cells were rinsed and washed 3 times for 2 minutes with Perm/Wash solution at room temperature followed by a 1-hour incubation with a 1:200 dilution of the corresponding secondary antibodies (Donkey anti-chicken-Cy3, Millipore, AP194C; Goat anti-chicken-AF488 ; Donkey anti-rabbit-AF647; Goat anti-rabbit-AF488, Goat anti-mouse-AF488) in Perm/Wash buffer. Cells were rinsed, washed twice for 2 minutes with Perm/Wash Buffer and then incubated for 5 minutes with Perm/Wash buffer containing 2 drops per ml NucBlue Dapi stain (Invitrogen, MA). The buffer was removed, and the cells were cover-slipped in Prolong Dimond antifade mounting medium (Invitrogen, MA). Images were acquired with an AxioObserver fluorescence microscope with Apotome (Zeiss). For the simulated emission depletion (STED) microscopy (super-resolution), the cells were plated on coverslips that were placed in wells of 6-well plates. The staining was performed as described above but with a 1:200 dilution of the Abberior Star 635P secondary antibody in Perm/Wash buffer. Cells were rinsed, washed twice for 2 minutes with Perm/Wash Buffer and then incubated for

5 minutes with Perm/Wash buffer containing 2 drops per ml DyeCycle Violet stain. The coverslips were mounted face down on glass slides with Prolong Dimond antifade mounting medium (Invitrogen). Images were acquired with a TCS SP8 confocal microscope with STED 3x and White Light Laser (Leica) with a 100x oil objective. 405 nm and 594 nm lasers were used for excitation and 775 nm laser for depletion. Image resolution obtained was 19.8 μm by 19.8 μm using a zoom factor of 6x.. Exemplary immunofluorescent detection of Oct4/Pou5f2 in BiPS cells shows that the cells were positive for the pluripotency factor Oct4 (FIG. 1C).

RNA isolation and RNA-seq

[0180] For RNA-seq, RNA was extracted from BiPS cells at passage 22 and BEFs at passage 3. RNA was extracted with the RNeasy RNA isolation kit (Qiagen, Germany) following the manufacturer's recommendations including the DNase digest (Qiagen, Germany) and eluted in 50 μl RNase/DNase free H_2O . The libraries were prepared with the SMART-Seq v4 Ultra Low Input kit (Takara Bio, undifferentiated cells) or the Stranded Total RNA with Ribo-Zero Plus kit (Illumina, differentiated cells) and 100 bp paired-end sequencing reads were (PE100) were generated by Illumina sequencing (NovaSeq 6000 S1) to a depth of 50 million reads (100 million total reads).

RNA-seq mapping and visualization

[0181] The quality of the reads from the RNA sequencing was analysed with FastQC v0.11.9 (Andrews, 2010), and visualized using MultiQC (Ewels et al., 2016). With the mean phred score of around Q35 across each base position no filter or processing was performed. To carry out the differential expression analysis, the genome of *Rhinolophus ferrumequinum* was used as reference genome, RefSeq assembly accession GCF_004115265.1, assembled and annotated by the Vertebrate Genomes Project (www.vertebrategenomesproject.org). The reads were mapped with HISAT2 v2.2.1 (Kim et al., 2019), the .sam files resulting from each mapping were converted into .bam files and indexed using samtools v1.10 (Li et al., 2009). The reads were mapped against each gene using featureCounts v2.0.1 (Liao et al., 2014) and the differential expression analysis was performed with DESeq2 v1.10.1 (Love et al., 2014). To visualize the RNA-seq data in the UCSC genome browser, bigwig files were generated using the bamCoverage command from deepTools (www.deeptools.readthedocs.io/en/develop/content/tools/bamCoverage.html; Ramirez et al., 2016).

MA Plot

The MA plots were generated based on the DESeq2 (see above) results with the ggmaplot function (www.rpkgs.datanovia.com/ggpubr/reference/ggmaplot.html) from the R package ggpubr (www.rpkgs.datanovia.com/ggpubr/). Genes are indicated by dots, plotted by their log₂ fold change between bat fibroblast and pluripotent stem cells and the log₂ mean of normalized counts (ratio of means). Blue dots indicate genes with an adjusted p value (or FDR) of <0.05 and a fold change of 2 (log₂ fold change of 1), red dots indicate genes with an adjusted p value (or FDR) of <0.05 and fold change of -2 (log₂ fold change of -1). Dotted lines are drawn at fold change of 2/-2 (log₂ fold change of 1/-1).

[0182] RNA-seq analyses revealed the induced expression of canonical pluripotency-associated genes (**FIG. 1D**).

[0183] However, closer data inspection revealed that the expression profile did not necessarily match any known pluripotency state. Instead, factors indicative of the so-called naïve pluripotent state (Klf4, Klf17, Essrb, Tfcp2l1, Tfe3, Dppa, and Dusp6) were expressed alongside genes typically found in the more advanced *primed* pluripotent cells (*e.g.*, Otx2, Zic2). Double immunostainings detecting four of the most commonly used primed/naïve factors, Otx2/Tfe3 and Tfcp2l1/Zic2, respectively, showed co-expression of naïve and primed markers in most cells (**FIGs. 2K-M**). No methylation in the promoters of Nanog, Pou5f1, or Sox2 was detected, which might be related to under-annotation of the *Rhinolophus ferrumequinum* genome at this point in time. Germ cell factors such as Dnmt3l and Dazl were absent. Thus, while cellular heterogeneity might be at play, their uniform appearance makes it most likely that bat stem cells occupy a novel, yet-to-be-characterized pluripotent default state.

ATAC-seq

[0184] To analyze the effects of the reprogramming approach on the bat chromatin and epigenetic structures a global epigenetic landscape survey using ATAC-seq was performed. ATAC-seq and bioinformatics analysis to detect open chromatin in bat fibroblasts and bat pluripotent stem cells was performed by Active Motif, CA from 100,000 cryopreserved cells (ATAC-seq service). In brief, nuclei were isolated and libraries of open chromatin were prepared with the Nextera Library Prep Kit (Illumina) by Tn5 tagmentation. The tagmented DNA was purified using the MinElute PCR purification kit (Qiagen, Germany), amplified with 10 cycles of PCR, and purified using Agencourt AMPure SPRI beads (Beckman

Coulter, CA). 42 bp paired-end sequencing reads (PE42) were generated by Illumina sequencing (using NextSeq 500) to a depth of at least 83 million total reads and mapped to the GCA_004115265.2 genome (Ensembl, annotation version 102) using the BWA algorithm with default settings (“bwa mem”). Alignment information for each read was stored as BAM file. Only reads that passed the Illumina’s purity filter, aligned with no more than 2 mismatches, and mapped uniquely to the genome were used in the subsequent analysis. Duplicate reads (“PCR duplicates”) were removed. Genomic regions with high levels of transposition/tagging events were then determined using the MACS2 peak calling algorithm (Zhang et al., *Genome Biology* (2008) 9:R137). To identify the density of transposition events along the genome, the genome was divided into 32 bp bins and the number of fragments in each bin was determined. The data were then normalized by reducing the tag number of all samples by random sampling to the number of tags present in the smallest sample. Peak metrics between samples were compared by grouping overlapping Intervals into “Merged Regions,” which are defined by the start coordinate of the most upstream Interval and the end coordinate of the most downstream Interval (= union of overlapping Intervals; “merged peaks”). In locations where only one sample has an Interval, this Interval defines the Merged Region. Intervals and Merged Regions, their genomic locations along with their proximities to gene annotations and other genomic features were determined and average and peak (*i.e.* at “summit”) fragment densities were compiled. The sequencing tracks (number of fragments in each 32 bp bin stored as .bigwig file) were visualized with the UCSC genome browser.

[0185] The global epigenetic landscape survey using ATAC-seq revealed significant chromatin configuration changes when bat fibroblasts transitioned into the pluripotent state (**FIG. 1E**). Generally, there were strict correlations between newly opened sites and gene expression and conversely closed regions and gene shutdowns (**FIG. 1F**). Similarly, mapping the DNA methylome by RRBS-seq exposed significant CpG methylation changes across the genome after reprogramming (**FIG. 2G-H and**).

Reduced representation bisulfite sequencing (RRBS) of bat iPSCs

[0186] Reduced representation bisulfite sequencing of bat fibroblasts and pluripotent stem cells was performed by Active Motif, CA (RRBS Service, Active Motif, CA). Briefly, 500,000 cells were provided as a frozen pellet. Genomic DNA was isolated, and 100 ng were digested with TaqAI (NEB, MA) at 65°C for 2 hours followed by MspI (NEB, MA) at 37°C

overnight. Following enzymatic digestion, samples were used for library generation with the Ovation RRBS Methyl-Seq System (Tecan, Switzerland) following the manufacturer's instructions. In brief, digested DNA was randomly ligated, and, following fragment end repair, bisulfite converted using the EpiTect Fast DNA Bisulfite Kit (Qiagen, Germany) following the Qiagen protocol. After conversion and clean-up, samples were amplified resuming the Ovation RRBS Methyl-Seq System protocol for library amplification and purification. 75 bp single-end sequencing reads (SE75) were generated by Illumina sequencing (using NextSeq 500) to a depth of at least 27 million reads (total of 54 million reads), with at least 2.9 million covered CpGs. The reads were mapped to the GCA_004115265.2 genome (Ensembl, annotation version 102) and the percentage of methylation at CpG sites across the genome was calculated. To visualize the methylation ratios aligned to the genome with the UCSC genome browser, the methylation ratio files containing the methylation ratio for each chromosomal position were first converted to bed files, that were then used to generate bigwig files with the bedGraphToBigWig v4 tool (www.encodeproject.org/software/bedgraph2bigwig/). Correlation scatter plots were generated to show the level of methylation at common CpG sites. To visualize the global differences between bat fibroblast and pluripotent stem cells, the RRBS methylation data were combined for all samples based on chromosome position, the ratios of the duplicates were averaged and the methylation ratio for each chromosomal position was plotted using the ggplot2 function "stat_density_2d_filled" with fill based on density. Only chromosomal positions that were present in all replicates were included in the analysis.

[0187] Similarly, mapping the DNA methylome by RRBS exposed significant CpG methylation changes across the genome (**FIG. 1A and 2G**) after reprogramming.

Chromatin immunoprecipitation sequencing (ChIP-seq)

[0188] 5 million cells were fixed cells in 1% formaldehyde by adding 1/10 volume of freshly prepared Formaldehyde Solution (11% formaldehyde, 0.1 M NaCl, 1 mM EDTA, pH 8.0, 50 mM HEPES, pH 7.9) to the existing medium. Cells were agitated for 15 minutes at room temperature and the fixation was stopped by addition of 1/20 volume of 2.5 M glycine solution (final concentration of 0.125 M) to the existing medium and incubation at room temperature for 5 minutes. The cells were scraped off the wells, collected by centrifugation at 800 g and washed with 10 ml chilled 0.5 % Igepal in PBS per tube by pipetting up and down. Cells were pelleted by centrifugation as before and resuspended in 10 ml chilled PBS-Igepal containing 1 mM PMSF. Cells were collected as before, and the cell pellet was snap-frozen in

liquid nitrogen. Further processing, chromatin immunoprecipitation and bioinformatics analysis to detect H3K4me3 and H3K27me3 was performed by Active Motif, CA (HistoPath ChIP-seq service). In brief, chromatin was isolated by adding lysis buffer, followed by disruption with a Dounce homogenizer. Lysates were sonicated and the DNA sheared to an average length of 300-500 bp with Active Motif's EpiShear probe sonicator. Genomic DNA (Input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by SPRI beads clean up (Beckman Coulter, CA) and quantitation with Clariostar (BMG Labtech). An aliquot of chromatin (20 µg) was precleared with protein A agarose beads (Life Technologies, CA). Genomic DNA regions of interest were isolated using 4 µg of antibody against H3K4me3 (Active Motif, CA) or H3K27me3 (Active Motif, CA). Complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation overnight at 65°C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation. Illumina sequencing libraries were generated from the ChIP and Input DNAs with the standard consecutive enzymatic steps of end-polishing, dA-addition, and adaptor ligation. After a final PCR amplification step, 75-nt single-end (SE75) sequence reads were generated by Illumina sequencing (using NextSeq 500) to a depth of at least 36 million reads per sample and mapped to the GCA_004115265.2 genome (Ensembl, annotation version 102) using the BWA algorithm with default settings. Duplicate reads were removed, and only uniquely mapped reads (mapping quality ≥ 25) were used for further analysis. Alignments were extended in silico at their 3'-ends to a length of 200 bp, which is the average genomic fragment length in the size-selected library and assigned to 32-nt bins along the genome. The resulting histograms (genomic "signal maps") were stored in bigWig files. To find peaks, the generic term "Interval" was used to describe genomic regions with local enrichments in tag numbers. Intervals were defined by the chromosome number and a start and end coordinate. Peak locations were determined using the MACS algorithm (v2.1.0) with a cutoff of p-value = $1e-7$ (Zhang et al., 2008). Signal maps and peak locations were used as input data to Active Motifs proprietary analysis program, which creates Excel tables containing detailed information on sample comparison, peak metrics, peak locations and gene annotations. No normalization was performed on the H3K27me3 data, while standard normalization was applied to the H3K4me3 data. The tag number of all samples (within a comparison group) was reduced by random sampling to the number of tags present in the smallest sample. To compare peak metrics between 2 or more samples, overlapping Intervals were grouped into "Merged Regions," which are defined by the start coordinate of the most upstream Interval

and the end coordinate of the most downstream Interval (= union of overlapping Intervals; “merged peaks”). In locations where only one sample has an Interval, this Interval defines the Merged Region. The sequencing tracks (number of fragments in each 32 bp bin stored as bigwig file) were visualized with the UCSC genome browser.

[0189] ChIP-seq analysis showed that histone marks associated with active (H3K4me3) and developmentally repressed genes (H3K27me3) showed many changes (**FIG. 1G**, Approximately 18.2% of the bat stem cell genes were associated with a “bi-valent” domain (H3K4me3 and H3K27me3; **FIG. 1H**), a pluripotency chromatin hallmark initially found in human and mouse pluripotent cells. Interestingly, while there was overlap between human and bat bivalency genes there were also some bat- or human-specific genes (**FIG. 2E**). Generally, there were strict correlations between newly opened sites and gene expression, and conversely, closed regions and gene shutdowns during the reprogramming process that also corresponded to the absence or presence of histone modifications, respectively (**FIG. 1I**). However, there are instances when there were simultaneously active and repressive epigenetic marks, most likely as a result of spontaneous differentiation in the cultures (**FIG. 2F**).

[0190] Collectively, the results establish that the bat pluripotent stem cells are reprogrammed both transcriptionally and epigenetically.

EXAMPLE 5 Three germ layer differentiation

[0191] This example illustrates the further functional characterization of the reprogrammed bat IPS cells. After reprogramming, cells were analyzed in pluripotency assays for pluripotency potential.

[0192] The differentiation of bat pluripotent stem cells was carried out with the STEMdiff Trilineage differentiation kit (StemCell Technologies, MA) following the manufacturer’s protocol. Cells were plated at the desired densities in mTeSR medium (StemCell Technologies, MA), and plated on Vitronectin-coated (StemCell Technologies, MA) cell culture plates. After 5 days (endoderm or mesoderm) or 7 days (ectoderm) in culture as directed by the manufacturer. For the ectoderm differentiation, the floating three-dimensional structures were then replated and grown for 4 additional days in fibroblast medium. The cells were stained with antibodies detecting the appropriate lineage markers as described above or cells were collected (surface area of 10 cm² per replicate) for RNA isolation and RNAseq after addition of 600 µl lysis buffer RTL (part of the RNeasy kit; Qiagen, Germany).

[0193] Results show that the bat iPSCs differentiate into ectodermal, mesodermal, and endodermal fates (**FIG. 4A**). In each case, the cells responded to the altered culture conditions by shifting their morphology profoundly. The differentiated iPSCs turned positive for Pax6 (ectoderm), T (mesoderm) or AFP (endoderm). Since the cells used in this experiment were at an advanced passage (passage 37, an equivalent of about 6 months of continuous culture), the results also suggest that pluripotency can be maintained long-term.

Embryonic body differentiation

[0194] To analyze the bat stem cells' developmental plasticity, the cells were subjected to embryoid body (EB) differentiation. Briefly, bat pluripotent stem cells grown on irradiated mouse embryonic fibroblasts from a total area of 60 cm² were washed with PBS, treated for 10 minutes with Gentle Cell Dissociation Reagent (StemCell Technologies, MA), collected by centrifugation and resuspended in 12 ml differentiation medium consisting of DMEM/F-12 (Life Technologies, CA), 10% fetal bovine serum (Sigma, MO), 0.1 mM MEM Non-essential amino acids (Life Technologies, CA), 2 mM GlutaMax supplement (Life Technologies, CA), Penicillin-Streptomycin (10 U/ml and 10 µg/ml, respectively; Life Technologies, 15140122) and 100 µM 2-mercaptoethanol (Fluka, NC). The cells were then transferred to one uncoated 60 cm² petri dish (Corning, 351029). After 3 days in culture, as much as possible of the medium (about 2/3) was carefully exchanged without disturbing and removing the floating EBs that had formed. The floating EBs were collected after 3 more days (total of 6 days) in culture, fixed in Cytofix/Cytoperm fixation buffer (Becton Dickinson, NJ) overnight, and then stained with antibodies against as described above to detect differentiation markers of all three germ-layers by immunofluorescence. For RNA isolation and RNA-seq, EBs were formed as described, collected, resuspended in 6 ml differentiation medium, and distributed into three wells of cell-culture treated 6-well plates (10 cm² each). After 2 more days in culture, the cells were washed with PBS, lysed with 600 µl buffer RTL (part of the RNeasy kit; Qiagen, 74104) and RNA was isolated as described above.

[0195] In the assay, cells differentiated and formed the for EBs' typical spherical arrangements. They subsequently matured into elaborate three-dimensional structures that were positive for all three germ layer markers (**FIG. 4B**). EBs were also analyzed by RNA-seq as described in **Example 4**. The RNA-seq analysis of RNA isolated from the monolayer differentiation and EB formation confirmed the respective cell fate changes (**FIG. 4C, FIG. 5A-D**).

Teratoma formation

[0196] To assay the potential of the bat iPSCs to form teratomas in vivo, cells were injected into immunocompromised mice and then analyzed. Briefly, two 6-well plates (12 wells) of bat pluripotent stem cells grown on irradiated mouse embryonic fibroblasts were scraped off in 2 ml DMEM/F-12 medium (Life Technologies, CA), collected by centrifugation and resuspended in 500 µl DMEM/F-12 medium. 100 µl of the cell suspension were injected into the hindleg muscle of 8-week-old male Fox Chase SCID Beige Mice (Charles River, MA). Tumor tissue that had formed after 16 weeks was harvested, fixed in 10% Formalin (Fisher Scientific, MA) overnight and then transferred to 70% ethanol. The tissue was embedded in paraffin and hematoxylin and stained with eosin of 5 µm sections. Images were acquired with an AxioObserver microscope (Zeiss) and analyzed.

[0197] The analysis showed, that the bat iPSCs formed a particular tumor (teratoma) at the injection site after four to five months albeit infrequently (33%) and very small (2-4 mm). The tumors were comprised of immature tissue with epithelial, neural and stromal characteristics (**FIG. 4D**). Transcriptional profiling of pivotal genes previously reported critical for teratoma formation (**FIG. 4G**) revealed that while some genes are downregulated in bat iPSCs in comparison with mouse iPSCs (like Eras), other genes like the hyaluronidases (HAS) and ADP ribosylation factors (ARFs) are indistinguishable between the experimental groups, making it likely that the anti-tumor effect seen in the rudimentary teratomas is a complex phenomenon. While the host mice were severely immunocompromised and immune-related tissues were not analyzed the immaturity and delay in growth may suggest a yet to be characterized anti-tumorigenic property of bat stem cells similar to, for instance, the naked mole rat which could also underlie the extended healthspans and cancer resistance reported in bats.

Blastoid differentiation

[0198] To analyze the potential of the iPSCs to form embryoid structures, the cells were subjected to a modified blastoid protocol. Cells were harvested and plated as described for the embryonic body formation above. After 3 days in culture, 100 ng/ml BMP4 (R&D Systems, 314-BP-010) were added to the medium. 24 later the supernatant was diluted with 2/3 of fresh medium and transferred to two fresh uncoated petri dishes. The medium was exchanged after 3 more days in culture and floating blastoids were harvested 4 days later (total of 12 days of differentiation). The blastoids were fixed in Cytofix/Cytoperm fixation

buffer (Becton Dickinson, BDB554714) overnight, and stained as described above to detect the expression of Oct4 by immunofluorescence microscopy.

[0199] Further analysis showed, that bat blastoids recapitulate critical aspects of preimplantation embryos, including an Oct4-positive inner cell mass, the cystic cavity and a bilayered epithelium consisting of trophoblastic and yolk sac cells (**FIG. 3E**). Replating these embryo structures resulted in their attachment to a flattened trophoblastic epithelium to grow and an expansion of the inner cell mass (**FIG. 3F**). These differentiation studies exemplify the unique potential of pluripotent bat cells to recapitulate important developmental events and serve as a powerful model to study the unique physiological adaptations of bats, including their reduced cancer phenotype.

[0200] Embryonic stem cell lines were derived from these outgrowths, confirming these embryoids' blastocyst nature.

[0201] The differentiation studies exemplify the unique potential of the described pluripotent bat cells to recapitulate important developmental events and serve as a powerful model to study the unique physiological adaptations of bats.

Example 6 Analysis of the distinct characteristics of pluripotent bat stem cells

[0202] To assay distinct characteristics of pluripotent bat stem cells, gene expression patterns in bat stem cells were analyzed such as the ground state transcriptome and then compared to other species. Transcriptome profiles of pluripotent stem cells from an assorted set of species (Bats, mouse, pig, dog, marmoset, human) and different cell types (EF, iPSCs, MEF, ESC) were assembled and principal component analysis was performed to obtain a high-level overview of the number of commonalities and differences between bats and other mammals (**FIG. 5A**)

Principal Component Analysis (PCA)

[0203] The DESeq2 output files of the RNA-seq analyses described above were subjected to a Variance Stabilizing Transformation (VST) using within-group-variability (Anders and Huber, 2010) to compare the bat pluripotent stem cell transcriptional profile with that of other species. The first two principal components of this result were plotted using the ggscatter function (<https://rpkgs.datanovia.com/ggpubr/reference/ggscatter.html>) from the R package ggpubr (www.cran.r-project.org/web/packages/ggpubr/index.html). The datasets used in the PCA were: GSM4616525, GSM4616526 and GSM4616527 (dog iPS), GSM4617887, GSM4617889, GSM4617890, GSM4617891, GSM4617895, GSM4617900 and

GSM4617901 (marmoset iPS), GSM4616532 (human iPS), GSM4616535 and GSM4616536 (pig iPS) from study GSE152493 (Yoshimatsu et al., 2021), and GSM1287734, GSM1287745 and GSM1287746 (mouse ESC) and GSM1287736, GSM1287747 and GSM1287748 (mouse iPS) from GSE53212 (Carter et al., 2014), as well as GSM2718393 and GSM2718399 (mouse iPS) from GSE101905 (Knaupp et al., 2017).

[0204] PCA showed that bats were unique to all mammals, even the more distant ones like dogs, clustered together in the PCA plot, while bats formed a separate distinctive group (**FIG. 5A**) despite including other closely related laurasiatherian mammals. Further analysis of the gene signature that contributed the most to the bat-specific gene expression profile in the PCA analysis was performed. The “leading edge,” was extracted, corresponding to the top 5% of the genes that fortified the difference in principal component 1 (**FIG. 5B**) when comparing bat with mouse pluripotent stem cells, corresponding to 674 genes. The list covered genes belonging to a broad spectrum of transcription factors, kinases, metabolic and homeostatic enzymes. For instance, it included the HMG-CoA synthase HMGCS2, the apolipoprotein APOA1, the cyclin CCNT1, plasminogen PLG, the pluripotency factors OCT4 and Nanog, Tmprss2 which is required for SARS-CoV-2 entry in humans and the ubiquitin ligase NEDD4 among many other categories. Given the broad spectrum of categories it was analyzed if the leading-edge genes were enriched for any particular biological pathway in gene ontology analyses. The leading-edge genes were further enriched for developmental controllers, proteins targeting membranes, including the endoplasmatic reticulum, lipid and cholesterol biosynthesis, and fibrinogen production. However, the most prominent groups were viral gene expression, viral transcription, and many sets of genes activated or suppressed after viral infection (**FIG. 5C**).

[0205] When analyzing the enrichment of any KEGG pathway, by far the most significantly enriched category was “Corona virus disease” (**FIG. 5D, FIG. 6A**). It almost seemed like bat stem cells executed a program normally activated after a virus infection. Interestingly, out of the set of leading-edge genes, only a total of eight genes showed significant evidence of positive selection in *R. ferrumequinum*, five of which showed at least one highly probable BEB site with no visual issues in the alignment region, while three genes (designated with ‘*’) did not (*AARD*, *COL3A1*, *FAM111A*, *LAMB3*, *MUC1**, *NES**, *RGS5*, *RSPH1**) (**FIG. 6B**). Two of these genes, *COL3A1* and *MUC1*, have roles in collagen formation in connective tissues; protect against pathogen infections and showed evidence of selection in another bat species suggesting unique, bat-specific adaptations in these genes.

The results might indicate that the unique bat signature is likely the consequence of the present viral sequences and that most of the coding leading edge genes are not under positive selection pressure.

[0206] Further, data were analyzed for the enrichment of transcription factor footprints in the mapping of open chromatin regions to these genes in the ATAC-seq data. Surprisingly, only two transcription factor motifs were significantly enriched, Klf5 and Ctcf. Notably, however, these factors accompanied the majority of the genes in this set. Klf5 is a canonical pluripotency factor, which is essential for early embryogenesis and self-renewal of pluripotent stem cells. The recruitment of Klf5 binding sites to a new set of genes makes it likely that bat stem cells acquired novel features under the influence of this transcription factor. Ctcf, on the other hand, contributes to the establishment of higher-order genome structures (topologically associating domains), which are evolutionarily stable.

[0207] The leading-edge genes showed that they were under a purifying and positive selection. Of the 655 orthologous genes analyzed, a significant intensifying, purifying selection was observed in only five (Rsph1, Nes, Col3a1, Rgs5, and Lamb).

MEME-ChIP

[0208] First, the ATAC-seq regions were identified that showed a $\log_2$ fold change of 5 between bat fibroblast and pluripotent stem cells and an adjusted p value of less than 0.1 that were within 10 kb (i.e., any interval within 10 kb upstream or downstream) of any gene that is part of the top 5 % of genes contributing to the differences in PC1 in the PCA analysis described above. The DNA sequences corresponding to these ATAC-seq regions were extracted from the GCF_004115265.1 reference genome and used in a MEME-ChIP motif search to identify sequence motifs (6-15 bp in width) for protein binding sites that are enriched in this set of genes (Machanic and Bailey, 2011; www.meme-suite.org/meme/tools/meme-chip). The sequence motifs with a p-value below 0.05 were then used in a FIMO analysis to identify the genomic positions and gene association of these motifs within the gene set. The number of genes associated with each motif within the gene set was then plotted against the factor known to bind to the and labeled with the protein known to bind to the motif

Evolutionary selection analysis

[0209] To explore evidence of positive selection in *R. ferrumequinum* for the 674 genes identified as part of the “leading” edge in the PCA analysis described above, all gene alignments were extracted that were available for these transcripts (n = 491) and had previously been annotated (Jebb *et al.*, 2020), in addition to annotating 169 alignments that had been made available as part of BAT1K but were currently unannotated. These alignments contained a maximum of 48 species from all eutherian mammalian superorders, with the species tree published by Jebb *et al.* (2020) used for all selection analyses. A total of 660 of these alignments contained representative genes for *R. ferrumequinum* and were analysed for positive selection using the branch-site models in the codeml package of the PAML suite of software (Yang, 2007). Positive selection was inferred using likelihood-derived dN/dS (ω) values under both a null (foreground and background ω constrained to be less than 1) and alternative (foreground ω can vary) model. The *R. ferrumequinum* lineage was designated as foreground branch to detect unique instances of taxon-specific positive selection. A likelihood ratio test (LRT, $2 * \ln L_{\text{alt}} - \ln L_{\text{null}}$) was used to compare the fit of both models, with a p-value calculated assuming chi-squared distributed LRTs. P-values were corrected for multiple testing using the Benjamin-Hochberg False Discovery Rate (FDR) method via ‘*padjust*’ implemented in R. Any significant gene showing a p-value greater than 0.05 with $\omega > 1$ was explored further. Significant sites showing positive selection were identified using Bayes Empirical Bayes (BEB) scores with a probability > 0.95 . All significant genes were subject to a visual inspection of the alignment, to rule out potential false positive results having occurred due to misaligned sequences. In addition to *R. ferrumequinum*, the *Myotis myotis* (n=637 representative genes), *Homo sapiens* (n=652), *Mus musculus* (n=628), *Canis lupus* (n=593) and *Felis catus* (n=603) lineages were also independently designated as foreground branches for all genes containing a representative sequence shared with *R. ferrumequinum*. This served as a means of determining whether positive selection identified in *R. ferrumequinum* was truly unique to the species lineage or a consequence of bat-specific, Laurasiatherian-specific, or eutherian mammal-specific instances of sequence evolution.

Gene ontology and KEGG pathway analyses

[0210] Gene ontology and KEGG pathways that are enriched within a group of genes were identified with the Enrichr tool (Xie *et al.*, 2021; www.maayanlab.cloud/Enrichr/). The odd ratios were then plotted with ggplot2 (Wickham, 2016; www.cran.r-project.org/web/packages/ggplot2/index.html) with the odds ratio displayed on the x-axis, the

dot size reflecting the gene count (number of genes present in the top 5% of PC1 contributing genes) and the dot color reflecting the p-value.

Protein Interaction Network in bat iPSCs

[0211] In order to understand if the leading-edge genes that make horseshoe bats unique were enriched for any particular functional gene ontology category (**FIG. 5C - D**). The genes of the Corona virus disease related KEGG pathway were retrieved from the PathCards database (www.pathcards.genecards.org).

[0212] The differential expression analysis was performed between bat (this study) and mouse iPS cells (GEO accession number: GSM1287736, GSM1287747 and GSM1287748 from Study GSE53212 (Carter et al., 2014) using DESeq2 (Love et al., 2014). The Corona virus disease-related genes were then illustrated with Cytoscape (Version 3.8.2, Shannon et al., 2003) using the STRING protein query with a 0.8 confidence score cutoff. The nodes were colored based on the log2FoldChange with a negative (blue) fold change indicating down-regulation and a positive (red) fold change indicating upregulation in bat pluripotent stem cells. Bold borders indicate proteins that were present in the top 5% of PC1 in the PCA analysis described above.

[0213] When analyzing the enrichment of any KEGG pathway, by far the most significantly enriched category was “Corona virus disease” (**FIG. 5D, FIG. 6A**). It almost seemed like bat stem cells executed a program normally activated after a virus infection. Interestingly, out of the set of leading-edge genes, only a total of eight genes showed significant evidence of positive selection in *R. ferrumequinum*, five of which showed at least one highly probable BEB site with no visual issues in the alignment region, while three genes (designated with ‘*’) did not (*AARD*, *COL3A1*, *FAM111A*, *LAMB3*, *MUC1**, *NES**, *RGS5*, *RSPH1**) (**FIG. 6B**). Two of these genes, *COL3A1* and *MUC1*, have roles in collagen formation in connective tissues; protect against pathogen infections and showed evidence of selection in another bat species suggesting unique, bat-specific adaptations in these genes. The results might indicate that the unique bat signature is likely the consequence of the present viral sequences and that most of the coding leading edge genes are not under positive selection pressure.

EXAMPLE 7 Identification of virus like structures in bat iPSCs

[0214] This example describes the identification of virus like structures in bat iPSCs.

[0215] Briefly, bat iPSCs were imaged with differential interference contrast microscopy and Image-based flow cytometry. Images of the bat iPSCs highlighted prominent cytoplasmic vesicles. Bat stem cells were observed to be packed with small, luminescent vesicles that filled a significant proportion of the cytoplasm (**FIG. 7A, FIG. 8A**).

Electron microscopy and Immunostaining

[0216] In order to analyze the vesicles, ultrastructural studies were performed using electron microscopy. Cells were grown in chambered Permanox slides (LabTek, MI) on irradiated mouse embryonic fibroblasts as described above for 5 days and then further processed by the Biorepository and Pathology core at the Icahn School of Medicine at Mount Sinai. Briefly, the cells were rinsed once with DPBS and fixed overnight with 2% paraformaldehyde and 2% glutaraldehyde in 0.01 M sodium cacodylate buffer at 4 °C. Sections were rinsed in 0.1 M sodium cacodylate buffer, followed by a quick rinse with ddH₂O. Cells were post fixed with 1% aqueous osmium tetroxide for 1 hour, followed with an En bloc stain of 2% aqueous uranyl acetate for 1 hour. Sections were washed again in ddH₂O, dehydrated through graduated ethanol (25-100%), infiltrated through an ascending ethanol/epoxy resin mixture (Embed 812, EMS), and then covered with pure resin overnight. Chambers were separated from the slides, and a modified #3 BEEM embedding capsule (EMS) was placed over defined areas containing cells. Capsules were filled with pure resin and placed in vacuum oven to polymerize at 60 °C for 72 hours. Immediately after polymerization, the capsules were snapped from the substrate to dislodge the cells from the slide. Semithin sections (0.5 - 1 µm) were obtained using a Leica UC7 ultramicrotome (Leica, Buffalo Grove, IL), counterstained with 1% Toluidine Blue, cover slipped and viewed under a light microscope to identify successful dislodging of cells. Ultra-thin sections (85 nms) were collected on 300 hexagonal mesh copper grids (EMS) using a Coat-Quick adhesive pen (EMS). Sections were counter-stained with uranyl acetate and lead citrate and imaged with a Hitachi 7700 Electron Microscope (Hitachi High-Technologies) using an advantage CCD camera (Advanced Microscopy Techniques). Images were adjusted for brightness, contrast, and size using Adobe Photoshop CS4 11.0.1.

[0217] Data analysis showed that the vesicles were lipid or glycogen-filled vesicles and autophagosomes (**FIG. 8B**), all reported previously in bat inner cell mass cells and other pluripotent stem cells. The most prominent vesicles, some surrounded by lipid membranes, contained a significant number of structures resembling viruses-like particles (**FIG. 7B**).

[0218] Interestingly, the virion structures did not belong to a uniform set of virus categories. While some exhibited features of (endogenous) retroviruses, other virus-like particles were packed in highly electron-dense material and resembled DNA viruses. Finally, numerous intermediate assemblies were much smaller than the more “mature viruses” but could also be defective exogenous retroviruses and many of them were embedded in double-membrane structures (**FIG. 7B**). Some of the virus-like particles must have been shedding into the supernatant as significant levels of retroviral activity (1.21×10^{10} viral particles per mL) were detected in the culture medium. These observations suggest that bat cells either produce active particles through endogenized sequences in their genome or through persistent infection that was already present in the BEFs. Previously, ERV-like particles have been reported in naive pluripotent stem cells in mice and humans, and western blotting and immunostaining revealed high quantities of ERV antigen in the cytoplasm of bat stem cells (**FIG. 7D**, and **FIG. 7F**). Additionally, bat stem cells were positive for coronavirus antigen in western blots and immunostaining (**FIG. 7C**, and **FIG. 7E**) and stained positive with an antibody raised against double stranded RNA viruses (**FIG. 7G**), suggesting endogenous infection and expression of endogenized viruses or fragments of endogenized viruses on an unprecedented scale, not seen in other tumor or stem cell lines.

Image-based flow cytometry (ImageStream)

[0219] Cells were seeded onto 6-well plates and separated from irradiated MEFs via two-stage trypsinization after four days. Wells were dosed and incubated with 0.25ml prewarmed (37°C) trypsin which was removed and discarded at 4 minutes. An additional 0.25ml trypsin was added and the plate was again incubated. After eight minutes cells were removed and pelleted via centrifugation. The cells were washed twice in PBS containing 0.5% BSA, fixed and permeabilized with Cytotfix/Cytoperm. The Primary antibody was added at a dilution of 1:200 in wash buffer incubated overnight at 4°C. The cells were washed twice with 0.5% BSA/PBS, resuspended in wash buffer containing the secondary antibody at a 1:200 dilution. Cells were then resuspended in wash buffer, the secondary goat anti-mouse AF568 antibody and incubated for 1 hour at 4°C. The cells were washed as before resuspended in 0.5% BSA/PBS containing two drops/ml DyeCycle Violet to stain the nuclei.

[0220] Imaging was conducted with the ImageStream MkII, at 60x magnification with the extended depth of field mode for probe resolution. Images were acquired using the INSPIRE 2.0 software at the lowest flow speed. Fluorophores were excited by the 405 nm

and 568 nm lasers at 60mW and 100 mW, respectively. Cells in focus were gated via histogram of brightfield gradient R.M.S. values and an aspect ratio vs. area plot was used to select the population of single cells. 5000 individual images of focused single cells were taken. Gating was refined further post-acquisition via the IDEAS 6.2 software suite by the same methods and plots, yielding n=1846 (BiPS). This software was used also for image processing, in which a set of custom masks defined by logical operators were used to denote vesicles and sensitively assess probes. For vesicles, it was observed that they may be selected from other cell component by contrast (bright and dark) and also by aspect ratio, and therefore are defined here by “Dilate(Range(Dilate(Range(System(Peak. (Threshold(M01, BF, 70), BF, Bright, 1), BF, 20), 0-5000, 0.4-1), 1), 0-5000, 0.4-1), 1) Or Range (AdaptiveErode(LevelSet(M01, BF, Dim, 5), BF, 75), 0-5000, 0.5-1).” BF and BF2 represent each brightfield image taken of a single cell from each of the two cameras, M01 and M09 represent the corresponding channel masks for each channel and the remaining terms represent mask modifiers and their associated values in the IDEAS software. For resolving immunofluorescence, “Peak(System(M05, Ch05, 3), Ch05, Bright, 1)” where Ch05 represents the staining of interest and M05 represents the corresponding channel mask. Modification was necessary to sensitively include all representative fluorescence, and to distinguish individual foci. The nuclear mask corresponding to DyeCycle Violet staining was defined “Object(M07, Ch07, Tight)” and the cytoplasm was defined through subtraction of the nuclear and vesicle masks from the cell mask through the logical operator available in the software (“Not”). Vesicle-nucleus overlap was determined in favor of vesicles by excluding them from the nuclear mask (“Not”). Probe localization was then defined according to these entities using the respective definitions and the operator “And.” Statistics for foci were generated using the Spot Count feature with a connectedness of 4. Prism 9 was used for graphs and statistics.

[0221] The results show that the bat stem cells were positive for coronavirus antigen in western blots and immunostaining (**FIG. 7H and I**), and double-stranded RNA in immunostaining (**FIG. 7J**). The latter is considered a hallmark for the presence of replicative genomes from positive-strand and double stranded RNA viruses. Super-resolution imaging showed that the dsRNA was present in aggregates (micron-order) throughout the cytoplasm but essentially absent from the nucleus. Further, ImageStream analysis indicated a close quantitative relationship between viral antigens and the intracellular vesicles. Based on these findings, it appears that pieces of endogenous viruses are being expressed at a scale that has

not been observed before in any other tumor or stem cell lines originating from other animals and humans.

EXAMPLE 8 Identification of retroviral sequences in the bat pluripotent stem cell

[0222] This example describes the identification of retroviral sequences in the bat iPSC.

Retrovirus assay

[0223] 2 ml of tissue culture medium were collected, and retroviral particle concentrations were determined using the QuickTiter Retrovirus Quantitation Kit (Cell Biolabs) according to the manufacturer's instructions.

Reverse Transcriptase Assay

[0224] Reverse transcriptase enzyme levels were determined with the colorimetric reverse transcriptase kit (Roche) per the manufacturer protocol. Cells lines represented were lysed in RIPA buffer, frozen at -80°C, thawed on ice, collected and resuspended in the kit lysis buffer (10 µL pellet in 40 µL lysis buffer per colorimetric well). Incubation duration (15h at 37°C) was selected for maximal sensitivity to the limit of the kit (1-5 pg RT). Absorbance at 405nm was measured by microtiter ELISA plate reader. Sample absorbance measurements were fitted to a linear regression of the measured HIV-1 RT standards ($Y=2.549X$) to obtain RT concentrations in units of ng/well. The results show, that some of the virus-like particles shed from the BiPS into the supernatant as substantial levels of viral particles (1.21×10^{10} viral particles per mL as determined in a retroviral assay and 0.3 ng/well in a direct reverse transcriptase assay) were detected in the culture medium.

Plaque Assay

[0225] Supernatants were centrifuged at 10000 rpm for 5 min to remove cellular debris, and the cleared lysates transferred to new tubes. Lysates were then diluted in 10-fold dilutions 6 times. Quantification of infectious titer was then performed by plaque assays in comparison to SARS-CoV-2 infection as positive control. Briefly, Vero-E6 cells were plated as confluent monolayers in 12 well dishes. Media was removed, and wells washed in 1ml of PBS. 200ul of diluted lysates was then added per well and allowed to incubate for 1 hour at 37°C. After viral adsorption, lysates were removed from the well and cells were overlaid with Minimum Essential Media supplemented with 2% FBS, 4 mM L-glutamine, 0.2% BSA, 10 mM HEPES and 0.12% NaHCO₃ and 0.7% agar. 72h post infection, agar plugs were

fixed in 10% formalin for 24h before being removed. Plaques were visualized by staining with TrueBlue substrate (KPL-Seracare) and viral titers calculated and expressed as PFU/ml. Immunostaining with an antibody detected the endogenous retrovirus protein Herv K or a Pan Corona antibody in *Rhinolophus ferrumequinum* embryonic fibroblasts. Immunostaining with a Pan corona antibody in *Myotis myotis* fibroblasts or induced pluripotent stem cells (iPS) is shown in FIG. The results show that inoculated Vero cells with cell culture supernatant of the bat iPSCs in the plaque assay did not detect any measurable cytotoxic effects in contrast to acute infectious virus particles that served as positive controls (SARS-CoV-2 particles).

Metapneumovirus (MPV) infection of BiPS and mES cells

[0226] 50,000 mouse ES cells (R1) or BiPS cells were plated per well of a 12-well plate on irradiated CF1 mouse embryonic fibroblasts using mouse and bat culture medium respectively. After 24 hours, culture medium containing human Metapneumovirus with GFP (MPV-GFP) (ViralTree) with a final multiplicity of infection (MOI) of 3. Medium was changed daily, and samples were dissociated at 3 and 5dpi using trypsin/EDTA and the infection rate was determined by fluorescence activated cell sorting (FACS).

[0227] In line with the pro-viral environment that was observed transcriptionally, bat stem cells infected with an exogenous Metapneumovirus (MPV) in comparison with mouse stem cells revealed a particularly permissive environment for viral persistence, further underscoring the supportive nature of bat stem cells for viruses. These results suggest that bat stem cells execute a program that in other mammalian cells is activated only after a virus infection.

EXAMPLE 9 Identification of viral sequences in the bat pluripotent stem cell transcriptome

[0228] This example describes the identification of viral sequences in the bat IPSC transcriptome.

[0229] Endogenization of an unusually varied group of viral genomes has occurred in bats (for example described in Banerjee et al. 2020; Katzourakis and Gifford 2010; Jebb et al. 2020). Endogenized viral sequences are reactivated and tolerated by all pluripotent stem cells (Grow et al. 2015). As a result, bat pluripotent stem cells should express and tolerate a particularly wide range of endogenized viral sequences. First, endogenous retroviruses, which are abundant and diverse in bat genomes (Jebb et al. 2020; Hayward et al. 2013; Skirmuntt

and Katzourakis et al. 2019) were analyzed. As a starting point, anchor points of retroviral sequences that had been previously mapped (Jebb et al. 2020) were picked. To obtain a broader portrait of the virus-like particles and approximate their identity more specifically, RNA-seq data was re-analyzed and additional long-read RNA sequencing (iso-seq) was performed.

Iso-Seq library preparation and sequencing

[0230] Cells were lysed in 400 µl Trizol reagent (Life Technologies) and total RNA was extracted using the AllPrep DNA/RNA Mini Kit (Qiagen) including a DNase digest to remove any potential contamination from carryover of genomic DNA using RNase-free DNase (Qiagen,) according to the manufacturer's instructions. The extracted RNA was then purified using 1.8X RNAClean XP beads (Beckman Coulter) to remove any molecular impurities. Iso-Seq SMRTbell libraries were prepared as recommended by the manufacturer (Pacific Biosciences). Briefly, 300 nanograms of total RNA (RIN > 8) from each sample was used as input for cDNA synthesis using the NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module (NEB,), which employs a modified oligodT primer and template switching technology to reverse-transcribe full-length polyadenylated transcripts. Following double-stranded cDNA amplification and purification, the full-length cDNA was used as input into SMRTbell library preparation, using SMRTbell Express Template Preparation Kit v2.0. Briefly, a minimum of 100 ng of cDNA from each sample were treated with a DNA Damage Repair enzyme mix to repair nicked DNA, followed by an End Repair and A-tailing reaction to repair blunt ends and polyadenylate each template. Next, overhang SMRTbell adapters were ligated onto each template and purified using 0.6X AMPure PB beads to remove small fragments and excess reagents (Pacific Biosciences). The completed SMRTbell libraries were further treated with the SMRTbell Enzyme Clean Up Kit to remove unligated templates. The final libraries were then annealed to sequencing primer v4 and bound to sequencing polymerase 3.0 before being sequenced on one SMRTcell 8M on the Sequel II system with a 24-hour movie each. After data collection, the raw sequencing subreads were imported to the SMRTLink analysis suite, version 10.1 for processing. Intramolecular error correcting was performed using the circular consensus sequencing (CCS) algorithm to produce highly accurate (>Q10) CCS reads, each requiring a minimum of 3 polymerase passes. The polished CCS reads were then passed to the *lima* tool to remove Iso-Seq and template-switching oligo sequences and orient the isoforms into the correct 5' to 3' direction. The *refine* tool was then used to remove polyA tails and concatemers from the

full-length reads to generate final full-length, non-chimeric (FLNC) isoforms. The FLNC isoforms were then clustered together using the *cluster* tool to generate final, polished consensus isoforms per sample.

[0231] Briefly, the existence of viruses in the *Rhinolophus ferrumequinum* transcriptome was explored by analyzing the RNA-seq and Iso-seq data based on a metagenomic approach using Kraken2 v2.1.2 (Wood et al, 2019). First, the adaptors in the RNA-seq data were removed with Trimgalore v0.6.7 (Krueger et al., 2021) and all replicates for corresponding datasets were joined in one file. The reference library “RefSeq complete viral genomes / proteins” was downloaded and a custom database was built to identify matches within the processed RNA-seq or Iso-seq. To eliminate false positive hits that could be due to matches with any cellular transcript such as oncogenes that are carried by some viruses, a second analysis was performed after eliminating all reads from the RNA-seq and Iso-seq datasets that matched any annotated *Rhinolophus ferrumequinum* transcript. To do this, the Iso-Seq FLNC isoforms or RNA-seq trimmed fastq sequences were first mapped to the “*Rhinolophus ferrumequinum* genomic rna exons RefSeq” file “GCF_004115265.1_mRhiFer1_v1.p_rna_from_genomic.fna” using gmap/gsnap (doi.org/10.1093/bioinformatics/bti310). The sequences with no mappings were then used to identify viral sequences using Kraken2 as before.

Mapping of RNA-seq reads to bat genomes and quantifying expression of ERVs

[0232] To trim adaptors and generate quality metrics of the fastq files, Trimmalore v.0.6.6 (www.github.com/FelixKrueger/TrimGalore), a wrapper for Cutadapt (www.github.com/marcelm/cutadapt) and FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) were used. Then, reads were mapped to the genome of *R. ferrumequinum* (Bat1K assembly HLrhiFer5) using HISAT2 v.2.2.1 (PMID: 31375807) suppressing unpaired alignments for paired reads (--no-mixed), suppressing discordant alignments for paired reads (--no-discordant), and setting a function for the maximum number of ambiguous characters per read (--n-ceil L,0,0.05). Output files were then filtered to remove any unmapped reads (-F 4), sorted and indexed using samtools (PMC2723002). Aligned reads were then assembled into transcripts using stringTie v2.2.1 (PMC4643835) in stranded mode (-rf). To generate a Ballgown readable expression output with normalized expression units of fragments per kilobase of transcript per million mapped fragments (FPKMs), the Bat1K annotation of known endogenous retrovirus (ERVs) for *R. ferrumequinum* (PMID: 32699395) (www.genome.senckenberg.de/) were also used as input in stringTie. Output counts were post-process and plotted with a custom R script.

Mapping of Iso-seq reads to bat genomes and identifying ERVs

[0233] Iso-Seq transcripts were mapped to the genome of *R. ferrumequinum* (Bat1K assembly HLrhiFer5) using minimap2 (PMC6137996) in mode for long-read/Pacbio-CCS spliced alignment (-ax splice:hq), giving priority to known splice sites from an input annotation (Bat1K), to find canonical splicing sites GT-AG in the transcript strand (--junc-bed -uf), with a cost of 5 for a non-canonical GT-AG splicing (-C5), and excluding from the output any secondary alignments (--secondary=no). Output files were then filtered to remove any unmapped reads or those not aligned to the primary alignment (-F 260), sorted and indexed using samtools (PMC2723002). Aligned transcripts to the genome were intersected with known ERVs.

De novo assembly of potential virus-derived RNA-seq

[0234] The trimmed reads that were identified by Kraken2 v2.1.2 to map to viral sequences with a confidence score of 0 as described above were classified as either mammalian or non-mammalian using the VIRION database (Carlson et al., 2022) based on their viral taxonomic ID assigned by Kraken2. The data were converted to FASTA format using the Seqtk v1.3 program and the reads were assembled using the Trinity v2.12 software. To check and gather successful assemblies that had produced at least one contig, a custom BASH script was applied for both groups of mammalian and non-mammalian viruses.

Mapping transcripts to viral and mammal databases

[0235] To determine if the assembled transcripts represented an expressed viral sequence, all transcripts were mapped to a database of viral genomes using BLAST. The viral database consisted of genomes whose host species contained either 'human' or 'vertebrate' as specified in the NCBI database. Initially this list contained over 17,000 genomes. However, this was reduced to 3,922 genomes by taking only unique virus/strain names. An additional non-mammalian virus database was generated by combining all genomic sequences of viruses identified by Kraken2 and classified as non-mammalian via VIRION.

[0236] Transcripts were also mapped to a combined database of bat, human and mouse genomes to both confirm their presence in the bat and to exclude the possibility of false positives through contamination. For each of these transcripts, expected values for both bat and viral genome BLAST results were combined into a single metric via the following formula: $\text{Log}(\text{bat-expected value} + 1 \times \text{virus-expected value} + 1)$. A threshold of less than 0.3,

representing a combined e-value of less than $1e^{-50}$ for both viral and bat hits, was used to rule out potential false positives. In addition, SQUID (www.eddylab.org/software.html) was used to shuffle the 63 (bottom-up) and 82 (top-down) sequences while preserving the dinucleotide distribution (parameter -d) to obtain a conservative threshold to distinguish bona fide viral homology from matches by random chance. Shuffled sequences were mapped to both the bat genome and viral genome databases, with the same BLAST threshold applied. All transcripts passing this threshold were extended by 5000 bp flanks within the bat genome and these regions were subsequently mapped to the viral database to confirm their presence in a viral genome.

[0237] The resulting sequencing reads were mapped against a virus database, using a metagenomic classification tool (Kraken). Mapping of the RNA-seq data revealed the expression of a widely diverse set of retroviral families in bat pluripotent stem cells, which was undetectable in BEFs. The results revealed a taxonomically highly diverse “zoo” of assigned viruses belonging to several significant viral families (**FIG. 9A-C**, **FIG. 10A**). They included, but were not limited to, Paramyxoviridae, Rhabdoviridae, Filoviridae, Bornaviridae, Flaviviridae, Coronaviridae, Picornaviridae, and Retroviridae (**FIG. 9A-C**, **FIG. 10A**). Similarly, viral sequences in BEFs were analyzed, notably yielding some viral sequences but to a much lesser degree (**FIG. 10B**). This finding is surprising as post-implantation tissues typically do not exhibit endogenous viral activity, underscoring pro-viral environments that bats create. Hence, the metagenomic analysis strongly suggests the remarkable possibility that bat stem cells harbor a significant number of viral-like sequences.

[0238] The potential for confounding effects that might impact the metagenomic assessment could be three potential sources for distortions: (i) statistical stringency, (ii) cellular genes containing viral-like sequences (e.g., oncogenes), and (iii) potential xeno sequence pollution originating from the feeder cells. To address the first point, progressively higher statistical stringency was used, yielding an expected decrease in matches. However, even under the most binding conditions, it still resulted in a sizable number of hits. To exclude potential cellular genes misinterpreted by the classification algorithm as viruses, the RNA-seq and iso-seq were depleted from all sequences that match exons, which only marginally affected the number of hits. Finally, some of the classified sequences were checked for murine origin as was the case for several retroviruses. Somatic tissue-derived cells, such as mouse fibroblasts, do not express endogenous viruses in measurable quantities.

Hence, the ability to readily detect such sequences may suggest the intriguing possibility that the BiPS cells triggered their activation and expansion or even the infection of the BiPS cells. While confounding effects could affect the metagenomic classification process, it is highly likely that a significant body of proviral sequences inhabits BiPS cells.

EXAMPLE 10 Assembly of novel viral sequences

[0239] This example describes the assembly of novel full-length viruses, shorter viral insertions, and novel, more distant viruses based on the sequencing data from BiPS cells.

[0240] As a starting point, anchor points of retroviral sequences that had been previously mapped were picked. Curation of the RNA sequences predicted to match those genomic sequences allowed the identification of not only previously described full-length bat retroviruses (RFeRV, **FIG. 10C**) but also an undiscovered full-length retrovirus sequence, RFe-V-MD1 (**FIG. 9D**, SEQ ID NO:1). The RNA sequencing also readily revealed short integrated viral sequences, for instance, Columbid/Falconid herpesvirus and Sindbis virus (**FIG. 9E**, **FIG. 10A**). In this case, the metagenomic classification tool pointed to this sequence. Upon closer inspection, it was found that the transcripts came from a genomic region immediately adjacent to a LINE-1 sequence. Furthermore, it was discovered that some of the sequences formed stem-loop structures, thus suggesting a potential functional role of the RNA (**FIG. 9F**). Another case at point was a region residing in the first intron of the XPA gene (a DNA damage and repair factor) on chromosome 12. A BLAST search with the fragment showed homology to two human herpesvirus 4 isolates (HKD40 and HKNPC60), the human respiratory syncytial virus (Kilifi isolate), and a fragment of about 500 bp that was identified at the end of a SARS-CoV2 isolate in an infected patient (**FIG. 11C**, **FIG. 9G**). Additionally, a protein translation search discovered homologies to an RNA-dependent DNA polymerase of the lymphocystis disease virus and the erythrocytic necrosis virus (**FIG. 11B**). Finally, expression data in conjunction with the bat genome was analyzed for more distant viral sequences using metagenomic classification taxonomies. Analysis for spike protein-like sequences found distant matches, a nearly 50% identical sequence to either RaTG13 (**TABLE 4**) or the Scotophilus bat coronavirus 512 (**TABLE 3**) covering most of the spike encoding sequences (**FIG. 9H**). A phylogenic analysis revealed that these genomic sequences mostly resembled the spike protein-encoding genomic portion of human coronavirus 229E and the human coronavirus OC43, respectively (**FIG. 11D**). In both cases, a flanking LINE-1 sequence was present. This suggests that potential LINE elements are directly involved in the homing of viral RNA.

TABLE 2

Identifier	Fragment/Read ID	Source	Size	Identified	Homology	Summary of result
RFe-V-MD1	m64019_210624_011 637/39584940/ccs	Iso-seq RNA	6088 bp	Overlap of Iso-seq sequence with previously predicted gag sequence of an endogenous retrovirus	Full length endogenous retrovirus	Iso-seq sequence overlapping with a predicted retroviral gag sequence allowed for identification of a novel full retroviral sequence.
RFe-V-MD2	m64019_210624_011 637/330171/ccs kraken:taxid 93386	Iso-Seq;	3350 bp	Kraken analysis of Iso-seq data and sequence alignments	Columbid alphaherpe svirus 1; Tax ID: 93386	Kraken analysis of Iso-seq reads identified homology with Columbid alphaherpesvirus 1. A subsequent Blast search confirmed a partial alignment with the Columbid and Falconid herpesvirus 1 as well as the Sindbis virus. The homologous sequence codes for a 24 aa stretch that has 79% homology with hypothetical proteins CoHVHLJ_080/FaH\HV1S18 80 of the Columbid or Falconid herpesvirus, respectively. Part of the sequence that shows homology to the Sindbis virus defective interfering particle di-2 which has been shown to inhibit viral replication in infected cells in vitro (Monroe SS, Schlesinger S. RNAs from two independently isolated defective interfering particles of Sindbis virus contain a cellular tRNA sequence at their 5' ends. Proc Natl Acad Sci U S A. 1983 Jun;80(11):3279-83. doi: 10.1073/pnas.80.11.3279. PMID: 6304704; PMCID: PMC394024) and can form a hairpin structure.

Identifier	Fragment/Read ID	Source	Size	Identified	Homology	Summary of result
RFe-V-MD3	m64019_210624_011637/128451663/ccs kraken:taxid 85655	Iso-Seq	7955 bp	Kraken analysis of Iso-seq data and sequence alignments	Ranid herpesvirus 1, Tax ID: 85655	Kraken analysis of Iso-seq reads identified reads that show homology with the Ranid herpesvirus 1. Alignment analysis revealed that the particular Iso-seq read matches a genomic DNA fragment in the first intron of the Rhinolophus ferrumequinum XPA gene (a DNA damage and repair factor) on chromosome 12 that is known to harbor a predicted LINE-1 sequence. Closer inspection of this Iso-seq read revealed homology with two Human herpesvirus 4 isolates (HKD40 and HKNPC60), the Human respiratory syncytial virus (Kilifi isolate) and an about 500 bp DNA fragment that was identified at the end of a SARS-CoV2 isolate from an infected patient. Additionally, a BlastX search discovered homologies to an RNA-dependent DNA polymerase of the Lymphocystis disease virus and the Erythrocytic necrosis virus.
RFe-V-MD4	m64019_210618_193151/159712964/ccs kraken:taxid 693999	Bat genome	6404 bp	Kraken analysis of genomic reads	Scotophilus bat coronavirus 512; Tax ID: 693999 NCBI Reference: NC_009657.1	Genomic sequence found that has 42% Identity and 42% Similarity with the Scotophilus bat coronavirus 512.
RFe-V-MD5	hub_1489433_GCA_004115265.2_dna range=chr1:38151239-38156098	Bat genome	4860 bp	Target analysis of RFe genome with spike protein coding sequence of bat RaTG13 coronavirus	Bat coronavirus RaTG13 Tax ID: 2709072; NCBI Reference: MN996532.2	Genomic sequence found that shows 44% identity and 44% similarity with RaTG13 coronavirus.

Identifier	Fragment/Read ID	Source	Size	Identified	Homology	Summary of result
RfRV	Bat1k: scaffold_m29_p_34:1 ,856,366-1,866,014 / GCA_004115265.2:c hr13:14,355,027- 14,363,924	Cui J, et al., J Virol. 2012 Apr;86(8):4288- 93.	9649 bp	Transcriptio n profile in RNA-seq in genomic region that overlaps with the previously identified endogenous retrovirus	Previously identified endogenou s retrovirus	Transcription profile in RNA- seq in genomic region that overlaps with the previously identified endogenous retrovirus

TABLE 3

Alignment of identified sequence with the Scotophilus bat coronavirus 512 genomic sequence.					
Sequence 1	NC_009657.1				
Sequence 2	m64019_210618_193151_159712964_ccs				
Matrix	EBLOSUM62				
Gap penalty	16				
Extend penalty	4				
Length	6654				
Identity	2802/6654 (42.1%)				
Similarity	2802/6654 (42.1%)				
Gaps	383/6654 (5.8%)				
Score	10094				
NC_009657.1	21507	CAATTGCTTGGTTGCATTGCCTAAGTTG--CAAG-GTCTTACTACCACTC	21553		
m64019_210618	1	CTACTGCAGTATTCTCAGCTAGAGTTGTGCTGGGACTCACAGTCACCTT	50		
NC_009657.1	21554	-TGTCTTTTGACTCACCACITTAATGTGCCTGGGTT--TTCCTGTAACGGC	21600		
m64019_210618	51	GAGGAACTTTACAAACC--TTTACAGGCCTGGACTCCTCCCTGAAGGTTT	98		
NC_009657.1	21601	GCCAATGGTTCTAGCTCAGCGCAAGCCTT-TCGTTTAAACGTCAATGATA	21649		
m64019_210618	99	TTGACTGAGTCAATCTAATAAG--GCCTGGACATTGTGTATTAGAAATA	146		
NC_009657.1	21650	CTAAGTTGT-TGTGTGGTGTGGCGTGTTCATT-GAACACCGTCGATG	21697		
m64019_210618	147	GTTCCCGAGTTTCTGATACAACCTTGTTCAAAAGTACTGAATGTATA	196		
NC_009657.1	21698	GTGTTAATGTTTCTATTGTGTGCTCCAATAATGCAACACAGCCCACTAGG	21747		
m64019_210618	197	AGTGTAATGTATCACTTCACAGATTTC-AAGCGTAAGAAACCTCTAGA	245		
NC_009657.1	21748	TCAA--ACAACCTGCAGGAAGACCTGCCTTACTATTGCTTCCTAACACT	21795		
m64019_210618	246	AAAGGTACAGTTAGTSAGGGGTACTTACTTCATCTCTCTCTCTTTCGCAACA	295		
NC_009657.1	21796	AGTAGCGGCACTAATCACACTGTTAAGTTTCTTTCACTTTTCCCGCCAAT	21845		
m64019_210618	296	CGCTTAGGC-----TGACT-TGAACCTGTCTTTACCAGTGGGCTCGAGAAG	339		
NC_009657.1	21846	CATTGCTGAGTTTGTGATCACCAAATATGGCAATGTCTATGTTAATGGCT	21895		
m64019_210618	340	TGAAGGGATATCATTGGCGGTATCTGCTGGCCTACAGAAGTACAC--CT	387		
NC_009657.1	21896	ATATCTATTTGAGAACTAGACCATTGACAGCCGTGCACTTGAACGCATCC	21945		
m64019_210618	388	GTTTCCTTTT-----TGCCAT---CAGTCATTCACTGGCGCTCAGAG	426		
NC_009657.1	21946	TCTCATTGCGAGGACGTAGCAGGGTTTGGACTATTGCCGCCACAAACTT	21995		
m64019_210618	427	ACTCTTTAGACATTTGCTGATGGGTAGTCTACTAATAA-GTAGAATCCGA	475		

Alignment of identified sequence with the <i>Scotophilus bat coronavirus 512</i> genomic sequence.					
NC_009657.1	21996	CACGGATGTGCTTGTGAGGTGAACACACAGG-CATTGAGAGTTGTTG	22044		
m64019_210618	476	CCC-----TTGAAGAAAGAGTTTTCATCTCTGTTACGCTCAGGTCGTTA	521		
NC_009657.1	22045	TATTGTGACACGCGCTGAAAACASTGTCAAATGTTACAACTCTCTTTTGA	22094		
m64019_210618	522	GATCAATAAATGTTTACCACCAG---CATGCTTTTCTGTCAGCAGTAAGA	568		
NC_009657.1	22095	ACTGGAGGACGGGTTTATTCCATGACTGCAGATAATGTTTATGCAGTAA	22144		
m64019_210618	569	AATCATGAAC-----CTTCCTTTT-----AGTTGAAGCT-GTGCAGATAG	606		
NC_009657.1	22145	CTAAGCCCCACACGTTTGTGACTTTGCCACGTTTAAATGACCATGGGTTG	22194		
m64019_210618	607	ACTGTCTCAATATGCTGTTTAATT-CCTACA-----GCCCTGG---C	645		
NC_009657.1	22195	GTAAATGTTACTGTGGGTGGTAACTTTGACAGTTTCATACCCACCAAAGTT	22244		
m64019_210618	646	ATAATCAGGAGGGTAGGTTTAAACATCAAAACATCAAGAAC-CTGGAGTT	694		
NC_009657.1	22245	CACCTGCTAATGGCACCTTAGTTAATAACGGCACTGTGGTGTGTGTCACCT	22294		
m64019_210618	695	CGTACAAAACACAAACGGAACGTGTTT-CAATAATTTAGAATCAG-CGGTC	742		
NC_009657.1	22295	CTAATCAG---TTCACCCCTAGACACGACTTTATGGTAGGTTATTCTGCT	22341		
m64019_210618	743	CTATGCAGGAATTGAGTATTTGATTA-ACAATCTGTGAAAAATAAATGCA	791		
NC_009657.1	22342	GATATGCGTAAGGGTATATTTGAGTACTCTAGTACATGCCCTTTCAATAG	22391		
m64019_210618	792	AATGGACTGTGGTGTGAATAAGTTTGA AAAATTCCTGAAGTGGGTAAAA	841		
NC_009657.1	22392	AGAACTATCAATAACTACCTTACGTTTGGTCGTATTTGTTTCTCTACTT	22441		
m64019_210618	842	TGAGTCTA--AATGACAAGCTGTC-TTTATTGCAAGCTGCGGCCCAATTT	888		
NC_009657.1	22442	CACCGCGCGGACGGTGTCTTGGCAATTGAAGTACTATGTTTGAACACCAAT	22491		
m64019_210618	889	TTGGATAAAACGTAGGATACCAAGAAAGGAAAGATTTACATACAAATA	938		
NC_009657.1	22492	GGAGCCCTTT-CACACCTGGCTGGCACCTTGTATGTTCAACATACAAAGG	22540		
m64019_210618	939	TTACATTATACACAAACCAATTTGGAACA-GCAAAATATAAAATCCCAAG-	986		
NC_009657.1	22541	GTGACATAATAACTGGGTACACCCAAACCAATTGCAGGGTTTGAATGACATT	22590		
m64019_210618	987	-TGCCCTATAGACT---AAATGTGTCTCCTGGATATGTTTAATTGCTC	1032		
NC_009657.1	22591	TCTGAATTGCACCTAGACACGTGCACCACTTACACCAATTTATGGTTTATAG	22640		
m64019_210618	1033	CATGTATTGATCA---ACGCTGACAAATTTGGAGACTCAGTAGAATGTG	1079		
NC_009657.1	22641	GG-GTGACGGTGTATTAGGTTGACCAATCAAACTTTCTTGTGACGGTGTG	22689		
m64019_210618	1080	GATGTCAAGCTTCTCTAAAGAAATCAACCTGGTCCTTCTTGATCACTTAC	1129		
NC_009657.1	22690	TA--CTACACTTCAGAGAGTGGTCAGTTATTAGCT--TTTAAAGAAATGCA	22735		
m64019_210618	1130	TGTGCTGTGCCTTTTAGATCAGACATGTGTTCTCTAGTTTGCTACTGTCC	1179		
NC_009657.1	22736	CTACAGGGCAGATTTATTCTGTGTACACCTTGCCAACTGGTTGACGAGTT	22785		
m64019_210618	1180	CCAC-----CTGGGTTCTTCACTTACTTT-GGTACCTTCTTTGTCTCCAT	1224		
NC_009657.1	22786	GCTTTTGTGAGGATAGGATTGTTGCGGTC-ATTACTAGTGTCTAATAATA	22834		
m64019_210618	1225	AAGCAC-TGGAGGTTACGATTCTTGTCTTATAGTATACGAGGTCTGACAA	1273		
NC_009657.1	22835	CTGGGTTCTTTAATTCAC-CAAGAACATTTCCAGGCT-TCTATT-----	22876		
m64019_210618	1274	TTACATTGCTGAATTCATTCTAGAAAAAGTAATGCATATCTCATTTGGTGA	1323		
NC_009657.1	22877	ATCACTCTAATGACACCACCAATTCACCTCACCAGACTTGTTTACTCT	22926		
m64019_210618	1324	ATATCACTGTGGTCACCTTCAAA-GTACTCCCTTGGGAAGCTGTGCACT	1372		

75

Alignment of identified sequence with the <i>Scotophilus bat coronavirus 512</i> genomic sequence.					
m64019_210618	2230	GTTTACTCCACTAGTACACTGGCATT	TTATCTTGGCATTATCCTCATAG	2279	
NC_009657.1	23862	GTTAGATCAGATAACTGCTGATGCGCAAGTTGACAGACTTATCACAGGTC		23911	
m64019_210618	2280	ACTTGSACTAACACGTCC---CTGATTTCACCTCCACTCTTG---CCAGGTT		2325	
NC_009657.1	23912	GGCTTGCAGCTCTTAATGCCCTTTGTTGCCAGTCACCTACCAAGTATGCA		23961	
m64019_210618	2326	TACCAAGAAAT-TTAATG--TTTGT-CATTGTTCTAATTCAGCT--CA		2369	
NC_009657.1	23962	GAAGTGCAAGCTA-GTAGGACATTGGCCAAAGCAAAAGTTAACGAGTGT-		24009	
m64019_210618	2370	GACATTCTTGCATGCAACACAAAAACACACACACACAAATATGAATGCC		2419	
NC_009657.1	24010	GTTAACTCACAGTCCCCAGAT----ACGTTTCTGTGGTGTGAAGGGG		24055	
m64019_210618	2420	ATTACGCAACACCGCCACATGTCCACACGGACACAGCTGTGAGATTTATA		2469	
NC_009657.1	24056	AACATA--TTTCTCACTCACCCAAAGCTGCTCCACAGGCTCTGATGTT-C		24102	
m64019_210618	2470	TACCAAGGTTATGAAACCTTATCGAGCTGTTTGTACAGTGTGCCAATGT		2519	
NC_009657.1	24103	CTACACACCGTTTTACTACCTAATGGTTTTATTAACGTTACAGCAGTTAC		24152	
m64019_210618	2520	AAACGCAAGGTGGCAAGTTCTCGAACTTAATTTTAAG--ACCCCAIATTT		2567	
NC_009657.1	24153	AGGTTTATGTGTGTGATGAGACCATAGCTATGACATTACGTCAGAGTGGAT		24202	
m64019_210618	2568	TGACAAACATTTCAAGATTTTCAATACA---GTCAATGT-TCTATGTTGAC		2614	
NC_009657.1	24203	TTGCTCTGTTTGTGCAAAATGG-TAATTATCTCGTG-TCACCGAGGAAAA		24250	
m64019_210618	2615	TTATTATGAGTATTTTATTTAAATTTTTTTATTGTGCTATATAGGGGAAC		2664	
NC_009657.1	24251	TGTTTGAACCTCGGAGACCTGAAGTTGCTGATTTTGTGCAAGTAAAAACA		24300	
m64019_210618	2665	AGTGTGTTTCTCCAGGGCCCATCAGCTCCAAGTCATTGCCCTTCAATCTA		2714	
NC_009657.1	24301	TGCACGATTAGTTATGTTAATCATCACCAATAACCAGTTGCCCTSACATTAT		24350	
m64019_210618	2715	GGTGTGGAGGGCAGACCT--CAGCTCCAA-GTCCAGTCGCGCTTTTTCAA		2761	
NC_009657.1	24351	TCC--AGATTATGTAGACGTTAATAAGACTATAGATGAGATTTTGGCCAA		24398	
m64019_210618	2762	TCTTTAGTTGCAGGGGGCGCAGCCACCATCCCATGCGGGAAITGAACCA		2811	
NC_009657.1	24399	CCTACCTAATAATACTGTGC---CTGATTGGCACTTGATGTCTTTAATC		24445	
m64019_210618	2812	GCAACCTTGTGTTGAGAGCTCACAGCTTAACCAACTGA-GCCATTAGGC		2860	
NC_009657.1	24446	AAACATTTCTTAATCTCACTGGTGAGATTGCAGACCTTGAAGCGCGATCT		24495	
m64019_210618	2861	CACCCCAACA-AAACGTATTGTTTA--TTTCAGAAGTGATACAGAAAATT		2907	
NC_009657.1	24496	GARTCCCTTAAAAACACATCAGAGAACTTAGACAGTTGATCCAAA-ATA		24544	
m64019_210618	2908	AGGTGAAAAGAGAAAAAATCA----TTCATATTCCCAATATCCAAAGACA		2953	
NC_009657.1	24545	TTAACAAACACACTTGTAGACCTTCAGTGGCTTAATAGGTTGAGACCTTT		24594	
m64019_210618	2954	AAAAACAGCACTGCTTCACATTTTAATAAATTCCTTAAAGTGTCTTCT		3003	
NC_009657.1	24595	ATTAAGTGGCCGTGGTACGTGTGGTTGGCTATTGTTATAGCTCTTATTTT		24644	
m64019_210618	3004	CTTTATT-----TAATCTCTACACTACACTTTTGAAAACAGACAAATTTA		3048	
NC_009657.1	24645	GGTTGTTTCACTGCTTGTGTTCTGCTGTATATCTACAGGTTGTTGCGGTT		24694	
m64019_210618	3049	GGTTTAGTCTCTGGCAATGATTT-CTCCCTGTCTTTTAGAAGTT-----TT		3093	
NC_009657.1	24695	GTTCGGTGTGTGTGCTTCTGTTTCTCAGGTTGTTGTGCTGSAACATAAA		24744	
m64019_210618	3094	CTTGACTGTGCTGACTATTAAA-CATTGATATCTTCAAC-TTCTGAC		3141	
NC_009657.1	24745	CTT---CAACATTACGAACCAATAGAAAAGTTTCATGTGCAATAATGTTT		24791	

Alignment of identified sequence with the <i>Scotophilus bat coronavirus 512</i> genomic sequence.				
m54019_210618	3142	CTTAAACCCCTTCTAGTCTCAGTCAATATGATCAATATGCCCGCTTCT	3191	
NC_009657.1	24792	CTTGGTCTGTTCCAGTATACTATTGATACTGCAGTTGAGCACA-CTGTAG	24840	
m54019_210618	3192	C-----CCAT--CTATGAGCTGATAAATCCCAAATAACATCTTCTATAT	3234	
NC_009657.1	24841	AACATGCTAACTTGTCCCAAGAAGAGGCTTTGATGTTGGAAGAAAACATC	24890	
m54019_210618	3235	AGACTACTTTCTGATCATTA-AATCCATTTTCCATTTCCTTACAGCATA	3283	
NC_009657.1	24891	GTTCTCTGAGACAAGCTACACATGTTACTGGATTTTGTCTACCCAGTGT	24940	
m54019_210618	3284	AT-CCAC-CTGGATCTCTTAAATTTAAAC--ATCCATGCTGCCCTTTC	3329	
NC_009657.1	24941	TTTTGTTTACTTCTTTGCACTGTTTAAAGGCTTCAAGCTACA-AACGTAAT	24989	
m54019_210618	3330	TTCTGTACTCTCTTCAGATTAGTTT---GATTGCAAGTAAGAGAAAGTCAA	3377	
NC_009657.1	24990	TTGCTGCTATTTTTCAGCAGTTTGTAGCTTTTATTAATTTATGCACCCAT	25039	
m54019_210618	3378	AATCAAAT-TGTGTAGAAAAACAAAAACA--AAAACTCAAATAACCT	3424	
NC_009657.1	25040	TTTAATATTTTGTGGTGCATCTTCGACGCTTTTA-TAGTAGTCGCAACA	25088	
m54019_210618	3425	TTTGGTTCCAGGATAAGACTCGTTGGATGCAGAAGCTCCAAGTCTCACTG	3474	
NC_009657.1	25089	TTGACTTCTCTGCTATTGTTTTTGAACCTACTACTCATGGCGTTATAAAAC	25138	
m54019_210618	3475	ATCATGCGCCATTTCTGTTTGTCTAGTTTCTTCTCATTTTC-TCTTTTTTT	3523	
NC_009657.1	25139	TTATAAATTTCTTATTTACAACTCTTCCACACTTATGTTTTTACATGG-T	25187	
m54019_210618	3524	TTTTTAATTTCACTTCAGAAATGCTGG-TCAATTTGTGCCCAAAATGACT	3572	
NC_009657.1	25188	CATGCCAATTATATAATGGCAGGC---CCTATGTAATGCTGAAGGTGGA	25235	
m54019_210618	3573	ACCACCATCTCCCTAAACTGCATGCTTCCAGATTCTAACCAGGCAGAAAA	3622	
NC_009657.1	25236	AGCCATTACGTCA-CATTGGTACTGATATAGTACCATTCTGACCCGAA	25284	
m54019_210618	3623	GAACGTGTGACGCTTCTGTTTTTACTATTTTCTTAGTATT--TCCACCAAA	3670	
NC_009657.1	25285	GTAATCTCTATCTTCCCATTCGTGGTAGTGTGAG-TCAGATATCCAACT	25333	
m54019_210618	3671	GTTTGTGACAGTCACTCAGAT--TAGAACGGCTAAGGTCACATGTCCAACA	3718	
NC_009657.1	25334	GTTGAGAAGTGTGAGT---TGTTAGATGCTAATTAC--CTCTA-----C	25373	
m54019_210618	3719	CTGAACCAATACGTTAGCCAGAGAGATGCAATGAACGCTCTGTTTAGC	3768	
NC_009657.1	25374	ATTTTCTCCAGTTGTCAAGTCGTTGGTGTACTAATTCAGGTTTGTAG-G	25422	
m54019_210618	3769	CGTTTCACATCATCGCAGGGCTCATGGGTACCTCCCACGGGCTGAGAGTG	3818	
NC_009657.1	25423	AGATTCAACTAGACGAATATGCTACAATTAGTGAATGATAATCGTGTAGT	25472	
m54019_210618	3819	GGGGAAAGAGGGACAGAACCTCAAATGAAACACAGAGCTGCTGTGAGAA	3868	
NC_009657.1	25473	TGTAAATGCGATTCTCTGGCTTTTGTACTCTTTTGTGTC-TAGTTATT	25521	
m54019_210618	3869	AAAGCAAATGGATGTCAAGAATTAAACAATAATACCTGACCCCTCTTTAT	3919	
NC_009657.1	25522	AGCATTAATTTCTCCCAAC---TTATAAACCTTTGTTTACTTGCACCCG	25568	
m54019_210618	3919	TGAATGGCACTCACTCATCCAGTCCAAAACCTGGCATCATGTGAGACCA	3968	
NC_009657.1	25569	GTTGTGTAATAACGTTGTTTATAAGCCTGTTGGAAAAGTATACGGAGTAT	25618	
m54019_210618	3969	CATTACTCTGACCTCTGCTT-----CCATAATCACATCTCTTTGTATGAT	4013	
NC_009657.1	25619	ACAAGTCTTATATGCGAATTCACCCCTTGACATCTGACATTATTCAAGTA	25668	
m54019_210618	4014	TCTCTTGTCTACCTCTTTCACCTTACAAGGACCCCTTGAGATTACA-ATGGA	4062	

Alignment of identified sequence with the <i>Scotophilus bat coronavirus 512</i> genomic sequence.					
NC_009657.1	25669	TAAACGAAATGTCTTCGAACCAATCCGTTCTGTAGAGGAGGTGATTAA	25719		
m64019_210618	4063	TCCACACAGATAA-TACAAAACAATCTCCCATCTCAATAGTCTTAATTT	4111		
NC_009657.1	25719	ACACCTCAGAAATGGAACCTTTTCATGGAATATCATACTTACAATACTCT	25768		
m64019_210618	4112	AATCATTGTACAAGGTCCATTTTGCTGTATAAAGTAACATGTTAACATA	4161		
NC_009657.1	25769	TAGTAGTGTTCAGTATGGACATTACAAATATTCAGGGTTCTCTATGGC	25818		
m64019_210618	4162	TTTCAGGGATTAGGATTAGCACATTTTGAGGGGCCATTATTTTGCTTGCC	4211		
NC_009657.1	25819	TTAAAGATGGCCATTCTTTGGCTTCTTTGGCCACTTGTCTGGCCCTTTC	25868		
m64019_210618	4212	ACACCCACA-TATTTCTTTAGAATCATCTTTAGCATAACCTAAT--TTTA	4259		
NC_009657.1	25869	CATCTTTGATGCCTGGGCCAGTTTTAATGTTAATGGGTTTTCTTCGCAT	25918		
m64019_210618	4259	GAAATGTGTT---CTGGCATTATGTTTATCTGGGTGCTTCTCTTTACTT	4306		
NC_009657.1	25919	TCAGCATCCTAA-TGGCCTGCGTCACAGCTGT-GCTGTGGATTATGTA	25966		
m64019_210618	4307	GCTTAACACTCTGTATCCTTCACTCTAGCACTCAACACCCACTCTGTCCC	4356		
NC_009657.1	25967	TTGT-TAACAGTATCAGGTTGTGGCGACGACCCATTCTTGGTGGTCCTA	26015		
m64019_210618	4357	TCATGCAACTTTGTGAGTTTCTCATG-CAAAACAACCTTGATTATTCAT	4405		
NC_009657.1	26016	CAATCCTGAAACGGACTCTATTCTGTCTGTCTGTGCTGGGTGGGCATG	26065		
m64019_210618	4406	TTCTGAGCAATAATGCCCAACTCTGTCTGGCACAACCAAGGAATTAATA	4455		
NC_009657.1	26066	TCTGCCTACCAATACCTTGGTGCACCCACGGGCGTAACGCTCACACTGCTT	26115		
m64019_210618	4456	ATTATAGTTCTAGAGTCCCTTAACCATCAACCTAAAAGCTTGATAGTTT	4505		
NC_009657.1	26116	AATGGCACATTCCTTCTAGAAAGGCTATCAG-GTTGCT-ACTGGCGTACAG	26163		
m64019_210618	4506	TGATCCCAAAATCCCAATTAATCTCAAAGTGTGCTGAGTGAATCACAA	4555		
NC_009657.1	26164	CTAAATAATTTACCTCGTTACGTAACAGTCGCCAAAGCTTCAACAACAAT	26213		
m64019_210618	4556	TGAATATATTTTACATTTGAAAGGAATTTGGCCAAAGTT-----CACT	4598		
NC_009657.1	26214	TGCTACACAGCGTGTGGGACGTTCCATGAATGCAAAATCAAGTACTGGCT	26263		
m64019_210618	4599	TTACCTTCTAAATTTCAAATAAGCCAA-TTGGACCACTGAATTTTAGTAT	4647		
NC_009657.1	26264	GGGCTTTCTTCTGTGAAGTCCAAGCATGGCGACTACTATGCTGCTGCGAAT	26313		
m64019_210618	4648	TTAATATAATGATGT-GCCATTGTTCTTAGTCAACTAAGAAACAAA-ACA	4695		
NC_009657.1	26314	CCAACAGAGGTTGTAACAGATAGTGAGAAAATTTACATTTAGTCTAAAC	26363		
m64019_210618	4696	CTAAATACCTTTTAAAAAGAGTTTAAAAAAGAGCTTAAAT	4745		
NC_009657.1	26364	AGAAACTTA-TGGCTTCTGTAAAATTCACACCTCGTGGTCTGTTCCAAGGG	26412		
m64019_210618	4746	GACTTCTTGGTTTCATCTGTTACAATCAAGTTTCAAGTGC-TTCTGAGA	4794		
NC_009657.1	26413	ACGTGTTCTCTGTCTCTTTTGTCTCCACTTAGGGTTACTGATGAAAAAC	26462		
m64019_210618	4795	AAGAAGTCTAGSAAAGAACAACTAAAAAACTGTGGACATTACAGAGCAAC	4844		
NC_009657.1	26463	-CACTTTACAAGTCTCTACCAATAATGCCGTCCCTCAGCGAATGGGAGG	26511		
m64019_210618	4845	TCTGAATACATGAATTGACACAGGTGTGCCTTAACTTTAATACTCTGTGT	4894		
NC_009657.1	26512	TAAG---GACCAACAAATGGATACTGGGTTGAACAACAGCGCTGGAGAAT	26559		
m64019_210618	4895	CACATTGACCCAATGTACCTCCTCAGCCAGTCTTC-GAGCTC-TGTTT	4942		
NC_009657.1	26560	GCGCCGCGGAGACAGAGTTGACCTGCCATCTAACTGGCACTTCTACTTCC	26609		
m64019_210618	4943	TCTCAT---GGTCCAGAGTCAATTTCTCAAACTGTAAGCACACATCCATCT	4990		

Alignment of identified sequence with the <i>Scotophilus bat coronavirus 512</i> genomic sequence.					
NC_009657.1	26610	TCGGTACTGGACCGGATTCTGATTTCGCTTTCAGAAAACGCACTGATGGT	26659		
m64019_210618	4991	GCAGAT-----GCAATGGGAT---CCCTACCAGCATCAATGTGAGCCT	5030		
NC_009657.1	26660	GTTTTCTGGGTTGCA-ATCGATGGTGCTAAGACCCAGCCAAACAGGCCTTG	26708		
m64019_210618	5031	TATGGCTGAAAGACATATCAGTAGTCCAATCACCA--CCCTGTACCCGC	5078		
NC_009657.1	26709	GGGTACGTAAGTCGTCTGAGAAGCCGTTGGTTCAAAAATTTAAGAACAAA	26758		
m64019_210618	5079	CCTTTTCAATGGAAG-CTCTGAAGCAGTCTCCCTCATAAGTGTGAACCTTG	5127		
NC_009657.1	26759	TTACCCAATAATGTGGAATCGTTGAACCTACACACCAACAACCTCCAG	26808		
m64019_210618	5128	AGAAGAAATAATCTGCCAAGAAGGATTCTCATGGTTAACTGAGCTCAAA	5177		
NC_009657.1	26809	AGCTAACTCAAGGAGTCGTAGTCGTGGTGGACAGTCCAAACAGCAGAGGAA	26858		
m64019_210618	5178	TTCTTAATAGAGTC-TAACAGCCATTCTTGACA---CAAGCCTCTCGCACA	5224		
NC_009657.1	26859	ATTCCCAAAACAGAGGT--GATAAATCCAGAAA---CCAGTCCAGAAACA	26903		
m64019_210618	5225	CTCTGCATTTTCAAGGAAAACCCACAGACTGAAATTTCCACCTCCCGAAT	5274		
NC_009657.1	26904	GGAGTCAATCTAATGATCGTGGGCTGACTCGCGAGATGACTTAGTGGCT	26953		
m64019_210618	5275	GTGCTCCTGCTGCTG---CCTAAGTCAACCATTTGTGAGGAATTCCTGATG	5322		
NC_009657.1	26954	GCCGTTAAAAAAGCACTT---GAAGACCTAGGAGTTGGTGTGCAAAAGCCA	27001		
m64019_210618	5323	GAACTCCTGATGGAATTCCAAAGGACTGAGACTAGTCCCATCCAATCAG	5372		
NC_009657.1	27002	AAA---GGC---AAAAACCCAGAGTG-GTAAAAAC--ACCCCTAAGAACA	27042		
m64019_210618	5373	AACTGTGGCGTTATATCCTCATTTGCATCTATACTGACCAATCAGAACTG	5422		
NC_009657.1	27043	ATCTAGGTCAGGCTCTGTGCA-ACGTGCAGAAAGCCAAAGGACAAACCCGAG	27091		
m64019_210618	5423	ATTCAACAACCAATCAGAACATATGATGCTGACT-GATCAGAACTGTG	5471		
NC_009657.1	27092	TGGCTCGTACTCCTAGTGGCGATGAGTCAGTTGAGGTTTGTGTTGGACC	27141		
m64019_210618	5472	TGATTTGGATTCTCATTTGCATAAAATGGACCAATGGGAACAGGGC	5521		
NC_009657.1	27142	CCGTGGTGGCACCAGAAATTTTGGTAGCTCCGAATTTGTGTC-TAAAGGT	27190		
m64019_210618	5522	ACTAACTTTCTCTGTAAAGGCCCTTCCCTTTGTCTGTGCTGCACTT	5571		
NC_009657.1	27191	GTGAATGCCCCCGGTTATGCTCAG---GCTGCTTCACTGGTACCCGGCG	27236		
m64019_210618	5572	TCGGTTTTTCTGTTTACCAACTGTTTCAAGTGAATAAAGTTATCCTCTT	5621		
NC_009657.1	27237	CCGCAGCACTGCTTTTGGTGGTAAATGTTGCCACCA---AGGAAATGG--	27281		
m64019_210618	5622	TC-CACACCTCATATTGGAACTTTTGTGATATGAGGTAGGCTATGGTC	5670		
NC_009657.1	27282	CTGATGGTGTGAAATCACCTATACATATAAAATGTTAGTCCCTAAGGAC	27331		
m64019_210618	5671	ACAATTCACAGAGGACCTTGAAGCTCAGTGAGATGATTTCAAATAAC	5720		
NC_009657.1	27332	GACAAGAACCCTTGAATCTTTCTTGCTCAGGTTGACGCATACAAGCTCGG	27381		
m64019_210618	5721	AGGAGGATTTCATCCAGATGATTTTGAGAAGGAAGAGTTACTGCCCCAGG	5770		
NC_009657.1	27382	CGATCCCAAGCCTCAGCGTAAAGTCAAAAGCTTCAAGAACCCTCAACCAA	27431		
m64019_210618	5771	AGTTCACTAA---GGAGTGAGGCTGCCCCAAACGTGAAAAGCTGTGT	5816		
NC_009657.1	27432	AACCTGCAACAGAGCCAGTTTA-TGACGAGCTTGCTGCAGATCCTACTTA	27480		
m64019_210618	5817	AGATGCCCTCGTGATATTTCACTGGAACAATA----TATTTTATTGTA	5862		
NC_009657.1	27481	CGCCAATCTTCAAGTGGGACACCACTCGAGGATGGTGTGAGATGATCA	27530		

Alignment of identified sequence with the <i>Scotophilus bat coronavirus 512</i> genomic sequence.					
m64019_210618	5863	GTTTGAAGAGTGAGGGTATCTAACAGGGAGGATGCTGCTGGACAGATT	5912		
NC_009657.1	27531	ACGAGGTTTTTGACACCCAGAATTGAATTCAACTAAAACAATGTACAGAA	27580		
m64019_210618	5913	CTCCAGT-----GTACACTACCGCCAGACAGC--ACACGGAGTTCACAA	5956		
NC_009657.1	27581	TTGTAGCTATTGCTTTTGGCTGAGCTTTTCGAGCACTGGCCATTTTGGC	27630		
m64019_210618	5957	AAG-ACAAAATGGTATGGTTGTGAGTAGGAATGCATTTTCTTTTTT---	6002		
NC_009657.1	27631	TCATTCTTCCAAATTTTGTGCTATATTTGATTGCATTTCCAAGGTGAG	27680		
m64019_210618	6003	-TTTTCTTCCTTTTCTTTTTTTTTTTTTTGGTATGCATTTTCTGATTAA	6051		
NC_009657.1	27681	TTTAAGCTGTCTACAGGACGTGGTGTTCCTTACATGTGCTGATTCC	27730		
m64019_210618	6052	TCTTCTGGAGTTGCTCAITGTACAGCATGAAAACACCTG--GAATTCT	6099		
NC_009657.1	27731	TTATTCTTGTGC-TCATATTCTTTCTTTCTTGGTGGCTTTTCTTACTG	27779		
m64019_210618	6100	TGTTTCTGACAGTGTCTATAGATCTTTGATCAGC-TATTTATGTTGCTC	6148		
NC_009657.1	27780	TTAGTGGTGTACATCGTTAA-ACATGATTGGCCCCCTGATGTGCTAT	27828		
m64019_210618	6149	AATGTCCACTTCCATAGAAAACAGTAGATGCAGCAGTCTAGTTCTCATT	6198		
NC_009657.1	27829	GTTAACCTCTACAGGCCCTTACATGATGGCTTAATCAGATTCTTATG-A	27877		
m64019_210618	6199	GCTCCACTCATCAAATTAACCAAGTCTGTATCTATCTGATGATGTGTATA	6248		
NC_009657.1	27878	CACCAGACTTTGCTGTCTTGGTTTTATCTTTCTTGTTCATGATCTTAACA	27927		
m64019_210618	6249	TATGTGTGTGTGGTGCATTAGATTGAGCTTGTGACAATAAACAATACT	6298		
NC_009657.1	27928	TG-GCTGCTGGGCATTGGAATCTTCCAATACTAGCGGT-CTTGGTCTTGC	27975		
m64019_210618	6299	TTTATTGACTGGGATACTGACCTACTTGTATATGTGCTGCCTCTTTAAAC	6348		
NC_009657.1	27976	ACACAACGGTAAGCCTGTAATAATGACAGTGCAAGCAGETTATTATTATA	28025		
m64019_210618	6349	CTTCCACTTTGTTTCAATAGAAATAGTATAAAAAACAAAAAGCTCTAGGAT	6398		
NC_009657.1	28026	TTGC	28029		
m64019_210618	6399	TTGC	6402		

TABLE 4

Alignment of identified sequence with the RaTG13 bat coronavirus genomic sequence			
Sequence 1	MN996532.2:21560-25369 Bat coronavirus RaTG13, complete genome		
Sequence 2	hub_1489433_GCA_004115265.2_dna		
Matrix	EBLOSUM62		
Gap penalty	16		
Extend penalty	4		
Length	3998		
Identity	1758/3998 (44.0%)		
Similarity	1758/3998 (44.0%)		
Gaps	281/3998 (7.0%)		
Score	6062		
21560-25369	8	TTTTTCTTGTTTTATTGCCACTAGTTTCTAGTCAGTGTGTTAATCTAACA	57
hub_1489433_G	134	TTGTTCACTATCTATTACATATTGAATTTTCAATAAATGTTAAGAGGCA	183
21560-25369	58	ACTAGAACTCAGTTACCTCCTGCATACACCA---ACTCATCCACCGTGG	104

Alignment of identified sequence with the RaTG13 bat coronavirus genomic sequence				
hub_1489433_G	184	GGTACAATTAA--TACCTCCA-CITTCAGATGAGAAAATTAAGGCAGAGA	230	
21560-25369	105	TGCTATTACCCCTGACAAAGTTTTCAGATCTTCAGTTTACATTAACTC	154	
hub_1489433_G	231	GGTTACATAATGTGCCAAGGTACCACACCTT--GATAAACAGC-AGCTG	277	
21560-25369	155	AGGATTTGTTTTTACCTTTCCTCTCCAA----TGIG-ACCTGGTTCC---	196	
hub_1489433_G	278	GGATTCTCACCCATCCAGTCAGCTTCAGANTCTGTGCACCTTACTACTAG	327	
21560-25369	197	ATGCTATACATGTTTCAGGGACCAATGGTATTAAAGGTTTGATAACCCA	246	
hub_1489433_G	328	ATGCTATATAGAAATTAATG--CCAAAACCTCTCAAAATCAGAGTCATGAGA	375	
21560-25369	247	GTCTGCCATTCAACGATGGCGTCTATTTGCTTCCACTGAGAAGTCTAA	296	
hub_1489433_G	376	GAAAAGCCAA--AGCCATCATGCCAATATTCTTAGGTTAGGTTAGGCTA	423	
21560-25369	297	TATAATAAGAGGATCGATTTTGGTACTACCTTAGATTGGAAGACCCAGT	346	
hub_1489433_G	424	TGTTAGGTTGCGTTTATTTTTT----ATCCCTTAATTCCTAATCT---T	467	
21560-25369	347	CTCTACTTATTTGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAA	396	
hub_1489433_G	468	CTACATTTAGGGGAAGAGATG-TGCTTCTATATTCATGAATGTTTATCAA	516	
21560-25369	397	TTTCAATTTTCTAATGATCCATTTTGGGCTGTTTATTACCAAAAAACAA	446	
hub_1489433_G	517	TG--AACATCGTATGGGACCATTAACCTGGACCCCTAAGAGATATGTTT	564	
21560-25369	447	CAAAAGTTGGATGGAAAGTGAGTTTACAGTTTACTCTAGTGCGAATAAT	496	
hub_1489433_G	565	TTGACATAATTCAATATCAATGATCAGCATT--CTCTT-TGGGTTGATTG	611	
21560-25369	497	GCACTTTTGTAGTATGTCTCTCAGCCTTTTCTTATGGACCTTGAAGGAAAA	546	
hub_1489433_G	612	GCCATGTCTTTATCATCTCCAGCTCTATAGAAGCTTCTT-ATGAAGAA	660	
21560-25369	547	CAGGGTAATTTCAAAAATCTTAGGGAATTCGTGTTTAAAGATATTGATGG	596	
hub_1489433_G	661	TATAGTCAGGACACACACACATACACACACGCGCGCGCGGATGGGGA	710	
21560-25369	597	TTATTTT-CAAAATATATTTCTAAACATAGCCCTATTAAATTTAGTGGGTGAT	645	
hub_1489433_G	711	CTCTTAACTAGCTCACCCTCACCACCAAAAGCCTCATCTAGAAGCACAGAT	760	
21560-25369	646	CTTCCCTCGTGGTTTTCAGCTTTAGAAC--CATTGG--TAGATCTGCCAA	691	
hub_1489433_G	761	TATCATGAATACTCTGGGGAGGGGGCATCATGGGGTGGCAATTCAAA	810	
21560-25369	692	TAGGTAT---TAACATCACTAGGTTTCAAACTTTACTTGTCTTACATAGA	738	
hub_1489433_G	811	GAGAAATGAGAAAAATCACAAGATCTTTAAATCAATGGGGATAGCGCTG-	859	
21560-25369	739	AGCTATTGACTCCTGGTGATTTCTTTCAGGTTGACAGCTG-----GT	783	
hub_1489433_G	860	-GAATTTTCCATCCTGAAGATTTTTCAGGGCTAAACCTCTGACTGAGT	908	
21560-25369	784	GCTGCAGCTTATTATGTGC-----CTTATCTTCAACCAAGGACTTTTCTA	828	
hub_1489433_G	909	TTTGTCTCTTTAAACAAAGAGCTGCTGCTGCTGCTGATACCTTATTTT	958	
21560-25369	829	CTAAAATATAATGAGAATGGAACCATACAGATGC--TGTAGACTGTG-C	875	
hub_1489433_G	959	CAAAAACGTTCTTTGTAACATCCAAATATTTTCATGAAAATGTTTC	1008	
21560-25369	876	ACTTGAC-----CCTCTTTCAGAAACAAAGTGACGTTAAATCCTTCAC	920	
hub_1489433_G	1009	TCTTACATGTGACCTCAATTTGACTCAGC-TGACCCTGTGACTACTTGA	1057	
21560-25369	921	TGTTGAAAAGGGAATTTATCAAACTCTAATCTTAG--AGTCCAACCAAC	968	

Alignment of identified sequence with the RaTG13 bat coronavirus genomic sequence				
hub_1489433_G	1058	-CTTGTGCTGCAACAAAGTCCACAGTTTCCTCCTGCAAGTCTTTCAATTT	1106	
21560-25369	969	AGATTCTATTGTTAGATTCCCAAAATATTACAAA-----CTTATGTCCT--	1011	
hub_1489433_G	1107	TCATTGTATGAGGTGTGATAAAAAAATACAGTGAATGTTAAATAAAAA	1156	
21560-25369	1012	-TTTGGTCAAGTTTTTAACGCC--ACCA--CAATCGCATCAGTTTATGCT	1056	
hub_1489433_G	1157	ATTTATTACAGTAAAAGACACATTACCATTAAATTCCTCCTCAAAATACTCC	1206	
21560-25369	1057	----TGGAACAGAAAGAGAATTAGCAACTGTGTT--GCTGATTACTCTGTCC	1102	
hub_1489433_G	1207	CCCTTGCTTTGAACACATGTATCCCTTTGTTTTGCCCACTTTCTGAAGCA	1256	
21560-25369	1103	TATAT---AATTCACATTCATTTTCTACCTTTAAATGTTATGGAGTGTCTC	1150	
hub_1489433_G	1257	GTTCTGGAAGTCTCTTTTATGAGTGTCTTTAATGTACTGTAGTGGCTG	1306	
21560-25369	1151	CTACTAAATTAAATGATCTCTGCTTACTAATGTTTATGCAGACTCAATTT	1200	
hub_1489433_G	1307	CTT-TGATGTCTGTAATCAATTCAAAAAGTTTACCTTTTGTGG-TCATTT	1354	
21560-25369	1201	GTGATTACAGGTGATGAAGTCAGACAAATTCGCCAGGACAAACTGGAAA	1250	
hub_1489433_G	1355	TTTCTTTAGGGAAGGCCAGCAGTCGCCCGTGCCAGATCTCGTTAATAA	1404	
21560-25369	1251	GATTGCTGACTACAATTATAAATACC--AGATGATTTTACTGGTGTGT	1298	
hub_1489433_G	1405	GCCGGATGAGGCACACCCATAATGTCTTTAGTTGACAGAAATGCTGTAT	1454	
21560-25369	1299	TATAGCTTGGAAATCTAAGCATATTGATGCAAAAGAGCGCGTAATTTTA	1348	
hub_1489433_G	1455	ACCAGAAGCGATGTTGGAGCATTGTCTATG--ATAGAGGATGATTACAGC	1502	
21560-25369	1349	ACTATCTTTACCGTCTCTTTAGAAAAGCTAATCTTAAACCTT-TGAGAG	1397	
hub_1489433_G	1503	ACACTGTAAAACACACCTTCTCTCAAGTGTATCTCACACCAACTGACTG	1552	
21560-25369	1398	GGATATCTCAACTGAAATTTACCAAGCA--GGCAGCAAACTTGTAAATGG	1445	
hub_1489433_G	1553	CACCAACCAAGTTGAACCTTGTACACATCATTTACTAAGGTTTGACATGC	1602	
21560-25369	1445	TCAAACTGGTCTAAATTGCTACTACCCACTTTATAGATATGGATTITACC	1495	
hub_1489433_G	1603	AGCTTCTTGTATGAATATCCCTGCTTTCCATTGGATGGCACTTAGCAG	1652	
21560-25369	1495	CTAC--TGATGGTGTG-----GTCAC-----CAACCTTATAGGGTAGTAGT	1535	
hub_1489433_G	1653	CAACGTTCACTGTATTGTTAATCACACCTCGTACTTATTCTGATGGAGA	1702	
21560-25369	1536	ACTTT----CTTTGAACCTTCTAAATGCACCAGCAACTGTTTGTGGACCT	1581	
hub_1489433_G	1703	AATTTTGTGTCAGTTGAGCA-CACTTTCCTCTCTCATCCTTTTATTTCT-	1750	
21560-25369	1582	AAGAAGTCTACTAAGTTGGTTAAAAATAAATCTG-TCAAT-TTCAACTTT	1629	
hub_1489433_G	1751	---CTGTCTA--GCTTTAGTTTGGGATGAGGAGGACAAAGTACTATTAT	1795	
21560-25369	1630	AATGGTTTAAAC--TGGCACAGGTGTCTCAGAGCTCTAATAAAAAGTTT	1677	
hub_1489433_G	1796	TATGAATTACAGTGGCTCTGGAGGCTCTCAAAATCCTGACTATGACACA	1845	
21560-25369	1678	CTACCTTTCCAACAATTTGGTAGAGACATTGCAGAGCTACTGAT--GCC	1725	
hub_1489433_G	1846	GAAAATTCGAAATAATTCACAGCAGGAGTACTATAGGACTTGGTCAGCT	1895	
21560-25369	1726	GTCCGTGATCCACAGACACTTGAGATTCTTGACATTACACCATGTTCTTT	1775	
hub_1489433_G	1896	TTCATTGAACATAACTCCACATCTATATTGGCTCTGCTTTTGCTTCTCA	1945	
21560-25369	1776	TGGTGGTGT-CAGTGTATAACA----CCTGGAACAAATGCCCTACACCAG	1821	
hub_1489433_G	1946	ATATTAAATGCTCAAAATATGTCAGTGTAGGCACATATTATTTATATCCCT	1995	

Alignment of identified sequence with the RaTG13 bat coronavirus genomic sequence				
21560-25369	1822	GTTCGTGTTCTTTATCAGGATGTTAACTGCA--CAGAAGTCCCTGTTGCT	1869	
hub_1489433_G	1996	CTGAAACATGTTTCTATTCAAGGATGCAGCATTCAGAAGACTCAGT--CC	2043	
21560-25369	1870	ATCCATGCAGACCAACTTACTCCCACTTGGCGTGTCTTACTCCACAGGTTT	1919	
hub_1489433_G	2044	AGCGAGTGACAGAAAAGACTTCC-CTTGGATTATCTATG----AGATTG	2088	
21560-25369	1920	TAATGTTTTTCAACACGTCAGGTTGTTAATAGGGGCT-GAACATGTC	1968	
hub_1489433_G	2089	TAATAGCTTATCTGCATAT-CTGCTCACTGAATACTGCCTCGATCATTCA	2137	
21560-25369	1969	AKTAACTCG-TATGAGTGTGACATACCTATTGGTGC-AGGAATAIGCGCC	2016	
hub_1489433_G	2138	TATATCTGGCTCACAAATGGGTAATCAATAAATGTGTGATGAATGGTCTAC	2187	
21560-25369	2017	AGTTATCAGA-----CTCAAACATAATTCACGTAGTGTGCCAGTCAAT-	2059	
hub_1489433_G	2188	AATTCCAGATTGACGCCCTAACTTGCTCATGATG-GCTTCCAGTAGTTT	2236	
21560-25369	2060	-CTATTATTGCCCTACACTATGTCACTTGGTGCAGAAAATTCAGTTGCTTA	2108	
hub_1489433_G	2237	TCTATCAAAGCC-ACATGTGGTCAGT---GTGCAGGATGAGGAGT--CGAG	2281	
21560-25369	2109	TTCTAATAACTCTATTGCCATACCTACAAATTTTACTATTAGTGTGACCA	2158	
hub_1489433_G	2282	CCCTTAAACTCAACT-CTAGAAGACCTACTCAAGCAGTTATTACAACAT	2330	
21560-25369	2159	-CTGAAATTCACCTGT---GTCTATGACAA-AGACATCGGTAGACTGT	2202	
hub_1489433_G	2331	GCTACAATACACAAGAACAAGACTTGTACATCAGAACAGGTTGTCTGA	2380	
21560-25369	2203	ACAATGTATAT-TTGTGGTGATTCAACTGAGTGCAGCAACCTTTTGTG-	2250	
hub_1489433_G	2381	AAAAGTTTTCTATTGGGGGAATGAAGCAAATTGAGCCTAAGTTTTCTGGA	2430	
21560-25369	2251	CAATATGCTA--GTTTTGACACAAATAAATCGTGCTTTAACTGGAATA	2298	
hub_1489433_G	2431	CAAAAAGAAAAGGCTGATTTACTCAGTTAA--GT-CTAAGACCAAGAA	2477	
21560-25369	2299	GCTGT-TGAACAGGACAAAAATACTCAAGAAATTTT-TGCTCAAGTTAAA	2346	
hub_1489433_G	2478	TAGTCTGAGAAAAACAAGATGTTACCTGATCTTATATGCACTCTATTAT	2527	
21560-25369	2347	CAAATTATAAGACAC---CACCAATTAAAGATTTTGGTGGTTTCAAT-TT	2393	
hub_1489433_G	2528	TATTTTGTCTTTCATGTCCTTGTAAATAGTGATGGTTTAAATGATCAT	2577	
21560-25369	2394	TTCACR--AATATTACCAGATCCATCAAAACCAAGCAAGGTCATTTAT	2441	
hub_1489433_G	2578	TTCATATAAAAATTAAGAAGTACATTTTTTAACCTCCTGTCAAAAAT	2627	
21560-25369	2442	TGAGGATTTACTTTTCAATAAAGTGACACTTGCT-GATGCTGGCTTCATC	2490	
hub_1489433_G	2629	TCTGCCTAAGGTACTTCTCAACACACACAGTGTAGTTGCTACC--CCTC	2675	
21560-25369	2491	AAACAATATCGGTGATTGCCTTGGTGATATTGCTGCTAGCGGATCTTATTG	2540	
hub_1489433_G	2676	CTTCAR---GGCTCTGTTTCATGCCCGTCTC-CTCCACGAAGACTTTTTTG	2721	
21560-25369	2541	TGCTCAAAAGTTCATGGCCCTTACTGTTCTGCCA-----CCTTTG	2580	
hub_1489433_G	2722	TTCTACACCTAGAAAGGCTCTGCCTACTCAGGCAGTTGTTATTACCTCG	2771	
21560-25369	2581	CTCACAGATGAATGATCGCTCAATACACTTCTGCACTATTAGCAGGTAC	2630	
hub_1489433_G	2772	ATTTCTACTATCAGATCTCTTCGTATTATCTTCTATATGACTAGGTCT	2821	
21560-25369	2631	AATCACTTCTGGTGGACTTTTGGTGCAGGTGCTGCTTTACAAATACCAT	2680	
hub_1489433_G	2822	CTCTCCCCCTCAACCACAATCTCTCTGAGGCTGGAAAT-TTGTGCACA	2870	

Alignment of identified sequence with the RaTG13 bat coronavirus genomic sequence				
21560-25369	2681	TTGCCATGCAAAATGGCTTATAGGTTTAAATGCTATTGCAGTTACACAGAAT	2730	
hub_1489433_G	2871	TTGCCTTGCACATAA-TAAAGGCTCCAGAGGTATCTGTCTAAACTGGCTT	2919	
21560-25369	2731	GTTCTCTATGAGA--ACCAAAAATTTGAT---TGCCAAACAGTTTAAATAGT	2775	
hub_1489433_G	2920	TATTTCCCTTGAGACTACAAGCACTTATCTGTGTGCCAGGCACCTTTTAGGTT	2969	
21560-25369	2776	GCT-----ATTGGCAAAATTCAGACTCACTTCTTCTC--TACAGCAAG	2816	
hub_1489433_G	2970	CCAGGGAAAAAGAGGTACAAAACAGACACAAACCCACCGTTATGGAGC	3019	
21560-25369	2817	TGCACTTGGAAACTTCAAGATGTTG---TCAACCAAAAT--GCACAAGC	2861	
hub_1489433_G	3020	TTACCTTTTAAATTAAGGTGGAAGGATGAACCTTTTTTGGTCTCTC	3069	
21560-25369	2862	TTTAAACACGCT--TGTTAAACA-----ACTTAGCTCCAATT--TTGGA-G	2902	
hub_1489433_G	3070	TAGAAAGTTGCAGCAGGAGACCATAGGAAATAGTATAAAATAGTTGAAAG	3119	
21560-25369	2903	CTATTTCTAGCGTGTAAATGATATCCTT-TCAGCTCTGCACAAAGTTGA	2951	
hub_1489433_G	3120	CACTGTGGAGTGTGAGTCAGGATACCTTGGTCTCATCTCTAATTGATGT	3169	
21560-25369	2952	GGCTGAAGTGCAGATTGACAGCTTGATCACAGGCAGACTTCAAAGCTTGC	3001	
hub_1489433_G	3170	ATCT--TGAGCACATTTCT---TTAAACATTTGGTCATCTGTTTCCCTGTA	3213	
21560-25369	3002	AGACATATGTGACTCAACAATTAATTAGAGCTGCAGAAATCAGAG--CTT	3049	
hub_1489433_G	3214	TGCCATATAGGAATCATATGCTTACTGGGAAAACCTGAA-TCAGAAAACAG	3262	
21560-25369	3050	CTGCCAATCTTGCTG-----CTACTAAATGTGAGAGTGTGT	3086	
hub_1489433_G	3263	ATGCAAAATCATGTTGGAGGGAACCTTCTCAACCTGATAAAAAGCATCTAT	3312	
21560-25369	3087	ACTCGGACAATCAAAAAGAGTGTATTTTGTGGAAAAGGCTATCATCTTA	3136	
hub_1489433_G	3313	GAAAAACCCACAGCTAACACCATACTTAAAGGTGAAAGACTGGAAGCCTT	3362	
21560-25369	3137	TGTCTTTCCCTCAGTCAGCACCTCATGGTGTAGTCTTCTTGCATGTGACA	3186	
hub_1489433_G	3363	CTTCCGAAGATCAGTAA-CAAGACAAGGATGTCTGCTCTCACCCTGCTA	3411	
21560-25369	3187	TATGTCCCTGCACAAGAAAAGAACTTCACAACCTGCTCCCTGCCATTTGICA	3236	
hub_1489433_G	3412	TTCAACATTTACCCGGA---AGTTCTAGCCAGGTTCTAAGTAAGAAAATGA	3459	
21560-25369	3237	TGATGGAAGGCACACTTTCACGCTGAAGGTGTTTTTCG---TTTCAAATG	3283	
hub_1489433_G	3460	AATAAAAAGAAATCAAGATTGGAATGAAGAGTACTAAAATATCTATTT	3509	
21560-25369	3284	GCACACACTGGTTTGTACACAAAGGAATTTTATGAACCAACAAATATT	3333	
hub_1489433_G	3510	TCATATGACATGACCTTACTTAGAAAATGCTAAGAAATCCACCCCAACC	3559	
21560-25369	3334	ACAACA--GACAACACATTTGTCTCTGGTAGCTGTGAT----GTGTAAAT	3377	
hub_1489433_G	3560	CCCACCCCAACAAAACCTATTAAAGCTAATAAGTGAATTCAGCAAGATTTC	3609	
21560-25369	3378	AGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCAAGACTTGATT	3427	
hub_1489433_G	3610	AGGATACAAGGTCAATACGCAAAAAAAGTTGTAT----TTCTATAAA	3655	
21560-25369	3428	CATTCAAGGAGGAGTTGGATAAATACCTTTAAAAATCATACATCACCTGAT	3477	
hub_1489433_G	3656	CTAACAAATGAACAATCTGAAAATGAAATTAATAAACA-ACACCATTATG	3704	
21560-25369	3478	GTAGATTTAGGTGACATTTCTGGCATTAAATGCTTCAGT----TGTCATAA	3523	
hub_1489433_G	3705	ATAGCATTAATAAAGAAATTAAGGAATAAATTTAGCAAGAGTGTAAACAC	3754	
21560-25369	3524	TTCAAAAGGAAATTGACCGCTCAATG-----AGGTTGCCAAAATCTAAA	3569	

Alignment of identified sequence with the RaTG13 bat coronavirus genomic sequence				
hub_1489433_G	3755	TTGTACGTGGAACCAACAAACATTGTTGAAAGAAATCAAAGACCTAAA	3804	
21560-25369	3570	TGAATCTCTCATCGATCT-CCAAGAACTTGGAAAGTATGAACAGTATATA	3618	
hub_1489433_G	3805	TAAAAATTTTAAAAATCCTGCCTTTGTGGATTAGAACACTTAATTTTGTTA	3854	
21560-25369	3619	AAATGGCCATGGTACATTTGGCTAGGTTTATAGCTGGCTTGATTGCCAT	3668	
hub_1489433_G	3855	AAATAGCAGTACTCC---TCAATTTGAATTATTCACAGCAATCCTACAAA	3902	
21560-25369	3669	AATAATGGTCACGATTATGCTT-TGCTGTA--TGACCAGTTGC-TGCAGT	3714	
hub_1489433_G	3903	AACTCTAGCTACCTTTATTTTCTGCAGAAATGACAAGCTGAGTTTAA	3952	
21560-25369	3715	TGTCTCAAGG-----GCTGTTGTTCTTGCAGATCTTGCTGCAAATTTGATG	3760	
hub_1489433_G	3953	TTTTACATGGAATGCAAGGAACCCAGAATATCCAAA-CAATCTTGAAR	4001	
21560-25369	3761	AAGACGA-CTCTGAGCCAGTGGTCAAAGGAGTCAAATTACATTACACA	3807	
hub_1489433_G	4002	AAAAGGAACAAAGTGGGAAGACTCATAC-TTCCTAATTTAAAAACTGA	4048	

[0241] To investigate the nature of the viruses identified by Kraken2 systematically in detail, pipelines that integrate these sequencing reads to identify viral-like sequences with high confidence were developed (FIG. 9A). First, a metagenomic classification method (Kraken2) was employed to detect possible viral sequences. Next, a two-pronged strategy for assembling the RNA-seq into transcripts that can be utilized for viral sequence analysis was used. The first strategy was bottom-up: a de novo assembly (using 4,707,164 of the total) of the RNA-seq reads was performed that classified them as viruses and separated them into putative mammalian or non-mammalian viruses based on the VIRION database and then verified that the respective transcripts map to the bat genome. Additionally, 5 kb flanks per transcript locus within the genome were extracted to determine the extent of each potential viral integration. Using the bat genome as a scaffold, the second method was a “top-down” approach and involved mapping the Kraken2 codified RNA-seq reads to the bat genome and then extracting the respective genomic sequences with or without adding 5 kb flanking regions on each side. Then BLAST was utilized against a mammalian and a non-mammalian virus database to discover viral hits. Importantly, to avoid viral matches by chance, all transcripts or genomic sequences to each database were mapped after randomizing them by dinucleotide shuffling.

[0242] When the pipelines were applied to the bat stem cell transcriptome data, 311 and 82 transcripts estimated to be mammalian viruses and 351 and 58 non-mammalian viruses (bottom-up and top-down, respectively) were obtained. Direct genome mapping yielded 56

hits (out of 63 transcripts, bottom-up; 25 unique) and 82 (all transcripts from top-down approach; 19 unique) mammalian virus hits against the *R. ferrumequinum* genome. After applying the BLAST threshold, 31 transcripts, with 13 transcripts shared between both methods, mapped to both a viral sequence and a locus in the bat genome. The BLAST step on extended sequences from both methods yielded a total of 16 sequences within the *R. ferrumequinum* genome that aligned with known viruses at high confidence. Validating this stringent approach, using the shuffled sequence data, no hits were found for the bottom-up sequences and only two top-down BLAST hits passed the threshold, indicating that the vast majority of the viral hits are not chance matches but reflect bona fide homology. Indeed, this was confirmed by manual inspection of the alignment hits, which showed numerous longer, well-aligning regions substantially exceeding the length and quality of the matches of randomized sequences. The results indicated a taxonomically diverse collection of attributed viruses from a number of major viral families. Included among them are Flaviviridae, Herpesviridae, Poxviridae and Retroviridae. Overall, this exhaustive analysis shows that bat stem cells contain a surprising diversity of sequences that resemble viral genomes. To implement an orthogonal metagenomic strategy, a direct alignment method using the Microsoft Research Premonition pipeline was employed. Using bat stem cell RNA-seq reads as input, this classifier positively recognized 419 different putative viral-like sequences. Again, the taxonomy included a number of important viral families, such as Paramyxoviridae, Flaviviridae, Retroviridae, Coronaviridae and Poxviridae. Manual examination of the expressed virus-sequence revealed a wide range of lengths ranging from (near) full-length viral sequences to specific viral protein encoding domains to short fragments of viral regulatory sequences. As before, the Premonition pipeline predicted sequences were mapped to the bat genome, extended 5000 bp flanks, and performed BLAST searches against the VirusDB and showed that a total of 13 extended bat genome sequences mapped to known virus genomes, 9 of which overlapped with the bottom-up/top-down approaches, indicating a high degree of consistency. Viruses linked to Hardy-Zuckerman 4 feline sarcoma virus, Friend murine leukemia virus, Porcine endogenous retrovirus E, and PreXMRV-1 provirus were examples. Consequently, both metagenomics pipelines methods reveal a significant number of endogenized sequences that resemble viral genomes with a final count of 20 high-confidence viral hits across all methods. Exemplary sequences of possible viral origin discovered with this method are listed in SEQ ID NOs: 1-349.

EXAMPLE 11 Identification of viral proteins useful in vaccine development

[0243] This example describes the identification of viral nucleic acid sequences and viral proteins present in the bat genome and in bat cells for the use in vaccine development.

[0244] Briefly, viral DNA and RNA sequences can be identified as described in **Example 8**, **Example 9**, and **Example 10**. The viral DNA or RNA sequences can be assembled into long contigs such as **SEQ ID NO: 1-349**. The contigs can be translated into amino acid sequences. The identified amino acid sequences can be compared to known nucleic acid sequences and proteins using methods like BLAST (www.web.expasy.org/blast) and the sequences can be aligned and translated into amino acid sequences of peptides and proteins. Vital viral enzymes such as the essential genes are replicase ORF1ab, spike (S), envelope (E), membrane (M) and nucleocapsid (N), RNA polymerases, kinases, and viral proteases can be identified using homology models and sequence alignment as described in **Example 10**.

[0245] In order to develop a vaccine, immunogenic CD8⁺ T cell epitopes in the identified viral virus proteins can be predicted using for example a machine learning platform such as described in Bulik-Sullivan et al. (2018) Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nature Biotechnology* 2018, 37(1). Predictions for these epitopes can be run for each HLA class I allele. Candidate CD8⁺ epitopes can be maximized for coverage of the prevalent HLA-types in a given population. The method described for generating candidate CD8/MHC class I epitopes can be used to generate peptides with sizes between 9 and 20 amino acids. Further, potential HLA-DRB, HLA-DQ, and HLA-DP MHC class II epitopes can be predicted. The predicted epitopes can then be displayed by MHCs and recognized by human T cells can be tested with methods such as mass spectrometry based HLA I and HLA II epitope binding prediction tools (*e.g.*, Immune Epitope Database and Analysis Resource, www.iedb.org). Epitopes such as for HLA-I or HLA-II can be scored and identified for peptide sequences derived from the identified viral viral enzyme. Top-ranking peptides can be prioritized based on expected population coverage (allele frequencies). Predicted peptides can be tested for T cell responses using PBMCs from human donors and MHC multimers loaded with peptides and ranked. Further assays of T cell reactivity (*e.g.*, interferon-gamma ELISpots, tetramers), which are stricter measures for T cell immunogenicity to epitopes, can be performed to further identify top immunogenic peptides.

[0246] The nucleotide sequences for the identified epitopes and peptides can be cloned into vectors with expression cassettes in order to express viral proteins for use in vaccines in

recombinant cell. Recombinant cells for example HEK cells or CHO cells can be transfected with these vectors to produce vaccines, such as adenovirus based vaccines. mRNA based vaccines can be synthesized chemically or enzymatically and packaged into lipid particles, nanoparticles or liposomes for further delivery to a subject.

References

- [0247] Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11(10):R106. doi: 10.1186/gb-2010-11-10-r106. Epub 2010 Oct 27. PMID: 20979621; PMCID: PMC3218662.
- [0248] Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>.
- [0249] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.
- [0250] Carlson CJ, Gibb RJ, Albery GF, Brierley L, Connor RP, Dallas TA, Eskew EA, Fagre AC, Farrell MJ, Frank HK, Muylaert RL, Poisot T, Rasmussen AL, Ryan SJ, Seifert SN. The Global Virome in One Network (VIRION): an Atlas of Vertebrate-Virus Associations. *mBio.* 2022 Apr 26;13(2):e0298521. doi: 10.1128/mbio.02985-21. Epub 2022 Mar 1. PMID: 35229639; PMCID: PMC8941870.
- [0251] Carter AC, Davis-Dusenbery BN, Koszka K, Ichida JK, Eggan K. Nanog-independent reprogramming to iPSCs with canonical factors. *Stem Cell Reports.* 2014 Jan 31;2(2):119-26. doi: 10.1016/j.stemcr.2013.12.010. PMID: 24527385; PMCID: PMC3923195.
- [0252] Dejosez M, Krumenacker JS, Zitour LJ, Passeri M, Chu LF, Songyang Z, Thomson JA, Zwaka TP. Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell.* 2008 Jun 27;133(7):1162-74. doi: 10.1016/j.cell.2008.05.047.
- [0253] Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016 Oct 1;32(19):3047-8. doi: 10.1093/bioinformatics/btw354. Epub 2016 Jun 16. PMID: 27312411; PMCID: PMC5039924.

- [0254] Huang Z, Whelan CV, Foley NM, Jebb D, Touzalin F, Petit EJ, Puechmaille SJ, Teeling EC. Longitudinal comparative transcriptomics reveals unique mechanisms underlying extended healthspan in bats. *Nat Ecol Evol*. 2019 Jul;3(7):1110-1120. doi: 10.1038/s41559-019-0913-3. Epub 2019 Jun 10. PMID: 31182815.
- [0255] Jebb D, Huang Z, Pippel M, Hughes GM, Lavrichenko K, Devanna P, Winkler S, Jermin LS, Skirmuntt EC, Katzourakis A, Burkitt-Gray L, Ray DA, Sullivan KAM, Roscito JG, Kirilenko BM, Dávalos LM, Corthals AP, Power ML, Jones G, Ransome RD, Dechmann DKN, Locatelli AG, Puechmaille SJ, Fedrigo O, Jarvis ED, Hiller M, Vernes SC, Myers EW, Teeling EC. Six reference-quality genomes reveal evolution of bat adaptations. *Nature*. 2020 Jul;583(7817):578-584. doi: 10.1038/s41586-020-2486-3. Epub 2020 Jul 22. PMID: 32699395; PMCID: PMC8075899.
- [0256] Kacprzyk J, Locatelli AG, Hughes GM, Huang Z, Clarke M, Gorbunova V, Sacchi C, Stewart GS, Teeling EC. Evolution of mammalian longevity: age-related increase in autophagy in bats compared to other mammals. *Aging (Albany NY)*. 2021 Mar 21;13(6):7998-8025. doi: 10.18632/aging.202852. Epub 2021 Mar 21. PMID: 33744862; PMCID: PMC8034928.
- [0257] Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019 Aug;37(8):907-915. doi: 10.1038/s41587-019-0201-4. Epub 2019 Aug 2. PMID: 31375807; PMCID: PMC7605509.
- [0258] Knaupp AS, Buckberry S, Pflueger J, Lim SM, Ford E, Larcombe MR, Rossello FJ, de Mendoza A, Alaei S, Firas J, Holmes ML, Nair SS, Clark SJ, Nefzger CM, Lister R, Polo JM. Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell*. 2017 Dec 7;21(6):834-845.e6. doi: 10.1016/j.stem.2017.11.007. PMID: 29220667.
- [0259] Krueger, F. (2012). A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisulfite-Seq) libraries. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

- [0260] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. Epub 2009 Jun 8. PMID: 19505943; PMCID: PMC2723002.
- [0261] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014 Apr 1;30(7):923-30. doi: 10.1093/bioinformatics/btt656. Epub 2013 Nov 13. PMID: 24227677.
- [0262] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PMID: 25516281; PMCID: PMC4302049.
- [0263] Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res*. 2015 Jul 1;43(W1):W566-70. doi: 10.1093/nar/gkv468. Epub 2015 May 12. PMID: 25969447; PMCID: PMC4489295.
- [0264] Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016 Jul 8;44(W1):W160-5. doi: 10.1093/nar/gkw257. Epub 2016 Apr 13. PMID: 27079975; PMCID: PMC4987876.
- [0265] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011 Jan;29(1):24-6. doi: 10.1038/nbt.1754. PMID: 21221095; PMCID: PMC3346182.
- [0266] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303. PMID: 14597658; PMCID: PMC403769.
- [0267] Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4; Available online at: <https://ggplot2.tidyverse.org>.

- [0268] Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019 Nov 28;20(1):257. doi: 10.1186/s13059-019-1891-0. PMID: 31779668; PMCID: PMC6883579.
- [0269] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007 Aug;24(8):1586-91. doi: 10.1093/molbev/msm088. Epub 2007 May 4. PMID: 17483113.
- [0270] Yoshimatsu S, Nakajima M, Iguchi A, Sanosaka T, Sato T, Nakamura M, Nakajima R, Arai E, Ishikawa M, Imaizumi K, Watanabe H, Okahara J, Noce T, Takeda Y, Sasaki E, Behr R, Edamura K, Shiozawa S, Okano H. Non-viral Induction of Transgene-free iPSCs from Somatic Fibroblasts of Multiple Mammalian Species. *Stem Cell Reports.* 2021 Apr 13;16(4):754-770. doi: 10.1016/j.stemcr.2021.03.002. Epub 2021 Apr 1. PMID: 33798453; PMCID: PMC8072067.
- [0271] Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, Jeon M, Ma'ayan A. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc.* 2021 Mar;1(3):e90. doi: 10.1002/cpz1.90. PMID: 33780170; PMCID: PMC8152575.
- [0272] Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137. doi: 10.1186/gb-2008-9-9-r137. Epub 2008 Sep 17. PMID: 18798982; PMCID: PMC2592715.

EQUIVALENTS / OTHER EMBODIMENTS

[0273] While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure that come within known or customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

SEQUENCE LISTING

SEQ ID NO:	Sequence
1	RFe-V-MD1 GGAGAGAATTGATCAGAACTCCTGTCTGTCTCCGGTCTTTGTGTCTCCCATTTTCCTCCCTTCTAGGTG CCTTCGGGGTCCCTCGTGTAGTGTCCCGCGGGTCGGGACAACTGGCGCCCCAACGTGGGGCCTGAAGTCTCC TAGAAGACGAGACGCCTGAGTTCGTCCGGTCTAAGGAGCTGCAGCATATTTCTTCTTTGATCACCATAAG ACTACCCAACTTGTGGGAGATCTGTACAGGTAAGCGGACGACTCCTTCAAAAAAATGGGACATATATTTG TGTCTAGACTTATGTATAGTCTACAGGCTTCCCTCAGACACTTAGACTAGGGTTCCCTAACCTGTTG TCCAGTCTCCCTTTTATCTGCTCTCAGCTCACTTTGGGTTTTAGTCGTTCAACACGAGACAGTTTTC TAGGTGTTTGGGACCGTTTGGCGAGATTTTGCCTGCTTACTTTGAGCTCCAATCGTCCACCCAGAGGAT TTCCCGACCGGTTGAGTCCCGACTGGCTTTGCCTGAGGGTCGTTACCAGCCGCGTGCCTCTCGGGATC CGTGTGGCGGATTATACCAACCGATTGCTCAGCTAAGGGCTTTTCTCCTCTCACCCCAACACCCCGT GGCTCCGGCCGGGTGAGTCCCAAGACATTGCTGCGGGTCGTTACCATCCGTGCCGTCTCGTTTGGG TCCATGTTGGTGGATTGTACCAACCGACTGCCATGTGAGGAGAGTCTTTATTCCTTATCATAATGGGAC AAGAGGTTAGTGTTCATGACATGTTTATCTCAGGACTAAAAGAGTCTTACAAATAAGGAGAGTTAAAGT CAAGAAAAAGATTAGTTATCTTTTAAATTTCTTAAAGATGTTTGGCCTTGGCTCCCTCAGGAAGGA ACCATAGACCACAAAAGATGGAAAAGGATCGGAGATGCCCTTAATGACTTTTATAAACTTTTGGCCCTA AAAAGAATCCCGATCACTGCTTTCACCTATTGGAATTGCATTATGAGCTACTTATGGTACATCGCTACA CCCTGACATCGACCGAGTGATACAAGAAGGAAACACATTTTACAAAACGCTTCCCGCCCCCTCCTCCTC CTTACAGGTCCCTCTTCTAAGTCTCTCACGATTGAGATTCTATTTCTATTTCAATGCCTCCTGAAGAT CCTGAGACCACCAAAAAGATCCTAGTAAGCCTTATATCCTCCCTACCACCTAATGTCTCTGATCTTAA TGTAATTTAGCCACCTGAGGACGATCAGTTAAGCCCTGAGGACGAGGCTGATTTAGAGGAAGCTGCC GCTAAATATCATAATCCTGTCTGGCAGTTTCTGGCCTCTAATCAATTGCCCTCCCTATAATCCCCAAA TGCTTTAGCTCCTATCCAGATCCTGATCAAACTCTCCTCTCCCACCAAGTCCAACAATTACAAGAAC TGTTCAACTCAAAAACAACATCTAACTCTCCTTAAACAACCTCAACAATTAGATTTACAACCTCTCCTCT GCTGCTACTCAAAAATTTCCCCCCCCCTTCCATAAATCCTACAAAACATTTCCCCTCTCAATAAAAAA AACCTTATTAATCTTTTCCCGTTATTGAATTCGCCCAATAAAAACCTGAAGGAGCGAGTGAGATGAT GATAAAGACCCCGACAGAGACAATATAGAACCCCGCAAGACACTATAAACGCCTTGACTTAAAAACCACA AAAGAACTCAAAAAGCGGTGGACGAATATGGCCCCACGGCCCCCTTACACTCTCAATTTTACAATCCC TAGATGACCTCTGGTTAACCACCATGATTGGCACTATTGCGCCATGCCACCTATCGGGGGGCGATTA TGTTCTCTGGAAATCTGAGTTTCTGAGGCCGTAAAGAACTGCACACCGCAACGAGAAGCGGGAGGC GAGTGCACTGATTCGACCTATGATAAGTTGAGGGCTTTAAGCCCTACGATACAAATGAAGCTCAACTAC AATATCCATCTGGCCTTTTTCTCAAATTTACCTTGCCGCTACTAAGGCATGGAAAAAATCTCTCCCTA AGGGGCCGGCCACAACCTCAACTCACTAGTATTAGACAGAGGCCAGAGGAACCTTATGCTGACTTCATCAG TCGCCTAACCAATGCCACTGAAAGACTCCTTGGTAGCAGAACTGATAGTGATTTTTTCAACAATTA GCTTTTGAATAACCAATTTGCTGTGAGGCAGCCATCTGCCCTAGAAAAAAGGATTCACTCTTGATT ACATTGCGCTATGCACTGATATTTGGTCTGTCACCAATGGGCCTCGCTATCGGGGCGACTTTTAAAG ATTCACTTACTTAATCTGTCTAAAGGCAAAAACAATGTTTTTCTGTGGCCAGCCCGGACATTTGCCAA ACAATGCCCAACCCCTCGCCAGAACACCATTAGGCCAACCCACTCCACACCCATATTGCCCCGCGAGT ATGTCCCAGATGCAAGAGAGACAAACATTGGGCCAATCAATGTAGATCAAAAATAGATGCCCAACAAT CCTCTCCTGCCCCAGCAGGGGAACTTCTGAGGGGCCAGCCCCAGGCCCTACAGGAGAATCCAAACCT TGGGGCGACTCGGTTTGTCTATCCACAACAACTTTGTCCCATCTCAAGTCTCCTCCGAGCAACCCCTG GCAGTGTGGACTGGACCTCAGTCCCTCCTCCAAATCAATATTAACCTCCTGACATGGGACCTCAGATA CTACCTACGGGTGTACCGGACCCCTACCAACCAACACTTTTGGTCTAAAATTTGGAAGAGGTAGTTTCGAG CCTACAAGGCTATATATTTACCCTGGTGTATAGATAATGATTTTACGGGAGAAATACAGATTGTAGCC TCCCTCACTTCTCTCATTTCTATACAACCGGGACAGAGAATAGCTCAACTACTCCTCTCCCCTCC AGACCACCCATAAATCTGCCAACAATGAGCCTAGAAACAACAAAAATTTTAGATCCTCAGATGCTTATTG GATTCAAAATCTCTCCCCAATAAGCCCATGTAGATTTAAAACTTGATGGAAAAACCTTTTAAAGGCCT TATCGCACTGGTGTGATGCAACCATATTAGACAAAAAGACTGGCCGCTTTCTTGGCCCCCTTTCTGA CACACTTACTCACCTACAAGGCATAGGACAAACAATAACCCAGACAAAGTGCCAAAGTTCTTAACATGG CTAGATAAAGAAAATAACTCTGGCACAGTACAACCTTACGTTGTACCAACCCCTCCAGTAAATCTGTGG GGCCGTGACATATTATCCCAATGGGAGTAATCATGTTGAGCCCCAATTCGAAGATAACCATCCAGATGT TAAACAAGGGTTTCTCCAGGTGAGGATTAGAAAAACAAGGACAGGGAATAAAAAACCCCTGTCTAC TGCTTCACTGGCTGCTTCGATTAGGCTTAGGACATTTTCACTAGTGGCTCTGACCACTGCACCCCA TGTGACCTATATCTGGAAAGGACAACCTCGCCCATATGGGTGGATCAGTGGCCACTAAATTCAGAAAA ACTAAATGCTGCCAATCAGTTAGTGCAGAAACAATGGCGGCAGGGCATCTAGAGCCAGTAACCTCCCC CTGGAACACACCTATCTTTGTCGTAAAAAGAAATCTGGAATTTGGAGACTTCTCCAAGACCATAGGGAAG

	TCAATAAAACAATGATAATTATGGGCGCCCTTCAACCAGGCCTACCTACCCCTGGAGCTATTCCCTCGG GGATCCTTAAAAATCATTATGATCTCAAAGACTGCTTCTTCACTATCCCTCTACACCCTCAAGATAGAC AATGTTTTTGGCTTTCAGCATACCTATAACTAATTTCCAAGGGCCCATGCAGAGATTTCAAGTGAAGGTCT TACCTCAGGGGCATGGCCAACAGCCGACACTGTCAAATATTTGTTTGTCTGGCCATCGATCCCATTCG AACTCAGTGGCCCTCTCTTTATATTATTCATTATATGGATGATATCTTAATAGCTGGCAAGAATGGGTCT GTACTTCCCTCTCCCAATATAAAACAAGAAAAACCTCAGCCTTGTCCCGCTAAATGCTCTACTATTTACCC TATTATTTCATAGTTCTTGTTACAATACCTATAAAACATGTACAGAAAAGATAAATCCTCTTATATACGG CTGTCTGACAAGCACTGGTCCCGCTGTCCCTCATTCTGACTGGTCTAACACCCCTGCTGCGGTTGGCAT TTGGCTCCCATAAACCCGACCCCTGCGCGGCATCTAATATGTTAGAAAAAATATTTGCTGGGCAGATCG AATCCCCCTATACCATATGTTTCTGACGGCGGGGGTCCAGCCGATCTCCAATCCAATGAAAAACGCATTA AAAAATTTGCTAAATACAAAAGACCCCTTAACCTAAATTTACCTATCACCCCTTTGGCCACCCCTAAAAAA CCGGGGTCCAGTGGACATTCCTCAGACTTTTGACATTTCTAGTTCTACCCACAGTTATGCTTTCT GTTAATTCATCCTACGCCAGAGACTGCTGGCTGTGTTTACTACAAGGTACCCCTTTACCATTAGCTATAC CCTATCCCTTTGTACCTCTGACTACCAATAATTCATACAACATAGCTCTCCCTTTTATAGTCCAACC CCTTGGCTTTAACAATACCCCGTGCATCCTCTCTCCCATTCAAAACAATACTACAGAGGTTATATTTAGG AAGCCTCTCCTTTTACAAATGCTCCTCTTCAATTAATGTATCCTCTCCTATGTGTACACCCCAATGGATCG GTATATATTTGTGGAATAATTTATGGCCTACACCTATTTACCACAAAACCTGGACAGGAGTTTGTACCC TAGGCTCCCTCCTCCAGATGTATCCATCATTCCAGGAGATGAGCCAGTCCCTATCCCGACTTTCGAACA TATTGCAGGACGCACTAAACGTGCAGTCCATTTTATTCCCTTATAGCGGGTCTAGACATCACACGACACA CTTGCCACCGGGTCCGCGGGGATAGGAACATTCCTAGTACAATACCATAAATTATCTGGACAACCTCAT ATCAGATGTCAGGTACTCTCAGAACTAATCCAAGATCTTCAAGATCAGGTTGATTCCCTCAGCAAGAT TGTCCTCCAAAACAGGAGGGGATTAGATTTACTTACTGCAAAAAAGGGGCATCTGTCTGGCCCTCGG AGAAAAATGCTGTTTTTATGCTAACAAATCTGGAATTTGTTGTAACAGAGTCAAAAAAATACAAAAA GACTTGAAAAAAGAAGGGACCTCCTTTCCAACCTCTCTGGACCGGATTCAATGGACTTTTACCCTACT TACTACCCCTGCTTGGCCCCATACCTCGGGTCTTTATCCTACTATCACTGGGACCACATCCCTCCTCAA TAAACTCATGCGCTTTCTCAGACAACAAATAGAGGCCCTTGACGGCCAAGCCCATACAGGTCCATTACACC CGACGGGAGATGCAAGAGCGAGGAGATCCCTATCTCCAATAACAGGAGTCATAAAACAGGACTCCTCCC CTGTGAGATGAACTGGATAGCCAATGACGGGTAAAGAGGACAGCTCTCTAAGTAAACATTAATAAATCAAAA ACCTGTGCTGTACAGGTTTACAGAGATGGACTGTCCCAACCTAAGACAGGCACAGTTCCCTAGGTGG CTCAGAGCTCTTTTTTATAAAACAGAAACGGGGGGACCTGTAGTGGGCGGGTGCTGTAAAGGCACCAATC ACATGACTGAGAAGCATGAGATAGAGGAAGTTACTTGGGTCTTTAGATAACACCCACATTCTGTAAGGTA TGTCAGAGGGCTTAAGACCATCAGCCTGCGGCAACCCCTGCTTATGTTAATGCCCTCCACCAGCACAA AAATGTATAATAACCCATGATTGAGCTGCAATAAAGAGAGACTTGATC
2	RFe-V-MD2 GGAGACCTCGTCGCGCAGCGGAGCGGTGCACCAGCCGGTCTTCTGTTACTAAAGGACTCAGGTGGAGGTA GGTGTGCGTTGGGCGCTGATACTCGAGCTTGTGTGACCGGACTGCTTTTAAAGAAATAGACATTTACACA CATATATAATTTAAAAAGCAAACAACATTTTCAGGATGCATTACGTACCTTTATTGCTGTCTCTGCACT CTATTCAAGTGTCTGTTCTTTGTCTAGTTTAAAAATGTTGGTCTGACTCACTGTATGCTTTTCACT CTCAGATGGGTGCGAACACACATTTTAAAAAATGCTGTAAGAATCCGGGAAGTGGGTGGTACCACGTTT GACCGACTAGTGCCCCGTGTATACCTGCGTCAAACAGCACGTAGGTGTGAATGAGCCCAAGACCGGTCTC ACTGTGCTGTTGGCAGAAAAGAAATCCTTGGCAGTTTCTGACAAACTAAACAAAAAGGATGAAATTCAC AGAAAAATTAAGTTATAGCCCTGCCTTAGTTATGTATCTTTTGCACAATGACTAGGACTTTGGTAATAA CCTGTTTTGTTTCAACTTGAAAAATGCATAATGAATATCGTAGTATGTCATCAATAAATATTTCATGTATA ACATACCTTTTCAGTGACAGCAAAAGTTTGCATCCTACTGATGGACATTTTAAAAAGAAAAATATTTACTG AAGTTTAAACAATTACAAAAAGCATATGAAAGTGAACAACCTCAATATATTTACACAAAGCAAGCAGACC CACGTACCTAGCACCCACTGTGAGAACCAAAATCATTACCAGAATCCCAGAGACTCCTGCCAAAGGTAGT GGGACCTCCAGTCACCTACCTTCCAAGCGTAATAATTATCCTGATTTCTACCACGGTATTAGTTTCACT ATCCCTTCAGACCAAGGCTGTCTCCCATAAACCACTGAATTTCTTTTGTGCAACCACTTTTCTCTCCCTC TCCTCTCTCCCTTCTTATCCCTCTTCTCTCTTTCTCTGTTTAGGAGACCTGATTTCTCCATTGCAAAAA GTATTTTGGCCAACTTCTGTTTACCTGGAGGTCTGTCTTCTTGGCAAAGTTACTTTCTTGTCTTTGTA CAACAGGCAACTGTCTCTGTATCCTTCTTATCTGGAAGTGAAGAGAGTTAGAGTCTGTAGTCTGT GGCCGAGTGGTTAAGGCGATGGACTAGAAATCCATTGGGGTCTCCCCGCGCAGGTTGCAATCCTGCCGAC TACGGGGTTCTTTTCTTCCCGAACCGCGAGTGACTCGGCAAAACCCGTGGCTGAACCTTGCCGGGCCAGA GCTCCAGCGACGGGGAGGGAGGTTCCGCGAGGAGCATGGCCAGTTTCTGTCGCTCCTTCTTTTAGGA CAGCTCTTCGTGAATTTTCTCCCTATGATAAAGGGCTGCGGTCCCTGGGTGCGAGTCTCGGGTCAGCGA GAGATTTCAAGGGATCAGTGGGCCAGCAGCATCCTCGTTAGTATAGTGGTGAGTATCCCCGCTGTCA CGCGGGAGACGGGGTTCGATTTCCCGACGGGGAGGCAGTATGTTTGTGTTTGCACCTCAGTCACCTGTTT TGGAGTTCTCTGGAGACTCTGTGGTCCCTGCTAAGGACATGAATGCTACAGAGCTCTGTGTGGGTGCCACA GGTTCTGTGGGTCTTCCCTTGCAGCTCTCGGCGACCGCCCTGCAAGGCTCTGGGGACTGAATGGCAG GGGACCTTCTGTGAGCTCTTTTCAACTTGACCCTGCCCCCTGCCAGGCTTGTGCCACTCCCCGTTCTGC

	CGCTCTCTGATCAGAGAAACACTTCAGAGCGACTCTAACTACCAAACCTAGAGGGGAACCTAGGTTTT AAGTGACGACGAGACTTAGAACACTTACTGAGACTTAGTAAGAGTGTGGTTGTCTGCACGCGCCTCCCATT TGCAGAAAAGACCACTGGGGGCAATGTGCGAGATGGCAAAAAAATCCACGTGGGTCTTCAGGCCCTCCT TCCTCCTAGAGGTCACCTGGGAATGGGGACCGCCACAGGCTCAGCTGGGGCTCTTTACTCCATCCTGGG CAACTGCTGCCCTAGGCTCTTGACCCCAAGTGTGTAGGAAGGTGGTTAAGTGGTCTCGGACCTGTGG GAACAGGAGGCTCCAAGTTCAGGATACTGCTTTCAACAAGATCTGAAGCTCCTAGCAGTGTGCTTTTG AGTGTATGTTAGACTTTATGAACTAAAGCTTTCTGAAAGGAAAAAACCACCTGTTATAAGCCATGG CAGTCGAGACAGTGTGGCCCTTACTCAGGAATGGATACTAAACGGATGGAAACAGAACGCATCCTAAACA GATCCACTCATACAGCCATTTGGTTTTAAACAAAGGTGATGCCGCAATGCACTAGGGAAGACCGTTCTT TTCAATAAATTGAAATCAATAAATTGGTGGTTCAATTGGATATCAATATGGAAATAAATGAATTACAC ATACCCAACTCAGTCACACGGAAGTATATTTAAACATCAAAGGGAAGCAATAATGTTTTCTGAAAGGT AACAGGATAAATTTCTTATGACTTTGGAGTATGCAAGAATTTCTAAACAGCACAAAAGCAGCTCGTCAC AAAAGATAAGATATATGTATACATTACACTTCACCAATATTGGAACTTTTGTTCATGACTAGCCACCAG TAAGCAAGTACAAGGCAAAATGTAGAGCAGGTGTTGTATTACATGTACCTAATAAGAGACTGTGTCCCT AGACAGAGTTCTCCAGAGAAACAGAACCAATAAGAGGTATGCGTATGTAACAAGAGATCTGTTTTGAGGA ATTGGCTCACGCCATTCAATTCACCAATGTTTTGTGGCTTTCAGAGTATAACTTTTATCTATTTTGTCT AAATTTATCTCTATTTTATTTTGTCTATGATTTTAAATGGAAGTATTTACTTTTGTCTTTTCTTTTC CTGTGAAACATTAGGAGGCTGACACCTCCAGATGCAAGTATGAAGTGTGAAAGATAGCAGGGATTAAT GTCCGCTAGGAGGATACTCCATAAATGCAAGAAATATAGCCACACAGGGAGAGTTTGAAAAAAT GCTTCAGACTCATAGGATAATGGCACAGATAAAGTGAAGCATAACATAAATTGAAATGTGCACTGTTT AGCTGGCTAGGACTGAAGATGCTGATTGGAAGAAAGTGTGATCCATGTCTTTCCATGTACAAGATGCA GCTCATGGAACCTCGACCTTAAAGTGGTGCCTGTTTGTCTCAGAAGCAACAAGATAGAG
3	RFe-V-MD3 GAGAATTGGAGATGGCGGCGGCGCAGGGAACCTTCGAGGAACCGGCGGTTTCAGAACAGCCCGCTGAGCT GACTGCCCTCCGTGCGGGCGAGCATCGAGCGGAAGCGGCGGCACTGATGCTGCCACAGGCCCGGCTG GCGGCGCGGCCCTACCCGACGACGGAGGTTCGCGGCTACCGGAGGTTGGGGCCCTGGCGGCGCTGCCCT GCCCTTCTCCGGCGGCGGCGGCGGTCGCGCGTCCCGTGTGTGGCGTCTACGCCTCCGGACTCCAGCCCC GGGCTTTCTCACTGCACCTGGGCGGTCCAGCTGCGGTCTTTAGCTTTGGGGGTGCAGCCCCCTCTCCGT CTGGAGGTGCCCACTAGTGCCCGTCCGCGCGGCGAGCTCTCCCTTCTGTTCTCTCCGATAGCCTCCACC ATTCCCAGAGATGATGCTTGCAAAAACCTTTAGACCTGTAACCCATCTCAGTAATCTGCACCCGCTCT TCTTTCGTCTCAGAGGGCACATTCGGATCCAGCACAAATGCTTGCCACGCGCAAGGCACCAAGAGGAGC AGAGAGACAGTAGCCACCGCCTTCGCGGGGCTCACAGAGTAGCCTCTGTTGTGCTTCATATGTTTGATTC TCGGAGCTAACCTGGAAAATTAGGGCAGGGTTTGGTATCCGTGTTGGTGAGGTGGTGGTGGCGACAAGA AAAACGGGGTTTGCTTAGGTCCGTCTCAGTAAGTGCAAGGCTAATCAGGACTCGAATCGGGTCATCCG ACACTGGGTTCAGGGCCTTTCTTGCCACCAGCTGCCCTGCTACACAAAGCACCTCTCCTACCCTTAGG AAGAAAGGTGTTATTTGTCTGGATTTTCTCTCTCTCTCTTAGGGTAGCTCTTCGCTGCGTATCTGTC GTGTATGTATTAATATGTGAATTTCTCCACTGTGGTCAAATAATAATCTTCCCAGGGTGCCTAAAAAT AGTTTGGGTCTTCAGGGCTAGCTCTATAACGTGAAGTACATGTGTTCTTAAAGCTAATCCCATACTGTG GAGTAGTTGAGCACAGTTTAAAGCTGTGTTATCTACTATCTTTTGAACAGTCAGAGTAAGGAAGAGTG ACAGTCTGGGTCTGACTGCGTGTCTTGATATTGATACCTGAATCTGCAAATTCAGCCACCTTTAATA ATTCTGGTCTTGTCTTATTGCTTGTGTGTGTATGTTTAAATCTCTTTTTCAGCTTGAGGCATTCTAG AGTCAGGAGAAAAAGTTGTTCAATTTGCATTGATTAATATTTATGATTCTATAAAGGATTCTAGATCTGTA CAGACAGTCCCAACTTACAGTGATTTGACTTACGGTGGTGTGAATGTTATTTCAGTAGAAACCATACTTT GAATTTTGATCTTTCTGGGATAGCCATATGTAGTACTATACTCTTGGGATGCTGAGCCACAGCTCCCT GCTAGCCACGTGATCATGTGGGTAAACAACCGATACTCTACAGTATAGTATTAATGCATTTTCTTTTTT TAATGTTGTAAACATTAATAATATATAGAGCAGAGATGTGATTTCAAAAAACACAGTCATAAACAGAAC AAAATGTATTGGATGAAAAAAGACAGTGCGCATTTGGGAAGGGTGATAGTGGAAAACTATTTAACACAT CATTAATGCATTTTGACTTAAAAAATTTCTATTTATGATAGGTTTCTCTGGATGTAACCCATTATA AACTGAGGAGCATTTGTACTAAATGTAGAATGGATGCAAAATAGAGTATAAACTAGTATTAACTTCTGG TCATGGAAGCAAGGTAGAATGAATATTCTGTAAGATTCTTAGGCAGTTACCCAAGAGTGAAGTGTGT TGAGTATTGCATACAACCCGCTGTGCTTTTAAAGCTTAGGTAGGTACTGAGATTTTATCTTCGCAGTA GTTTTATTTCAATGTACTGTACAATTTTCCATTTTCTGTATGTGCTCTGACATACACCATGAAAAAGATG GGGAAGAACTTGCTTAGAATGTGGTGCTAAGAAGTGGTGTGAGGGCCTGGTGAAACAGCAAGGCATAGC AGCTGAGAAAACTGGCATGATTTAGCATTTGTTTCAGGATCTTGCTCTAGTTTCAGCCTTGACTACTTTAG CTTCCCCTCTTCTAATTTCTCATTGCACTCTTGGTCATTCCAGTTATGTGCTACACGATTCTGAAATCA ATATCATTTGGTATATTTATTTGATTTCTATCCATCCAGTAGATATTTCATGGAATGTTTAACTATCAGAA TTACAGAGATAAAACACTCAGTCTAATGGATGATATACAGCCACCACTTCCGGAACCTTAGAAGTTTCC CTAAAGCCACGTTTTAGTCAATCAGCAACCTCAGACATAACTACTGTTCTAACCATTGATTAGTAATA GTATCTTTTTTTGAACCTCATGTAATGGAATCATACAGTGCCTGGATAGTTTTGCTCAGCATAATATCT GCCAGATTCATCCATGTTGTTGCATGTTTGGTAGTTTATTTATATGCTATATAGTTATTTTTTTGTAT

TATACCACAATTCTCCATTTTTCCTTTTGGTGGATGTTTGGGTTGTTTGCAGTTTGGAGCTATCATGAA GAAAACTTTTGTGAACATTCTTTTAAAAATTTTCAATTACATTTGACACACAGTATTAGTTTCAGGTGTAT ATCATAGTGATTAGACATTTATACAACCTTACAAGTGATCACTCTGATTAAGTCTTTGTAGCCATCTGACAC CATACATAGTTATTATAATATTATTGACTATATTTCTTTTCCCATGACTGTTTATAAATTGGCAATTTGTAT CTTCTTAATCTCTTCACCATTTCATCCATTCCTCCACCCCTCCCATCTGGCAGCCATTGAGTTTGT CTCTATATCTATGAGTTTGTGTTTGTTCGTTTATCTGTTTTTTAGATTCCACATTTAAGTGAAATC ACATGGTATTTGTCTTTCTCTGTTTGACATTTCACTTAGTATAATATCCACTAGGTTCCATCCATGTGACA AATGACAAGATTTTGTGTTTTTATAGCTGAGTAATATTCCATTGTATACATATACCACATCTTCTTTGTGT ATTGCTCTGTGAGTGAACTTTGGTTACTTCCATATCTTGGCTGTTGTAATAATGCTGCAGTGAACATAG GGGTGTGTATATCTTTTCGAATTAGTATTTTGGATTTTTTTTTCAGATAAATACCCAGAAAGTGGAAATGCTG GGTCATATGGTAATCTATTTTAAATTTTTTGGGGACCTCCATACTGTTTTCCGTAGTGGCTGCACCAA TTTACAAGGTGCTTTTCTCTACATCCTTGCCAACACTTGTGTTTATTGATTTATTGATGATGGCCATTCTC TGACACGTGTGACATGATAGCTCATTGTGGTTTTAATTTGCATGTCCCTGATGATTAGTGACATTGAGTA TTTTTTTCATATGTCTATTGGCCATCTCTGTGTCTCTGGAGAAATGTCTGTTTCAGGTCCTCTGCCCATTT TTTAAATCAGATTGTTTCGTTTTGTGTGTTAAGTTGTATGAGTTCCCTATATATTTTGGATATTAACCC TTATTGGCATCTTCTCCCATTCAGCAGGTTATCATTTTGTGTTTGTCTAATGGCATCCTTCACTGTGCAAAA ACTGTTTAGTTTGTAGTAGTCCCATTTGTTTATTTTTTTTCTTTTGTGTTCCCTGCCAGAGGAAACATAT TCAAAGAAATACTACTAAAAGAGATGTAAAAGCGTTTACTGCCTATATTTTCTTCTAGGAGTTTACGGT TTTGGGTCTTAAATTTAACTCCTTAATCCATTTTATGTTTATTCTTATATGTATACAGTGATCCAGTTTC ATTCTTTTGCATGTATCTGTCTATAGTTTTCACACCATTTACTGAAGAGACTGTCTTTACCCAATTA TATATTTTTTGCCTCCTGTCTAGATTAATGACCATGTGGGCATGGGTTTATTCTGGGTCCTGTTCCAT TGATTTATGTGTCTGTTTTATGTGAGTACCATGATGTTTTGATTACTATGGTCTAGTAGTATAGTTTGA TATCAAGTAGCATGATACCTCCAGCTTTGTTCTTCTTATCAAGATCGCTTTAGCTATCTGGGGTCTGTT GTGGGGTCTACAAATTTTAGGGTTACTTGTCTGGTCTGTGAAAATGCCATTGGTATTTTGTATAGGAAT TGCATTGAATCTGTAGATTGATTTGGGTAGTATGAACATTTAATGATGTTAATTTCTTTTCTATTTCACAAA CATAGTATATGCTTCCATTTATTAGTATCTTAACTTTTCTTCTCAGTGTCTTACAGTTTTCACAGCACA GGTCTTTTACTTCTTAAATTCATTCTAGGTATTTTATCTATTTAATGCAATTTTAAATGGGATTGTT TTCTTAAATCTCTCTTTCTGATAGTTTGTATTGGTGTATAAAATGCAACCAATTTCTGAATPATTAATTT TGTGTCTGATACCTTACTGAATTCATTTATTAGTCTTAATGTTTTTGGTGGAAATCTTAAGGTTCTC TCTATATAGTATCATGTCTGTGAATAATGACAATTTTACTTCTTCTTTTCAATTTGGATGGCTTTT ATTTCTAGTCTGACTGCTGTGGCTGGGACTTCTAGTACTATGTTGAATAAAAGTGAAAGTGGCTTGTCC TGATCTTAAAGGAAAAGCTTTCAGCTCTTCACTACTGAGTATGATGTTAGCTGTGGGTTTGTCTTATATG GCCTTTATTATGTTGAGGTATTTTCCCTCTATTTCCCAATTTGCTGAGAGTTTATCATAAATAGATGTT GGATTTTGTCAAAATGCTTTTCTGCATCTATTGATATGATCATATGATTTTATCTTTTCAATTTGTTTAT ATAGTTTTATCACATTAATGATTTGCAAAATATTGAACCAACCTTGCATGCCAGGAATAAATCCCACTTAA TCATGGTGTATGAACCTTTTAAATGTACTGCTGAATTTGGTTTGTCTAATACTTTGTTGAGGATTTTGCAT CTATGTTGTTTCATCAGGGATGTTGGGCATTTTTTTTTTTTTTTTGTATTGTCTCTGGTTTGGTATCAGG CTAATGCTGGCCTTGTAATGAGTTTGGAGGCTTCCCTCCTTTTTCAGTTTTTGGAAATGTTTGGTAAAA TTTACCTGTGAAGTCATTTGGTTTCAAGGCTTTTGTGTTTGGGAGTTTTTGTATTACTGATTGATTTG TTAGCAGTTACTGGTCTGTTTCAAGTTTTCTGTTTACTGATTGAGCCTTAAATTTTCTGCTGATTCAAGCCTT GGAAGATTGTATGTCTAGCGATTTATCCATCTCTTCCAGTTTGTCCAATTTGTGAGCATATAGTTGTT CTAGTGTTCCTTATACCTTCTTGTATACCTGTGGTGTGAGTTGTGATCTCTTTTCAATTTCTGATTTTA TTTTGGCCCTCTCTCTTTTCTTCTGAGTCTGGCTAAAGGTTTATCAATTTGTTTTATCTTTTTCAGAGAA CCATCTCTTGTCTTTGTTTCTCTTTTCTATTGTCTTTTGTAGCTCTATTTTGTGTTTCTACTGATCTTTAT TATCTCTTCTCTTTTACACACTTTGGGCTTTCTCTTTTCTAGTTCCTTCAGGTATAAGGTTAGATTGT TTATTTGATATTTTTTTTTTGTGTTCTTGAGGTAGGCTGTATTGCTATAAATTTCTCTTAGAAGTGTCT TCGCTGTGCTCCATAGATTTTGGGCTGTGCTGTTTTTATTTGTCTCAAGGTATTTTTTGAATTTCTCCTT AATTGCATTGTTGACCCAGTCATTGTTTAGTAACATGTTATTTAGCCGCCATGTGTTTGTGTGTGTTTCA GTTTTTTTTCTTGTAAATGATTTCTAGTTTATACACAGAGAAGATGCTTGGTATAATTTCAATTTACTGAG ACTTATTTTGTGGCCTAACGTGGTCTATCTTAGACAGTGTTCCATGTGCACTTGAATATACTGCCGCTTT TTGGTGAAATATCCTAAATTTATCATTCAGTCCATCTGGTCTTATGTGTCATTTAATGTCACTCTTTCC TTGTTGGTGGGAATATAGTGATATTTTATTATGGTTTTTAAATTTGATTTTCCAGTGACAGTGATGATA ACTTTTTTCTAGTGTTTACTAGCTATTTGGATACCTCATTTGTGGAAGTCCCTATTTCAGGCTTTTGGCTT TTTTTTTTTTTTTTCAGTTGGGTAATTTGTCTTTTATTTATTTATAGGATTTTATTACATATTTCTGGATGT GAATCCTTTGTGAGATATGATCTTGTCAAATAGCTTCTCCAGTTTGCATCCTGTCTTTTCACTCTCCTA ATGGTGTCTTTGTGAATAGAGGTTCTTTTAAATCAAGACCAGTTTAAACAATATTTTTTCCCAATGGTT AGTACGTTAGGCCACTAAGAAAGTTTTAGCTATCTCAAGTTTCAAGAGTTATTCTCTGTGTGTTTTTATTT TTCTGGAAGCATTGTGTTTCAACATTCAAGATTATGATCCATTAAAAATGTTTTTGGTGTATATTGCAT GAAGTAGGTTAAAGTTCTTTATTGAAAAGACATTTTCTCTCACTGTTTTGTAGTGTCACTTTTGTCT ATAAATCCCAAGTGTCAATTTACTGAAAAGATTATTATTATTATTTTAAACCACAGAATTGCCTTGG AGCATTTGCTGTAAATTAATGACCAATATGTGTAGGCTATTTCAAGATTCTCTCTCTATTCCATTGAT

	CTCTTTGTTTGTCTTTGTGTGTCAGTATCACACTGTCTTAATTTATAGTAAATAGCTTTATAGTAAATCTTT AAAACCTCCAATTATTACATATAAATGTGAGAATCAGCTTGTGAGCGCCACCTCAAGGTCCCCCCCCC CCGATCCCTCCAACCTACTGAGGTTTTGACTGGGATCATATTGGAGAGATAAATTTGGGAGGCTGAGATC TTACAGTATTGAGGCTTCCAATCTGCACATGGTATATTTCTCCATTTATTTAGGTCTTTGATTTCTCTT ACTGGTGTCTTTCAGTGTAGACGTTTTATACATCTCTTCTAGGTGTTATTTCTTAATTTCTAATTTGTAGAT TCCAATGGATATTCTACATACATAATCATATATTTGTGAATAAAGACTGATCTATTGCCAGCCTTGATGC TTGTTTTGATTTCTTACCATCGTGCCTAGCTGGCACCTTCAGATAATGTTGAATGAAATGTAATAGTG GACAGTGCTTGTCTGTTTGATATATATTAATTTAGTGAAAGTTCCTGTTTCTACACGAGGGATCATAT GGGTTTACCTCGTTCAATTATTGACCACTTTTACTTATTTTTTTGTAGGCATGGCTAATGTAAAGCAGCC CCAAAGACAATTGACACAGGAGGAGGCTTCTTCTGGAAGAGGAGAGAGAGAAACATACAAATTGGAA AAGTTGTTCAACACAGGACCTGTTATGGAATTTGATTATGCGATATGTGAAGAATGTGGTAGAGACTT CATGGATTCTTATCTTATGAACCACTTTGATTTGGCGACTTGTGATAACTGCAGAGATGCTGATGATAAA CACAAGCTTATAACTAAAACAGAAGCAAAACAAGAATACCTTCTGAATGACTGTGATTTAGAAAAAAGAG AACCAGCTCTTAAGTTTATTGTGAAGAAGAATCCTCATCATTCACAATGGGGTGATATGAACTCTACTT AAAATTACAGATTGTGAAGCGGGCTCTTGAAGTTTGGGGTAGTCAGGAAGCATTAGAAGAAGCTAAGGAA GTTCGACAGAAAAACCGAGAAAAAATGAAACAGAAGAAGTTTGAT
4	RFe-V-MD4 AAGCAAATCCTAGAGCTTTTTGTTTTTTATACTATTCTATTGAAACAAAGTGGAAGGTTTAAAGAGGCAG CACATATACAAGTAGGTGAGTATCCAGTCAATAAAAGTATTGTTTTATTGTCAACAAGCTGAATCTAAT GCACCACACACATATATATACACATCATCAGATAGATACAGACTTGGTTAATTTGATGAGTGAGCAAAAT GAGAAGTAGACTGCTGCATCTACTGTTTTCTATGGAAGTGACATTTGAGCAACATAAATAGCTGATCAAA GATCTATAAGCACTGTGAGAAACAAGAATTCAGGTGTTTTCTGCTGTGACAATGAGCAACTCCAAGA AGATTAATCAGAAAAATGCATACCAAAAAAAAAAAAAAAAAAGGAAGAAAAAAGAAAAATGCATT CCTACTCACAACCATAACATTTTGTCTTTTGTGAAGTCCGTGTGCTGCTTTGGCGGTAGTGTGACACTGG AGAAATCTGTCCAGCAGCATCTCCCTGTTAGATACCCCTCACTCTTTCAACCTACATGAAATATATTTGT TTCCACTGAAATATCACGAGGGCCATCTACACAGCTTTTTTACGTTTTTGGCAGACCTCACTCCTTAGTG AACTCCTGGGGCAGTAACCTCTTCTCTCAAAATCATCTGGATGAATCCTCCTGTTATTTGAAATCAT CTCACTGAGCTTCAAGGGTCTCTTGTGAATTTGTGACCATAGCCTACCTCATATCAACAAAAGTTTCCAA TATGAGGTGTGGAAAGAGGATAAACTTTATTCACTGAACAGTTGGTAAACAGGAAAAACCGAAAGTGCA CACCAAGACAAAGGGAAGGGGCTTTTACAGAGAAAGTTAGTGCCCTGGTTCCCATTTGGTCCATTTTTT ATGCAAATGAGAAATCCAAATCACACAGTCTGATCAGTCAGCATCATATGTTCTGATTGGTTGTTGTGA ATCAGTTCTGATTGGTCAGTATAGATGCAAAATGAGGATATAACGCCACAGTTCTGATTGGATGGGACTAG TCTCAGTCTCTTGGAAAGTTCCATCAGGAGTTCCATCAGGAAGTTCTGACAATGGTTGACTTAGGCAGCA GCAGGAGCACAGTTCCGGAGGTGGAATTTTCAGTCTGTGGCTTTTCCCTGAAATGCAGAGTGTGCGAGAG GCTTGTGTGAGGAATGGCTGTTAGACTCTATTAAGAATTTGAGCTCAGTTAACCATGAGGAATCCTTCTT GGCAGATTATTTCTCTCAAGGTTCCACATTATGAGGGAGACTGCTTCAGAGCTTCCAATGAAAGGGCGG GTACAAGGGTGGTGATTGGACTACTGATATGCTTTTACGCCATAAGGCTCACATTGATGCTGGTAGGGAT CCCATTGTCATCTGCAGATGGATGTGTCTTTACGATTTGAGAATTGACTCTGACCCATGAGAAACAGAG CTCGAAGACTGGCTGAGGAGGGTACATTTGGGTCAATGTGACACAGAGTATTAAGTTAAGGCACACTGT TGTCAATTGATGATTGAGAGTTGCTCTGTAATGTCCACAGTTTTTTAGTTGTTCTTCTCTAGAACTTCTT TCTCAGGAAGCACTTGAACTTCATTGTAACAGATGAACCAAGAAGTCATTTTAAGCTCTTTTTTTTTT TTTAACTCTTTTTAAAAAGGTATTTTAGTGTTTTGTTTCTTAGTTGACTAAGAACAATGGCACATCATT ATATTAATACTAAAAATTCAGTGGTCAAATTTGGCTTATTTGAAATTTAGAAGGTAAGTGAACCTTTGGCC AAATTCCTTTCAAATGTAATAAATTTCAATTGTGATTCACTCAGCAACACTTTGAGATTAATTTGGGATT TGGGGATCAAAACTATCAAGCTTTTAGGTTGATGGTTAGAGGACTCTAGAACTATAATTATTAATTTCC TTGGTTGTGCCAGACAGATTGGGCATTATTGCTCAGAAATGAATAAATCAAAGTTGTTTGCATGAGAA ACTCACAAAGTTGCATGAGGGACAGAGTGGGTGTTGAGTGCTAGAGTGAAGGATACAGAGTGTAAAGCAA GTAAAGAGAAAGCAACCCAGAATAAACATAATGCCAGAACACATTTCTAAAATTAGGTTATGCTAAAGATG ATTCTAAAGAAATATGTGGGTGTGGCAAGCAAAATAATGGCCCTCAAATGTGCTAATCCTAATCCCTG AAATATGTTAACATGTTACTTTATACAGCAAAATGGACCTGTACAAATGATTAAATTAAGACTATTGAG ATGGGGAGATTGTTTTGTATTATCTGTGTGGATCCATTGTAATCTCAAGGGTCTTTGTAAGTGAAAGAGG TAGACAAGAGAATCATACAAAGAGATGTGATTATGGAAGCAGAGGTGAGAGTAATGTGGTCTCACATGAT GCCAAGTTTTGGAAGTGGATGAGTGAGTGCCATTCAATAAAGGAGGGTCAGGTATTATTTGTTAATTTCTT GACATCCATTTGCTTTATTCTGACAGCAGCTCTGTGTTTCATTTGAGGTTCTGTCCCTCTTTCCCCACT CTCAGCCCGTGGGAGGTACCCATGAGCCCTGCGATGATGTGAAACGGCTAAACAGAGCAGTTTCATGTCAT CTCTCTGGCTAACGTATTTGGTTCAGTGTTGGACATGTGACCTTAGCCGTTCTAATCTGAGTGACTGTCA AAACCTTTGGTGGAATACTAGGAAAAATAGTAAAAACAGAAGCTGCACAGTTCTTTTCTGCCTGGTTAGAA TCTGGAAGCATGCAGTTTAGGGAGATGGTGGTAGTCATTTGTGGTCACAAATGACCAGCATTCTGAAGGT GAAATTAATAAAAAAAAAAGAGAAATGAGAAGGAAGTAGCAAAACAGAAATGGCGCATGATCAGTGAGACT TGGAGCTTCTGCATCCAACGAGTCTTATCCTGGAACCAAAAGGTTATTTTGAGTTTTTTGTTTTT

	TTCTACACAATTTGATTTTGACTTTCTCTTACTTGCAATCAAACCTAATCTGAAGAGAGTACAGAAGAAAG GGCAGGCATGGATGTTTAAATTTAAAGACATCCACGTGGATTATGCTGTAAGGAAATGGAAAAATGGATT TAATGATCAGAAAAGTAGTGATATAGAAAGATGTTTATTTGGGATTATCAGCTCATAGATGGGAGAAAGC CGGGCATATTGATCATATTGAGTGAGACTAGAAGGGGTTAAGGTGAGAAGTTGAAGAATACCAATGTTT AATAGTCAGGCACAGTACAAGAAAACCTCTAAAAGACAGGGAGAAATCATTGCCAGAGACTAAACCTAAA TTTGTGAGTCTTCAAAGTGTAGTGATAGAGATTAAATAAAGAGAAGACACTTTAAGGAAATTTATTAATA TGTGAAGCAGTGCTGTGTTTTGTCTTTGGATATTGGGAATATGAATGATTTTTCTCTTTTACCTAAT TTTCTGTATCACTTCTGAAATAAACAAATACGTTTTGTGGGGTGGCCTAATGGCTCAGTTGGTTAGACTG TGAGCTCTCAACAACAGGTTGCTGGTTCAATTCGCGCATGGGATGGTGGGCTGCGCCCCCTGCAACTAA AGATTGAAAACGGCGACTGGACTTGGAGCTGAGCTGTGCCCTCCACACCTAGATTGAAGGGCAATGACT TGGAGCTGATGGGCCCTGGAGAAACACACTGTTCCCTTATATAGCACAAATAAAAAATTTAAATAAAAA CTCATATAAAGTCAACATAGAACATTGACTGTATTGAAAATCTTGAAATGTTTGTCAAAAATGGGGTCT TAAAATTAAGTTCGAGAACTTGCCACCTTGCGTTTACATTGGCAGCACTGTACAAACAGCTCGATAAGGT TTCATAACCTTGGTATATAAATCTCAGAGCTGTGTCCGTGTGGACATGTGGCGGTGTGCTGAATGGCAT TCATTATTGTTGTTGTGTGTTTTGTGTTGCATCGCAAGAATGTCTGAGCTTGAATTAGAACAATGAACA AACATTAAATTTCTGGTAAACCTGGCAAGAGTGGAAAGTGAATCAGGGACGTGTTAGTCCAAAGCTTAG AGGATAATGCCAAGAATAAAATGGCAGTGTACTAGTGGAGTAAACGTTTTTTCCGAGGGGAGAGAACGTG CAACTGATAAAGAGAGGTCAGGGCATCCAATAACGAGTAGAAGTATGAAAAAATTTGCAAAAATTCATC AAATGATCCATCAAAGTTATTGGCTGACTCTGAGAAGCATAGTAGTCCAAGGTAAATCAATAGAGAAAG ACAAAATCTGAAGTGAATCTTGCCATGAGGAAGATGTGTGCAAAAATGGTCCCGAAGTAGCTCACCGG TGAACAAAAACAAAAGAGAGTCCAAAGTTTGTCAAGACCTTTTGGAGAGGCAACATGACATTTTAGGCCAT GTTGTCACTGGTGATGAACATGGGTGTACCAATATGATCCTATAACAGAATGTCAAAGTACAAAATGGA AGTCAGCCAAATTTCTCCACGAAGAAAAAGTTCCATCAGTCCAAATCAAGGGTCCAAACGATGTTGCTGAC CTTTTTTGATATCAGAGGGATTATTCAATATGAATTTGTACCAACTGGACAAACAGTTAACCAAGTTTAC TATTTAGAAAGTGCAGAAAAGGCTGCGTGAAAAACTTCAGACGAAAATGGCCTGAACGATTTTCCAACAA TCATGGATTTTGCATCATGACAATACACCGGCTCACACAGTCTGTGAGGGAGTTTTTAACGACAAACAA ATAACCGTATTGGAACACCTCCCTACTCACCTCACCTGGCCCCAATGCCTTCTCTCTTTACCTGATGA TAAAGGAATATTGAAAGGAAAACATTTTGATGACATTCAGGACATCAAGGGTAACACGACGAGAGCTCT GATGACCATTCCAGGAAAAGAGTTCCAAAATGCTTTGAAGGGTGGACTAGGCGCTGGCATCAGTGCACA GCTTCCCAAGGGGAGTACTTTGAAGGTGACCACAGTGATATTCACCAATGAGATATGCATTACTTTTTCT AGAATGAATTCACGAATGTAATTGTGACACCTCGTATACTATAAGACAAGAATCGTAACCTCCAGTGCTT ATGGAGACAAAGAAAGGTGACCAAAGTAAAGTGAAGAACCAGGTGGGGACAGTAGCAAACCTAGAGAACA TGCTGTGATCTAAAAGGCACAGCACAGTAAAGTATCAAGAAGGACCAGGTTTGATTCTTTAGAGAAGCTTG ACATCCACATTCTACGTGAGTCTCCAAAATTGTCAGCGTTGATCAATACATGGAGGCAAAATTAACATAT CCAGGAGACACATTTAGTCTATAGGGCACTTGGGATTTTATATTTGCTGTTTCCAAATGGTTGTGTATAA TGTGAATATTTGTATGTAATCTTTCTTTCTTTGGTATCCTACGTTTATCCAAAATTTGGGCCGAG CTTGCAATAAAGACAGCTTGTCAATTTAGACTCATTTTACCACCTTCAGGAATTTTTCAAACCTATTTCAC ACCACAGTCCATTTGCATTTATTTTTACAGATTGTTAATCAAATACTCAATTCCTGCATAGGACCGCTG ATTCTAAATATTGAAACAGTTCCGTTCTGTTTGGTACGAACTCCAGGTTCTTGATGTTTTGATGTTTA AACCTACCTCCTGATTATGCCAGGGCTGTAGGAATTAACAGACATATTGAGACAGTCTATCGCACAGC TTCAACTAAAAGGAAGGTTTATGATTCTTACTGCTGCAAGGAAAGCATGCTGGTGTAAACATTTATTTG ATCTAACGACCTGAGCGTGAACAGAGATGCAAACTCTTTCTTCAAGGGTCGGATTCTACTTATTAGTAG ACTACCCATCAGCAATGTCTAAAGAGTCTCTGAGCGCCAGTGAATGACTGATGGCAAAAAGGAAACAGG TGTACTTCTGTAGGCCAGCAGATACCGCCAATGATATCCCTTTCACTTCTCGAGCCCACTGGTAAAGACA GTTCAAGTCAGCCTAAGCGTGTGCAAAGGAGAGAGATGAAGTAAAGTACCCCTCACTAACTGTACCTTTT CTAGAGGTTTCTTACGCTTTTGAAATCTGTGAAGTGATACATTACACTTATACATTACAGTACTTTTGAAA CAAGGGTTGTATCAGAACTCGGGGAACATTTCTAAATACACAATGTCCAGGCCCTATTAGATTGACTC AGTCAAAAACCTTCAGGGAGGAGTCCAGGCCGTGTAAGGTTTGTAAAGTTCTCAAGTGACTGTGAGTCG CCAGCACAACTCTAGCTGAGAAATACTGCAGTAG
5	RFe-V-MD5 TTCCCTCCTCCACTTACACCTGGAATGGTTGGATGGGTCCAGTGACATAGAAGGTGTGGTGGCTGGCAAA ATTCTGCCATACTTTGGGGTTACATGTATATAGATGTTAACTACTATACAGATGTGCCAGGCATTGTTCA CTATGTATTACATATTGAATTTTCAATAATGTTAAGAGGCAGGTACAATTAATACCTCCACTTTCAGA TGAGAAAATTAAGGCAGAGAGGTTACATAATGTGCCAAGGTACCACACCTTGATAAACAGCAGCTGGGA TTCTCACCCATCCAGTCAGCTTCAATCTGTGCACCTTAACTACTAGATGCTATATAGAATTAATGCCAA AACTCTCAAAATCAGAGTCATGAGAGAAAAGCCAAAGCCATCATGCCAATATTTGTTAGGTTAGGTTAGG CTATGTTAGGTTTCGTTTTATTTTTATTTCCCTAATTTCTAATCTTCTACATTTAGGGGAAGAGATGTG CTTCTATATTATGAATGTTTATGAATGAACATCGTATGGGACCATTATAACTGGACCTAAGGAGATAT GTTCTTGACATAATTCATTATCAATGATCAGCATCTCTTTGGGTTGATTGGCCATGTCTTTATCATCTC CACGTCCTATAGAATGTTCTTATGAAGAATATAGTCAGGACACACACACATACACACACGCGCGCGC

GCGATGGGGACTCTTAAGTAGCTCACCCTCCACAAAAAGCCTCATCTAGAAGCACAGAGTTATCATGAAA
 TACTCTGGGGGAGGGGGCATCATGGGGGTGGCAATTCAAAGAGAAATGAGAAAAATCACAGATGTTTAA
 ATCAATGGGGATAGCGCTGGAATTTTCCATCCTGAAGATTTTTTCCAGGGCTAAACCTCTGACTGAGTTT
 TGTTTCTTTAAACAAAGGAGGTGGTGGTGGTAGATTACCTTATTTTCAAAAACGTTCTTTGTAAACAT
 CCAAAATTTATTTCCATGAAAATTTGTTTCTCTTACATGTGACCTCAATTGTACTCAGCTGACCCTGTGACT
 ACTTGGAGTTGTGGTGAACAAAGTGCAACAGTTTCTCCTGGAAGTCTTTCATTTTCATTGTATGAGGT
 GTGATAAAAAAATACAGTGAAATGTTTAAATAAAAAATTTATTACAGTAAAGACACATTACCATTAATTT
 CTCTCAAAATACTCCCCCTTGCTTTGAACACATGTATCCCTTTGTTTTTGGCACTTTCTGAAGCAGTTC
 TGAAGTCTCTTTTATGAGTGTCTTTAATTGTACTGTAGTGGCTGCTTTGATGTCTGAATCAATTCAA
 AAGTTTACCTTTTGTGGTCATTTTTTCTTTAGGGAAGAGCCAGCAGTCGCACGGTGCCAGATCTGGTTA
 ATAAGGCGGATGAGGACACACCATAATGTCTTTAGTTGACAGAAATTGCTGTATACCAGAAGCGATGTTG
 GAGCATTGTCTATGATAGAGGATGATTACAGCACACTGTAAAACACACCTTCTCTCAAGTGATCTCACA
 CCAACTGACTGCACCAACAAGTTGAACTTGTACACATCATTACTAAGGTTTGACATGCAGCTTCTT
 GTATTGAATATCCCTGCCTTTCATTGGATGGCACTTAGCAGCAACGTTCACTGTATTGTTTAAATCACAC
 CTCGTACTTATTCTGATGGAGAAATTTTGTGAGTTGAGCACACTTCTCTCTCATCTCTTTTATTTTCT
 GTGTCTAGCTTTAGTTTGGGATGAGGGAGGACAAAGTACTATTATTATGAAATTACAGTGGCTCTGGAGG
 CCTCTCAATCCTGACTATGACACAGAAAATTTCTGAAATAATTCACAGCAGGAGTACTATAGGACTTGGT
 CAGCTTTGCATTGAACATAACTCCACATCTATATTGGCTCTGCTTTTGTCTCTCAATATTAAATGCTCAA
 ATATGTCAGTGCTAGGCACTATTATTATATCCCTCTGAAACATGTTTCTATTCAAGGATGCAGCATTC
 GAAGACTCAGTCCAGCGAGTGACAGAAAAAGACTTCCCTTGGATTATCTATGAGATTGTAATAGCTTATC
 TGCATATCTGCTCACTGAATACTGCCTCGATCATTCAATATCTGCTCACAATGGTAATCAATAAATG
 TGTGATGAATGGTCTACAATCCCAGATTGCAGCCCTAAGTCTGCTCATGATGGCTTCCAGTAGTTTTCTA
 TCAAAGCCACATGTGGTCAGTGTGCAGGATGAGGAGTCGAGCCCTTAAAACTCAACTCTAGAAGACCTAC
 TGAAGCAGTTATTACAACATGCTACAATACACAAAGAACAAGACTTGTACATCAGAAACAGGTTGTCTGA
 AAAAGTTTTCTATTGGGGGAATGAAGCAAATTTGAGCCTAAGTTTTCTGGACAAAAAGAAAGGCTGATTT
 ACTCAGTTTAAAGTCAAGACCAAAGAATAAGTCTGAGAAAAACAAGATGTTACCTGATCTTATATGCACT
 CTATTATTATTTTTGCTTTGCATGTCCCTTGTAAATAGTGATTGGTTTTAATGATCATTTTATATAAAAAAT
 TAAAAAGAGTACATTTTTTAACTTCTGTCAAAAATTTCTGCCTAAGGTACTTCTCAACACACACAGCT
 TAGTTGCTACCCCTCCTTCAAGGCTCTGTTTCACTGCCCCGTCTCTCCACGAAGACTTTTTTGTCTACACC
 TAGAAAGGCTCTGCCTACTCAGGCAGTTGTTATTACCTCCGATTTCCTACTATCAGATCTCTTCGTATTA
 TCTTCTTATATGACTAGGTCTCATCTCCCCCTCAACCACAATCTCTCTGAGGGCTGGAATATTGTGCACA
 TTGCCTTGACATAATAAAGGCTCCAGAGGTATCTGTCTAAACTGGCTTTATTTCTTGAGACTACAAGC
 ACTTATTCTGTGCCAGGCACCTTTTAGGTTCCAGGGAAAAAGAGGTACAAAACCAGACACAAACCCCTACCG
 TTATGGAGCTTACCTTTTTTAATTAAAGGTGGAAGGGATGAACCTTTTTTTGGTCTCTCTAGAAAAGTTGC
 AGCAGGAGACCATAGGAAATAGTATAAAATAGTTGAAAGCACTGTGGAGTGTGAGTCAGGATACCTTGGT
 CTCATCTCTAATTTGATGTATCTTGAGCACATTCTTAAACATTGGTCATCTGTTTCCCTGTATGCCATA
 TAGGAATCATATGGTTACTGGGAAAACCTGAATCAGAAAACAGATGCAAAATCATGTTGGAGGGAACTTTCT
 CAACCTGATAAAAAAGCATCTATGAAAAACCCACAGCTAACACCATACTTAAAGGTGAAAGACTGGAAGCC
 TTCTTCCGAAGATCAGTAACAAGACAAGGATGTCTGCTCTCACCCTGCTATTCAACATTCTACCGGAAG
 TTCTAGCCAGGTTCTAAGTAAGAAAAATGAAATAAAAAAGAAATCAAGATTGGAAATGAAGAAGTACTAAAC
 TATCTATTTTCTATGACATGACCTTACTTAGAAAATGCTAAAGAATCCACCCCCAACCCCCACCCCAAC
 AAACTATTAAAGCTAATAAGTGAATTCAGCAAGATTTTCAGGATACAAGGTCAATACGGAAAAAAAAG
 TTGTATTTCTATAAACTAACAATGAACAATCTGAAAATGAAATTAAAAACAACACATTTATGATAGCA
 TTAAAAAGAAATTAAGGAATAAATTTAGCAAAGAAAGTGAACACTTGTACGTGGAACAAACAAAACATT
 GTTGAAGAAATCAAAGACCTAAATAAAATTTTTAAAAATCCTGCCTTTGTGGATTAGAACACTTAATTTT
 GTTAAAAATAGCAGTACTCCTCAATTTGAATTATTCACAGCAAAATCCTACAAAAATCTTAGCTACCTTTAT
 TTTCTGCGAGAAATGACAAGCTGAGTTTAAATTTTACATGGAATGCAAGGAACCCAGAATATCCAAAA
 CAATCTTGAAAAAAAGGAACAAAGTGGGAAGACTCATACTTCTTAATTTAAAAAAGTACGCGCAAGCTAC
 AGTAATCAAGACTATGAGGTACTGGCATAAAGACAGACATATAAATCAATGGAATAGACTATGAGTCCAG
 AATAAATCCATGGTCAATTGATTTTTGATAAATGTGCCAAGACAAATCAATGGAAGAAAATAATCTTTTC
 AACAAATGGTGCTAAGACAACCTGGATATCCACATGCAAAGGATGAATTTTGAACCCCTACCTCACACCA
 TATACAAAAATTAGCTTGAATGGATCAAAGATATACAAATAAGTGTACAACTATAAAATTTGAAGAAA
 ACATAGGTGTAATCTCCATGACCTTGGATTAAAGCAATGTCTTCTTAGATACAAACATCAAAGCAACAAGCA
 ACAAAGAAAAACAATTGGATTTTATCAAAATTTGAAACTTTTGTGAGCCAACCTTCACAACCTCACACG
 GTGGCTCAGGTGGTTGGAGCGCCATGCTGGTTCGATTCCACAGTGGGCCAGTGCGCTGCATCTCTACAG
 CTAAGACTGTGAACAACGGCTCTCCTTGGAGCTGGGCTGCCACGGGCTGCCGTGGGCTACCATGTGCTGC
 CAGGAGCGGCTGGTGGCCAGCGTGAGTGACCGGCAGCCAGCGAGAACTGACATGAAGTGCTGTGAGTGGC
 CGAGAGGTCCAACCAAGTAACCGACTGCCTCAGCTGGGGGAGCGCAAGGCTCATAATACCAGCATGGGCC
 AGGGAGCTGTGTCTACATAGCTAGACTGAGAAACAATAGCTTACGCCGAGTGGTGGGGGAGCGGGAAG
 GGGAAAAACAACAACAACAACAACAAAA

6	RfRV AAATTAAGACTCACGTTAGGGAAGGCTGAGACAAGCAGCAGAAACCACTAGATAGGAACAAGAAATGTGA GGAAATCAAGGCAGGGAGCATGTGAAGTGGCAGGGAGGGGACAATGGAAGAGTGAACAGAGCAGAGGTG ACAGGCAGCAGAAGAGAAAGTGATTAGAAGAGAAGGTGGTACATTAAGCTGTTGGTAATAACAGAGACAA GAAATCGCAATAGAGGAAGAGTGTTGCTTCTGAAAGGAAAAAATCTAAATTAACCTAACTAAAAGCAATCT ACGATCACAACTCTACCTGTTAGGAGCAAATAGCACTATATACCTACATACCTCTGTCATCCACATGCA TTACAGTGCTGCCCTGGACAAACATGAGGGTGAATAAGTCCCGCTTTCCCTGGGAATGTCCAGTCTTA GCACGGAAGTCTGTATCCCAAGAAAAACACACACACAGTAGCAGTCTAATCAGGACAGTTGTTACCCCT GATTAGCATTGACTCAAAATAGCAGTGCAGTTTGGGGCTGGTCTGTAAAGTGTCCTTCTAGTGGTACTCA GGATTATTACTGCTTCACAGTAACCACACACATGCTAGTAAGTGTAAAGTCCGGAATGTCCCCCTCAG ACAGACAGAGAACCCGCACAATAATGCAAGTCACACAGCGAGGTTTATTACCAGCCGGCTGGGGTCCCCG TCTCTGCCCGACGCAGCGGGTTTTTAACAAGGACCCCAACACTTAAAGCCAAGGGGTATATAGTATT TTAAGGCCCTTCCATAGCTTCTGAGAGTACAAGATAAACTTACAGGACTTACATAGTTTCAAAGAGTATAA GATTTACTACAGGCAAGGAGACCAGGAGTATAAGATAAGCTACAGGACTTCTAGTGCCTATGCTGCAAC TGCCACATCCTGGAATTTTATAATTATGTTGTTTCAGGCTAGGGGCTGTTAACCCTGACCTGAAGATAG CAGTTCTCATGCTAACAGTCTCTTACATGCTAACAGTCTCTTACATTTCCCCCTGTGTGTTGTTCTATAA TGAGGAATCTTGCTCATGTACGGGCCAGCCGTAGAGGTCTCACGGGTGGGCACTGTCTTATACTGTTGT CTTAGGACCATGAGCTGGACAGTATTAACACGATCTTTTACAAAATTAATAAGCTTGTTTAATATGCAAG GGCCCAAGGTAGATAAAGCAACAGCAAGATAATTGGGCCACTAAGGTAGATACTAAGGTGGTGAGCCA AGGAGATCTGTTAAACCAAGCCTCAAACAGTCTCTCTGAGCTTCCCTTTCCCGTTTTCTCTTAGCTAAC CCTTCTCTCACTTTGCCATGGATTCTTTTACCACACAGTATGGTCAGCATAAAAAACAACATTCCTCCC CCAGCGCGGCACACAGTCCCCCTTGTGGAGGAACAACAATCAAGTCTCTTTTATTCTGGAGTACTAC CTCGGATAGAGAGGTGAGCGACTTCTCTAAATGACTAATCGAGGTTTCCACCTTTCTATGTCTCATCT ATGGCTGCCCTTAGGGAGGTCACTTCTGATTGTTGAGTAGCCAGGGAAGCTATGCCGGTCCCAGCTCCGG CTATTCCCAAACTGAACAGGTGGCAATGGTTAGTGCGGTGATGGGCTCTCTTTTACTTCTTGTGTCACT ATCCCAATGCGAATACATACTCTCCTCAGGGTGATAGAGGATGCGGGGCAGCACTGTTACTAGGACACAG AATTCAATTGGCGGCATTAAAGACCGAGGTAGATAAGCACGGGGTGAGGCCAGTCTTTGAACATATCCACC ATCCATCAGTTCTGGGGATTAAACCACTTAGTGTCATTTTCCAACTAGGGGAGCTGTCTATGGAGGCACA TAAACTTTGTTAGCCTGGGGCACCTTTCTAGACAGGTCCCATTGCCGCTCACTAGTTGCATGGTTAAG CCAATTTTACGATTTCCCATGAACACTGAGAAGGGTTCTTACCCTTAGAGGCCTTGAAGTGGCATTAA GTCTTATTGCCTCATAGAACGGGGCTTTTATATCATAGCACAGCCAACAGGAGGTTGTGAGGTTAGGACT GGTGGCATTAAAGGTCTCGTATACAGTGCGCACAGTTCGCGAATGAGTCTTTGGTTGGCTGCGTAGCA GGCGTCAAAGGAGTTACAGAGGTCTTTGGCTGGGTTCGCGCAGTCCCTCCTGCGGTGTTTTATCCCTTG ATATACCTGGGTTCTTGGTAGGGACGAGAGGGGCCAGCACCTTATTGGACCCACCTGAGTGCTGATCAT TTCCACCAGATAGTCAATAGTTAGGAGACCGCCGGGGTGGGGACCTATCCATGCCCAACGGCCTATATCT AATTGGAACCCCCAAGTTAATCCTGATAACCAACCACGCTCTCTTCGCGCCACGCTCTGGTTAACTGGA CAGGTACCTGGGGCACCGGTTGTGGGGTCCCTAAAGGAGAATTAACTAGATCCCTGTTCCCAACGTC CCACTGTCGGGGCCCGTCATAGGAGGTGACACAACCTCCAACCTACCACAATAGTAGCGGTCTGGGCCGCCA CAGGTTTTTCAATTGTTTCTTAGGTTCCCTGGGCAGGGCCAAAACCTTGTGCACTATGTCCTTGAGAGG TGTCATTACAGCCCGCTTGACCTGACTGAGTAGTCATACTGCCGTCCACGTTTAGTGCCGAAAATGTC ACGCAGGTGAAAAACAAGTCTGGCCACCACGTATTGATGGGGGAGTATGTGTGGTGCTATTAAAGGGTT GTTTGGGTCTGTCCATCTGTTAGGGTCCATGTTAGCTTATGGGGTTGGTGTGGGTGATCCCCGCGTGGC TCTTCTCCAGATATTGAGCAGAGTTAGGGCTAGCAGCCATTCCATCGTTAGCTGAGCAGGGGGCTTGA CGCTTCCCCGAGGTCGGGAGAGCTGCAGCTTCAAGGGTTATCAGGGTGTCGCGGTACGATCCACTCCTT AGCTTGCCTCTTCTCCAGCTGGCTGGCTCGGCGCAGTGAGAGTGATGGACCAAGGCCAATGCCGTCA ACCTTTAAGGCAGTGGGGGTAAACCAGATAACACATAAGGACCTTTCCATCTCTCTCCAGTGTCCGGG ACCTGTGTCTCCTTACCCATACCCAATCCCCGGAACGATGCCATGTTCCGGGTTTGGGGCGTCTTAAAT TTCATACAGGGAACTCACTAGGGGCCATATCTCATGTTGGACCCCTTGTAGGGCCTTAAACTGGCCAGA TAACTTGGGGCCACATTGGGGTCATGATCTGGTAGAGTACGAACAATAATGGGGGGTGGTGGCCCATACA GAATTTGAAAGGTGTCAAACCATGTACATATGGTGAGTTCGCGACCCGGAAGATGGCATAGGGTAAGAG GGTCAACCCAGTCCCCGCCAGTCTCGATGGCTAGTTTGGACAAGGTCTCCTTTAGAGTCCGATTCACTCTC TCTACTGCCCTGAGCTCTGGGATTATATTCACAATGTAACCTCCAATTGATCCCTATCGCTCGGGCTC GTCCTTGTAGGACGTTACTGATGAAAGCTGGGCCGTTATCGGAGCCTAAAACCTCAGGAACCCCATATCT GGGAATAATTTCTTCTAGTAATGCCTTAGCAACCACTTGGGCAGTCTCCCGTTTCTGTGGGAAGGCTTCC ACCCAGCCCGAAAAATGTGTCAACCATTAAGCAAGTACTTATACCCACACCTCCCAGGCTTTACCTCAG TAAAATCCACTTCCCAACTCCGTCCCGGCGCTCTTCCCGTACCTCGTACCTGTATGTTGGGGTCTTTT TCTACTGGGTCTCATAGCCTGGCACCCAATGCACTGATCTACAATCTCTTGAATCTGAGCCGCTTGTCCG GGAAACCGAGGCGGGCGGACTCGAGAATTGTGAGCAACTTCTTTTTCCTAAGTGGGTGGCTGTATGCA GGTTGGAGAGAAGAAACAGTCTAGCTGTGCCGCGAGTATCAATCTTCTTCTGTATCCCGATGCCACCC CTGCTGATCAGATTCCGGGCGAGTGGTGGTCTGGATCCATCGCAGGTCTTCTGGAGTGATGTCAGGTCCG GGGGCAGGCGAGGAGCTCAGGTGTGGGCAGGCTGAGTGCTAAAGCTGATGAAGCTACTGCCACTGCCT
---	---

TTGGCGGCTTCATCCGCTCGCCGGTTTCCTTTAGCTTCCGGGGTCTGGGCAGACTGGTGCCAGGGATGTG GACAACTGCGACTGCCCGGGCATTGTACAGCCATCAGCAGTCTTCGTACCTCAGGAAGATTGCGCAGA GTCTTTCCCTCCGCTGTAACAAAGCCTCTTTCCCGGTAGATAGCGCCATGCACATGGACAGTGCCAAAGG CGTAGCGGCTATCGGTGTAGACAGTCACTCGTCTCCCTTCGGCCCGTTCCAGCGCTTCCGCCAGCGGAT CAGTTTCGGCCTTCTGTGCTGATGTCCCCGGGAAAGCGAGGCACTCCAAATGATGTTTCCCCCTTGGTCT ACCACCGCTGCGCCTGCCCTCCGCACACCATCTATAACGAAGCTGCTTCCATCAGTGTACCATACCAACT CACTGTTGGGTAGTGCCTGTCTGGAGGTCCGGGCGCACCTGGGTGACTTCTGCCATGATCTCTTGGCA GTCATGCAGGGGAGCTCTCAGATCCGGGGTCCGCAGCAGGGTGGTGGATTAGAGCGGTGGGTTCAGCG AAGATGATCCGGGTGCATCTAGCAAGAGTCTTGGTAATGTGTTAGTCGGGCATTAGTCATCCACCTAC CAGGGGGATATTTAGGACCCCTCGATCGCATGGGGGTTACTACCTTCAGATGTTGCCCAAAAGTGAG TTTATCAGCATCCTTCACCATTAGGGCTACTGCCGCAATGATCCTCAAGCACGGGGGCCATCCTGCTGCA ACTGGATGATGCTTCTTGGATAAATAGGCAACCGGGCTTTCCAGGGCCCCAGACGCTGCATTAGCACCC CTTTGCTATTCCCCCTCTCTCATCAACAAAGAGAGTGAAGGGCTTCAGGGGGTCTGGCAATGCCAGAGC CGGGGCTCTTAGGAGAGCGACCTTGAGTTTCATCGTAGGCCCTTCTGTTGGTCTGACCCCCAGGCCCAAGG ACCTTATCCTTGGTTGCCTCATACAGAGGTTTTGCTATTTTCAGCATACCCCAAAATCCACAGCCGGCAGT AGCCTGTCGTCCCTAAAACTCACGGACCTCTCGTGCTGAGGTCCGGACTGGAAGTCAAGAAATAGTCTC TTTCATGGCCTCTGTCAGCCATCTGGCTCCTTTTTTTAGTTTATACCCAGGTAGGTGACTGTTTGCCTG CATATTTGAGCCTTCTTTGCACTGGCCCGATAGCCCAACTGCCCCAGCTCCTGGAGGAGGTCTCCAGTGG CCTGTGCGCATTCAGCTTCGAGGGGGCTGCCAGAAGCAAGTCACTACGTAAGTCAAGGAGCGTAACTGA ATTATGGCTCTGGCGAAACGAGTCCAAATCCTGATTTAGGGCTTCATTAAACAGAGTTGGAGAGTTTTTG AAGCCTTGGCGTAGTCTAGTCCAGTCAAGTCCGCGGGGTTCCCGTATTGCCATCATTCATTGCAAG CAAAATGTGTTGGCTGCTGGGTGCCAGGGCTATGCTAAAAAACCCATCCTTTAGGTCTAAGGTAGTATA CCAGACATGTGAAGGGGGCAAGTGAAGTATAGGTATAGGGTTGGGGACCGTGGATGGATGTCTTCA ACCCTCTTATTTACTTCCCTCAAGTCTGGACTGGCTTATAATCTTTCCCCCGGTTTCTTAACGGGGA GAAGTGGGGTGTTCAGGCAGAATGGCAAGGTTTCAGTATTCCAGCTTCCAGTAAACGGTTAATGTGCGG GGCAATCCCTTTCCGCGCCTCTGCAGACATGGGGTACTGGCGGATCCGGATAGGCTGGGCTGAGGCTTTA AGTTCACCACTACTGGTGCTCGGCGGGCCGCCCGGCCACACCGCTATTTTTGCCACGCTGAGGGT ATGTTTTAAGCCAATAATCCATATCACGGGGCCATTCTGTAGAGGGAGGGTGTAGGGGTGTCTGCTGAG GGCGAACAGGCGATGTTATCCACAAGAGACAGGGTCAAAATGTGGAGGGGCTGTCTTGGCCATCCAAAT AGCTTAATGCCATCCGGCTCAAAATGGATCTGAGCCCTGATCTTAGTCAGGAGATCCGCCCCCAATAAGG GGGACAGGCATTACAGGATAACTAGGAAGGAGTGGGTCACTTGGTGGCGGCTAAGTCTACTTGGCGCCT ACTAGTCCACCGATAAGCCTTGGACCCAGTTGCCCTTGCACCAAACTGGTTTTCTGAGATAAGGGCTCT GTGGGCTTATTCAAAACAGTACTGGGCTCCTGTGTCTACCAGGAATCCTACTGGCTTCCCCCTCCACAT ACGCAGTTACCCAAGACTCGGGGAGGGGATCCGAGTCCCGTCTCCCTAGTCACTCTCCATCCCCGCGAG CAGGACCCGTGCGTCTTGCCCTGTTTGGCCCTGGCGCTTGGGGCACTCCCTCTTCCAATGTCCATACTCT TTGCAGTTTGCACACTGTCCCTATCCAGTCCGGGCGCGGTCTCCACGGTCCGGCTGGTCCGGCCGGAC TCGGTCCCACCCCTGACTGTGCTTTGCACGCTGCCAACAGATCTTGCCATCTCCCTCTGCTGCTCTCT GTTCTCCCTACTCTGATGCTCTGTCTTCTTCTTCTGATTCTGTTCTCTGTAATTCCTGATTTCTTTCTA ATTCTAATCCTCCCTTTCTTCGGGAGTCTCTCGAGTGTGTAAGACTCTCTCCGCTACTTTTCAATAATCCC GGATAGACATTTCTCCAGTCCCTCCTGTTTGTACAATTTCTTCTAATATCTGGGGCAGCCTGGTTTAT AAAGGACATAATTACAGCCGACTGGTTTTCTCTGCCAGGGGGTCCAACGGGGTGTACTGTCTGTAAGCA TCATAGAGGCGTTCTAAAAACAGGGCCGGGCTTTCATTATCCCCCTGCATTATAGCTTTTACCTTGGCCA AATTGGTGGGGCGGCTGCCCGCGCTCGGAGACCTGCCATAAGAGTCTGGCGGTAGACTCGGAGACGCTC CCTACCTTCTGCGTTCCCAAAGTCCCAATCCGGTCTATTTCAGGGGAAAGCGCTCGTCTATCAGGTTCCGC AAGGTCTGCTGGTCTTCCGTTGTGCGCGGGGACATTTTTTCTGGCCTCGGTGAGGATTCGCTCGCGCTCCT CGGTGGTGAATAAGGTCTTAAGAAGCTGCTGGCAATCATCCCAAGTGGGACTGTGTGTGTGCATGACAGA CTCGAACAGGTCAAGTAGGCTTTTCGGGTCTCAGAAAAAGGAGGGTTTTGAGCCTTCCAGTTGTACAGA TCACTGCTAGAAAAAGGCCAGTACTGGTATGCTCGCTCCCATCTGGACAGTTCTCTCTAGTGCTCGCA CAGGGAGAATCGCGCCCCAGCGGAGGTGGAGAGGACGGTTCCTCTCTGGGTCTCTTCCCGCGCTCT GCGAGGCCCAATCCCTCGCCGGCCCTTGAAGTGGGGCAGGGGAACCCGGGGAGGGGAGCCATCGGG GAGGCGGTAGAGTGAAGCAGCTCGGAAGGAGCGGCAGCCTCAGGCGGGAGCGGCGCCATCGGGCTGGCG CGCGCGCGGCTGGAGCGGCGCCACCGCGCGTAAGGAGGGGGGTCTCTTCCAGTCTAGGTCTATCAG GGAGGGGTATATGTCTGACCCCTCTGAAGGATGGGTCCCTGGGTGGCTCTCTCGGGGGTTTCGGGCGG GCCGTGGGCACTGCCGACTGGCGGGAGGGTCCCGAAACGGTCAAGACTTTAAGGGGAGGGGGGATCTT CTGGCTTGTGGGGATAAAGGGCTTAAGCCAGGAGGGAGGACTCTCTACTAAGGCTTGCCACATCAAAAT ATAGGGATATTGGTCCGATGGCGCCGATTAATAATATCTCGGACCTTTTTAATAATGTCTAGGGAAAA GTTCCCTGGGGGGCCAGCCACATTAAAAAGTAGGCCATTCTGCAGAGCAGAATGTATCAAACTTACCTT TCTTCACTTCCACACCATGATTACGAGCCTTGGCGCGGATTTTCAGGAAAGTGGTTTCAGGAGCAGGGTCTT AGGTGTTTACCTGAACCTGTCCATAATTGTCAAAAGAGAAACCAAGAAAGGCAAAAGAAAGGACAAAA GACACAGTGCCAGCAAATACAACTTCGCACAGGACTCTTCAACACCCACCGGCCGGTCAACACACCA CATCCACAGGCGCGGTTTCAATCACACAGTCTCACCAGCTCAAGATCCTTACCTAGGGCCCGTCCAAA
--

	CGGCGTCCACTGTGGACGTGCTGGGCCACCTTCTCGTCGGGGACGTCTCCACGACTTCAAGTAACGAA GCCTCCAGGGTCGTAACCTGCACTTTCTTCCCGTGAGAATTCTCAACTGGGACCGGGCAGAGACCTGTT TCAGTCTCTCCGGTCGAGGACCTGTTTCAGTCCTCCCTGTTTGGGACCGGGCAGAGACCTGTTTCAGTC TCTCCGGTCGAGGACCTGTTTCAGTCCTCCCTGTTTGGGACCGGGCAGAGACCTGTTTCAGTCCTCCG GTCGAGGACCTGTTTCAGTCCTCCCTATTGGAGGTGGCCAAACCTCCTTCCGCGGTTCCCTATGTAAAC CTCGGTATCGGGAGTTGTCTGTTCCCTGAGGGGGGGCGTCCCGGGCGAGCCCCAAATGTTAAGATCCG GAATTGTCCTCCCTCAGACAGACAGAGAACC CGACAATAATGCAAGTCAACAGCGAGGTTTATTACCAG CCGGCTGGGGTCCCGTCTCTGCCCCGACGACGCGGGTTTTTAACAAGGACCCCAACACTTAAAGCCAAG GGGTTATATAGTATTTTCAAGGCCTTCCATAGCTTCTGAGAGTACAAGATAAACTTACAGGACTTACATA GTTTCAAAGAGTATAAGATTTACTACAGGACAGGAGACCAGGAGTATAAGATAAGCTACAGGACTTCTA GTCGCCATGCTGCAACTGCCACATCCTGGAATTTTATAATTATGTTGTTTCAGGCTAGGGGCTGTTAAC CCTGACCTGAAGATAGCAGTTCTCATGCTAACAGTCTCTTACATGCTAACAGTCTCTTACAGTAAGAAGT TCCAAGCCTGTGGTGGCAGTAAGTGAATTTCTTCTTTCAATAGACTATGAAGGAGGGACATTGCATT TGAAGTCACTCCATGAGTCATGATGCTCTTTATGTCCATTAAGGAGTAACTTTCTCTCTATTCACTAT TTCTTTCACTATTGTATAGGGTAACGTGTTTGGGGAGAAAAATCAATAAAATGCTTAAATAAAAGT TCCATGCTCATAAGGTTTTTATCTTCCATTATAGGAAAATGAATCTATATGGAAGGTCATTTTCTGA TGATGTTTTGTAAGAAGCATTATTCTATCAATCTATTAATAATATATTGATGCACTTTCC
7	Part of RFe-MD-2 sequence with Columbid/Falconid DNA homology TCCTCGTTAGTATAGTGGTGAAGTATCCCCGCTGTACGCGGGAGACCGGGGTTTCGATTCCCCGACGGGG AGGCAGv
8	Protein sequence of RFe-MD-2 fragment that shows homology with Columbid and Falconid herpesvirus homologous with hypothetical proteins CoHVHLJ_080/FaH\HV1S18_80 of the Columbid or Falconid herpesvirus PRRGIEPRSPA*QAGILTTILTRM
9	Part of RFe-MD-2 sequence with Sindbis virus (hairpin) homology TATAGTGTGAGTATCCCCGCTGTACGCGGGAGACCGGGGTTTCGATTCCCCGACGGGGAGGCA
10	Part of RFe-V-MD3 sequence with Human herpesvirus 4 isolate HKNPC6 homology TTCATCCATGTCACAAATGACAAGATTTTGTTTTTTATAGCTGAGTAATATTCCATTGTATACATATACC ACATCTTCTTTGTGTAATTCGTCTGTCAGTGAACCTTTGGTTACTTCCATATCTTGGCTGTTGTAAATAATG CTGCAGTGAACATAGGGGTGTGTATATCTTTTCGAATTAGTATTTTGGATTTTTTTCAGATAAATACCCA GAAGTGAATTTGCTGGGTCATATGTAATTTCTATTTTAAATTTTTGAGGGACCTCCATAGTGTTCCTG TAGTGGCTGCACCAATTTACAAGGTGCTTTTCTCTACATCCTTGCCAACACTTGTGTTTATGATTTAT TGATGATGGCCATTTCTGACACGTGTGACATGATAGCTCATTGTGGTTTTAATTTGCATGTCCCTGATGAT TAGTGACATTGAGTATTTTTTCATATGTCATTTGGCCATCTCTGTGCTCTGGAGAAATGCTGTTTCAG GTCTCTGCCCCATTTTTAAATCAGATTGTTTCGTTTTGTGTGTTAAGTTGTATGAGTTCCTTATATATT TTGGATATTAACCCTTATTGGCATCTTCTCCATTCAGCAGGTATCATTTTGTGTTTGTCTAATGGCATC CTTCACTGTGCAAAAACCTGTTTAGTTTGATGTAGTCCCATTTGTTTATTTTTTTCTTTTGTTCCTCTG CCAGAGGAACATATTTCAAAGAAATACTACTAAAAGAGATGTAAAAGCGTTTACTGCCTATATTTCTTCT TAGGAGTTTTACGGTTTTTGGGTCTTAAATTTAACTCCCTTAATCCATTTTGTAGTTTATCTTATATGATA CAGTGATCCAGTTTTCATTCTTTTGCATGTATCTGTCTATAGTTTTTCCAACACCATTACTGAAGAGACT GTCTTTACCCAATTATATATTTTTGCTCTCTGTCATAGATTAATTGACCATGTGGGCATGGGTTTATTTTC TGGGTTCTGTTCCATTGATTTATGTGCTCTGTTTTATGTCACTACCATGATGTTTTGATTACTATGGTCT AGTAGTATAGTTTTGATATCAAGTAGCATGATACCTCCAGCTTTGTTCTTCTTATCAAGATCGCTTTAGC TATCTGGGCTCTGTTGTGGGCTCTACAAATTTTAGGGTTACTTGTCTGGTTCTGTGAAAATGCCATTGG TATTTTGTATAGGAATTGCATTGAATCTGTAGATTGATTTGGGTAGTATGAACATTTTAAATGATGTTAAT CTTTCTATTTCACAAACATAGTATATGCTTCCATTTATTAGTATCTTAACCTTTCATTCTTCAGTGTCTTAC AGTTTTCCAAGCACAGGCTCTTTTACTTCCCTAAATTCATTCCTAGGTATTTTATCTATTTAATGCAATT TTAAATGGGATTTGTTTTCTTAATCTCTCTTTCTGATAGTTTGTATTGGTGTATAAAAAATGCAACCAATT TCTGAATATTAATTTTGTGCTCTGATACCTTACTGAATTCATTTATTAGTCTAATTTGTTTTTTTGGTGG AATCTTAAGGTTCTCTCTATATAGTATCATGTCATCTGTGAATAATGACAATTTTACTTCTCTCTTTTCA ATTTGGATGGCTTTTATTTCTAGTCTGACTGCTGTGGCTGGGACTTCTAGTACTATGTTGAATAAAAGTG AAAGTGGCTTGTCTCTGATCTTAAAGGAAAAGCTTTTCAGCTCTTCACTACTGAGTATGATGTTAGCTGTG GGTTTGTCTATATGGCCTTTATTATGTTGAGGTATTTTCCCTCTATTCCCAATTTGCTGAGAGTTTTTA TCATAAATAGATGTTGGATTTTGTCAAATGCTTTTCTGCATCTATTGATATGATCATATGATTTTTATC TTTCATTTTGTTTATATAGTTTATCATTAAATGATTTGCAAAATATTGAACCAACCTTGCATGCCAGGA ATAAATCCCACTTAATCATGGTGTATGAACTTTTTAAATGATGCTGAATTTGGTTTGTCTAATACTTTGT TGAGGATTTTGCATCTATGTTGTTTCTCAGGGATGTTGGGCATTTTTTTTTTTTTTTTGTATGTCCTCT GGTTTTGGTATCAGGCTAATGCTGGCCTTGTAATGAGTT
11	Part of RFe-V-MD3 sequence with Human herpesvirus 4 isolate HKD40homologyTAAATACCCAGAAGTGAATTGCTGGGTATATGGTAATTCTATTTTAAATTTTTTG

	AGGGACCTCCATACTGTTTTCCGTAGTGGCTGCACCAATTTACAAGGTGCTTTTCTCTACATCCTTGCCA ACACTTGTGTTTATTGATTTATTGATGATGGCCATTCGACACGTGTGACATGATAGCTCATTTGTGGTT TTAATTTTGCATGTCCTGATGATTAGTGACATTGAGTATTTTTTCATATGTCATTGGCCATCTCTGTGT CCTCTGGAGAAATGCTCTGTTGAGGCTCTGCCCATTTTTTTAAATCAGATTGTTTCGTTTTGTGTGTTAA GTTGTATAGTTTCTTATATATTTTTGGATATTAAACCCTTATTGGCATCTTCTCCCATTCAGCAGGTTAT CATTTTGTGTTTGTAAATGGCATCCTTCACTGTGCAAAAACCTGTTAGTTTGTATGTAGTCCCATTGTTTA TTTTTTTTCTTTTGTTCCTCCCTGCCAGAGGAAACATATTCAAAGAAATACTACTAAAAGAGATGTAAGG CGTTTACTGCCATATTTTTCTTCTAGGAGTTTTACGGTTTTGGGTCTTAAATTTAACTCCTTAAATCCATT TTTAGTTTATTCTTATATGTATACAGTGATCCAGTTTCATTCTTTGTCATGTATCTGTCTATAGTTTTTC CAACACCATTTACTGAAGAGACTGTCTTTACCCAATTATATATTTTTGCCTCCTGTCTATAGATTAAATGA CCATGTGGGCATGGGTTTATTTCTGGGTTCTGTTCATTGATTATGTGTCTGTTTTTATGTGAGTACCA TGATGTTTTGATTACTATGGTCTAGTAGTATAGTTTGATATCAAGTAGCATGATACCTCCAGCTTTGTCTC TTCTTTATCAAGATCGCTTTAGCTATCTGGGGTCTGTTGTGGGGTCTACAAAATTTAGGGTTACTTGTTC TGGTCTGTGAAAAATGCCATTGGTATTTTGATAGGAATTCGATTGAATCTGTAGATTGATTTGGGTAGTA TGAACATTTTAAATGATGTTAATTTCTTCTATTACAAAACATAGTATATGCTTCCATTATTAGTATCTTA ACCTTCATTCTTTCAGTGTCTTACAGTTTTCCAGCACAGGTCTTTACTTCTTAAATTCATTCCTAGTT ATTTTATTCTATTTAAATGCAATTTTAAATGGGATTGTTTTCTTAATCTCTCTTTCTGATAGTTTGTATTT GGTGTATAAAAATGCAACCAATTTCTGAATATTAAATTTTGTGTCTGATACTTTACTGAATTCATTTATT AGTTCTAATTTGTTTTTTGGTGGAAATCTTAAGTTCTCTCTATATAGTATCATGTCTGTGAAATAATG ACAAATTTTACTTCTTCTCTTTCAATTTGGATGGCTTTTATTTCTAGTCTGACTGCTGTGGCTGGGACTTC TAGTACTATGTTGAATAAAAAGTGAAAGTGGCTTGTTCCTGATCTTAAAGGAAAAGCTTTACGCTCTTCAC TACTGAGTATGATGTTAGCTGTGGGTTTGTCTATATAGGCTTTATTATGTTGAGGTATTTTCCCTCTAT TCCCAATTTGCTGAGAGTTTTTATCATAAATAGATGTTGGATTTTGTCAAATGCTTTTCTGCTATCTATT GATATGATCATATGATTTTTATCTTTCAATTTTGTATATAGTTTATCACATTAATGATTTGCAAAATAT TGAACCAACCTTGCATGCCAGGAATAAATCCCACTTAATCATGGTGTATGAACCTTTTAAATGTACTGCTG AATTTGGTTTGTCTAATACTTTGTTGAGGATTTTGCATCTATGTTGTTTCATCAGGGATGTTGGGCATTTT TTTTTTTTTTTTTGTATTGTCTCTGTTTTGGTATCAGGCTAATGCTGGCCTTGTAAATGAGTTTGAGAGC CTTCCCTCCTTTTTCAGTTTTTTTGGAAATGTTTTGGTAAATTTACCTGTGAAGTCATTTGGTTTCAAGGCTTT TGTTTGTGGGAGTTTTTTGATTACTGATTTCGATTTTGTAGCAGTTACTGGTCTGTTTCTGATTTTCTGT TACTGATTCAGCCTTAATTTCTGCTGATTCAAGCCTTGAAGATTTGATGTGTCTAGCGATTTATCCAT CTCTTCCAGTTTGTCCAATTTGTCAGCATATAGTTGTTCTAGTGTTCCTTAACTTCTTTGTATACCTG TGGTGTGAGTTGTGCTATCTCTTTCATTTCTGATTTTATTTTGGCCCTCTCTCTTTCTTCTGAGTCTG GCTAAAGGTTTATCAAT
12	Part of RFe-V-MD3 sequence with Human respiratory syncytial virus (Kilifi isolate) homologyTTTTCTTCTAGGAGTTTTACGGTTTTGGGTCTTAAATTTAACTCCTTAAATCCATTTTATGTT TATTTCTTATATGTATACAGTGATCCAGTTTCATTCTTTTGCATGTATCTGTCTATAGTTTTTCCAACACC ATTTACTGAAGAGACTGTCTTTACCCAATTATATATTTTTGCCTCCTGTCTATAGATTAATTGACCATGTG GGCATGGGTTTTATTTCTGGGTTCTGTTCCATTGATTTATGTGTCTGTTTTTATGTGCTAGTACCATGATGTT TTGATTACTATGGTCTAGTAGTATAGTTTGATATCAAGTAGCATGATACCTCCAGCTTTGTTCTTCTTTA TCAAGATCGCTTTAGCTATCTGGGGTCTGTTGTGGGGTCTACAAAATTTTAGGGTTACTTGTCTGGTTCT GTGAAAATGCCATTGGTATTTTGATAGGAATTCGATTGAATCTGTAGATTGATTTGGGTAGTATGAACAT TTTAAATGATGTTAATTTCTTCTATTACAAAACATAGTATATGCTTCCATTATTAGTATCTTAACTTTCA TTCTTCAGTGTCTTACAGTTTTTCCAAGCACAGGTCTTTTACTTCTTAAATTCATTCTAGGTATTTTAT TCTATTTAATGCAATTTTAAATGGGATTGTTTTCTTAACT
13	Part of RFe-V-MD3 sequence with SARS-CoV-2 homologyAGGTTTCATCCATGTACAAATGACAAGATTTGTTTTTATAGCTGAGTAATATTCATTGT ATACATATACCACATCTTCTTTGTGTATTCTGTCTGTGAGTGAACCTTTGGTTACTTCCATATCTTGGCTGT TGTAATAATGCTGCAGTGAACATAGGGGTGTGTATATCTTTTGAATTAGTATTTTGGATTTTTTTCAG ATAAATACCAGAAGTGAATGCTGGGTATATGGTAATTTCTATTTTTTAAATTTTTTGGGGACCTCCAT ACTGTTTTCCGTAGTGGCTGCACCAATTTACAGGTGCTTTTCTCTACATCCTTGCCAACACTTGTGTGT TATTGATTTATTGATGATGGCCATTCGACACGTGTGACATGATAGCTCATTGTGGTTTTTAATTTGCATG TCCCTGATGATTAGTGACATTGAGTATTTTTTTCATATGCTATTGGCCATCTCTGTCTCCTCTGGAGAAA TGCTGTTCAGGCTCTGCCCATTTTTTTAAATCAGATTGTTTCGTTTTGTGTGTTAAGTTGTATGAGTT CCTTATATATT
14	Part of RFe-V-MD3 with RNA-dependent DNA polymerase of Erythrocytic necrosis virus homology PTSLMNNIDAKILNKVLNQIQQYIKKFIHHD*VGFIPGMQGWFNICKSINVINYINKMKDKNHMIISID AEKAFDKIQHLFMIKTLSKLGIEGKYLNIKAI
15	Part of RFe-V-MD3 with RNA-dependent DNA polymerase of Lymphocystis disease virus homology

	MTSQVNFTKHSKLRREGSQTHLQGGH*PDTKTRDNTkkkkKC-PTSLMNNIDAKILNK VLNQIQQYIKKFIHHD*VGFI PGMQGWFNICKSINVINYINKMKDKNMIISIDAEKAF DKIQHLEFMIKTSLKLGIEGKYLNIKAI*DKPTANIIISSEELKAFPLRSG
16	Prediction of a potential new spike protein sequence (RFe-SP2) (M) FVVVVVVVFPRRLPHHSVSVYCFVLCRTQLPGPCWYYEPCAPPSSGRLLVGPPLGHSQHFMVSLAGC RSLTLATSRSWQHMVAHGSPPWQFSSRESRCSQSLRMQRTGPRGNRTSMALQPPEPPCEGCEGWLTKVFN DEIQLFSEFVACALMLYLRRHCLIQGHGDLHLCFLQVLLHLEFVYLSISSFLYMGVRVSKFILLHVDIQLSH HLLKRLFESSIELSWHIYQKSIDHGFILDSSILYMSVFMVPHSLDYCSFAVSFIRKYESSHFVPPFQD CFGYSGFLAFPPCKITQLVNFCKRIKVAKIFVGFAVNNSNGVLLFQNVFSTKAGFKFYLGLFLSTMFCCFP RTSVTLCLYSLISFCYHKWCCFLISFSDCSLLVYRNTTFFFPYPCILKSCIHLLALIVLLGWGLGVDLS AFSGKHVIKIVLVLHFSFFLHFHTNLARTSGRMLNSSGESRHPCLVTDLRKKASSLSPLSMVLAVGF SMLGFI RLKFPPTFASVFFSFPSNHMIP IWHGTGKQMTNVEMCSRYIKLEMRRPYPDHSTVLTSLIYYFL WSPAATFRDQKKVHPFHLLIKKVSSITVGFVSGFVPLFPWNLKVPGTVEVLVVSRRKSQFRQIPLPPLCAR QCAQYSSPQRDCGGGDETSYKKIIRRDLI VGNRRQLPEAEFPVNKKVFEETGMNRALKEGQLTCVCGST LGRIFDRKLNVLLENFYMKSLKPIITTRDMQSKNNNRVHIRSGNILEFFSDFGLRLKLSKSAFSFCE NLGSCICFPIPIENFRQPKVSDVQVLFVYCSMLLLQVFSVLPARLLILHTDHMWLKITGSHHEQVRAAIW ELTIHHTFIDYPLARYMNDRGSIQADMQISYNNLIDNPREVFFCHSLDVFMHLPIETCERGI IIVPSTDI FEHLILRSKSRANIDVELCSMQSPSPIVLLLI ISEFSVSSGFERPPEPLFHNSTLSSLI PNSTQKIKGE RKVCSTDRNFSIRISTPCDTIQTLLLSAIQWKGPDIQYKKLHVKPCVTSFNLFGAVSWVDLTLEPRCVLQC AVNHPLSQCSNIASGIQQFLSTKDIMVCPHPYPDLAPCDCWLFPKEMTTKGKLFELIQDIKAATTVQL KTLIKEDFQNCFRKWQKQRTDCVQSKGEYFEENWCVFYCNKFFITFTVFELSHLIQKTSRRKLLHFVPP QLQVVTGSAEYNGHMEKQFSWKFWMTKNVFENKVIYHHHLLCRNKTSQSEVPWKKSSGWKIPALSPIT SCDFSHFSLNCHPHDAPS PRVFHDNSVLLDEAFWWGASESPSRARVCVCVCSLYSSEQFYRTWRRHQS TQRECSLIMNYVKNISPGPVMVYPDVHSTFMNIEAHLFFPMKIRKLGEKIKRTHSLTPNKYVHWDGFGFSL MTLITFLALILYLSAQLIKLTGWVRIPAAVYQGVVPAHYVTSLPFSLKVEVLIVPASHYCENSIC NTTMEGTVLTSIYMPQSMAEFCQPPHLLCHWTHPTIPGVSGGGXXXXXXXKX
17	Prediction of a potential new spike protein sequence (RFe-SP1) (M) YCSISQLELCWRLTVTGTLQTFTGLDSSLVFEDVNLIRPGHCVFERNSSPSFYNPCFKSTECISVMYH FTDFKSVRLNKRYSGVLTSSLSFATRLGLELSLPVGSRSERDIIGGICWPTEVHLFPFCHQSFTGAQRLF RHLLMGSLISIRIRPLKKEFCISVHAQVVR SINVYHQHAFPAAVRNHEPSFLKLCDRLSQYVCLIPALTA SGGVTSKHQEPGVRTKTERNCFNNLES AVLCRNVFDQSVKNCKKWTVVISFEKFLKWVKMVSCLYCKLR PNFWIKRRI PKKGILHTNIHI IHNHLETANIKSQVPYRLNVS PGYVFASMYSTLTILETHVCGCQASL KNQTSWFLITYCAVFP RSDMCSLVCCPHLGSSLTLTVTFVSI STGGYDSCLIVYEQVLSIHSRKSNA LIGEYHCGHLQSTPLGKCLTDASASTLQSNFGTLFLEWSSELSSCYPCECHQNVFLSIFPLSSGKERRH WGPGEVSRREGVPRLFVWCWLT PSQLVAGVLSCKIHELLEKRSGHFRLKFTTQPFLLHFIVNLVNCSSWY KFIMNPNPSDIKKQQHRLDPFGLMELEFSSWRIGLPFCTLTFCYRIILVHPCFITSDNMANVMLPLQKVLT NLDSLLFLFTGELLRDHFCHLPHAKLFSSDFVELYFYLGLLCSSESANNFDGSFDEFLQFFSSVLLVIG CPDLSLSVARSLPSEKTFTPLVHCHFILGIIILIDLHVPDFTSTLARFTKKFNVCSLFFKL RHSCDATQK HTTIMNNAIQHRHMSTRQTDLTYKVMKPYRAVCTVLEMTQGGKFSNLILRPHILNISRFISQSPFV DLLVEYLNFFIVLYRGTVCFSSRAHQVIALQSPCGGHSSAPSPVAVFQSLVAGGAHHPMRELNQPPCC ELTPVPTPELGHPNKTYCLFQYRKLEKRNKHSYSQYPKTKTQHCFTFISLCKLLFISLHYTFENQIVSL AMISPCLELVFLYCALLNIGILQLLTLPFSHSISICPAFSLHADKSQINIFYIHYFLI IKSIFPPYSI IHVDVFKFKHPCLPFLYSLQISLIASKRKS KSNCEVETKTKNSKPFGRIRLVGCRSSKSHSCAISVLL VPSHFSFFFLISPECSFVTNDYHHLPLKHASRFPGRKELCSFCFYFPYSISTKVLTVTQIPTAKVTC PTLNQIRPERCNEILLCLAVSHHRAHGYL PRAESGGKRDRTSNETQSCCQNKANGCQELTNNTPSFIEWH SLIQFQNLASCETTLPLPLPSHLFVFSCLPLSLTRTLEITMDPHRYKTI SPSQSFNHLKYHFAVSNMLT YFRDDHILRGHYFACHTHIFLNHLHNLILEMCSGIMFILGCFSLLAHSVSFTLALNTHSVPHATLVSHAK QLFIHFAIMPNSVWHNQGNLFSPLTINLKAFLIPKSQINLKVLLSESQNYFTFERNLAKVHFTFISNKPI PLNFSIYNDVPLFLVNETKHNTFLKRVKKKKSLKLLGYFCYNEVSSASERSRRKNNTVDITEQLIHILT TVCLNFTLCHIDPNVPSSASLRALFSGHSESILKSSSTHPSADAMGSLPASMALWLKDISVQSPPLYPP FHWKLSLPLHKCEPEEIIICQEGFLMVNAQILNRVQPF LTQASRTLCSIGKSHRLKFPPELCSCCCLSQP LSGTSWNWSWNFQRTETSPIQSELWRYILICIYTDQSELIHNNQSEHMMMLTDQNCVIWISHLHKNGPNGNQ GTNFLCKRPLPLCLGVHFRFFLETNCSAESLSSFHTSYWKL LLLIGRLWSQFTRGPLKLSMIFKQEDSSR FEGRGYCPRSSLSRSEVCQKREKAVMALVIFQWKQYISLVERVRVSNREDAAGQISPVSHYRQDSTRSSQK TKWYGCEECIFLFFFFLFFFFFLVCI FLINLLGVACHSMKTPGLLVSQCLIFDQLEMLLNHVFHRKQM QSSSHLLHSSNPSLYLSDDVYICVCGALDSACQNNFTFDWTDLLVYVLP LTFHVSIEYKKQKALGFA
18	ORF number 1 in reading frame 1 on the direct strand extends from base 610 to base 837 TCTCACCTAGCAGGAAGGscadmtctcaggaccatcccatcacgagggtggaggattgggtgga tcagggtacataggcccaatacgtctgtgtctctctctctgcattgctgagggtcatcaatgtcatca

	gcaggtagaggggtccacatgtcgcaccaatcggttctggcagccaccgagggccttctgcatcctgtggaaaaacacaaacatgccctcgcccccatatga
19	Translation of ORF number 1 in reading frame 1 on the direct strand SHLAGRXXLRITIPYSRVEDWWIRYIGPIRLVFFCIAEGHQCHQQVEGPHVAPIVLAATEGLLHPVEKHKHALGPI
20	ORF number 2 in reading frame 1 on the direct strand extends from base 3349 to base 3699 Ccttgatgcccattgtaagagtgtgtggagcgcttttggcatccttctgtgccccctcagggcttgggtcaaatggaaagacccacttacaggctccttggaaggccagatccagtcctcatatggggcgagggcatgtttgtgtttttccacaggatgcagaaggccctcggtggctgccagaacgatgggtgcgacatgtggaccctctacctgctgatgacattgatgascadmccctctgcatcctgtggaaaaacacaaacatgccctcgcccccatatgaggactggatctgggccttgccaagagcctgt aagtgggtcctttccattgaccaaagcctga
21	Translation of ORF number 2 in reading frame 1 on the direct strand PWMPMVRVLWSAFGILLPLRLWSNGKTHLQALGKAQIQSSYGAEGMFVFFHRMQKALGGCQNDWCDMWTLYLLMTLMXXPSASCCKTQTCPRPHMRTGSGPCQEPVSGSFHLTKA
22	ORF number 3 in reading frame 1 on the direct strand extends from base 4186 to base 4740 aggggtccacatgtcgcaccaatcggttctggcagccaccgagggccttctgcatcctgtggaaaaacacaaacatgccctcgcccccatatgaggactggatctgggccttgccaagagcctgt aagtgggtccttccatttgaccscadmaattgaaagaccacttacaggcttttggcaaggccagatccagtcctcatatggggcgagggcatgtttgtgtttttccacaggatacagaaggccctcggtgctgccagaacgattgtgtgcgacatgtggaccctctacctgctgatgacattgatgaccctcaacaatgcagaagaagaccagacgtattggcctatgscadmATCCCATACAGCAGGGTGGAGGATtggtggatcaggtacataggcccaatacgtctgtgttcttctgtattgctgagggatcatcaatgtcatcagcaggttagaggggtccacatgtcgcaccaatcggttctggcagccaccgagggccttctgtattctgtggaaaaacacaaacatgccctcgcccccatatga
23	Translation of ORF number 3 in reading frame 1 on the direct strand RVHMSHQSFQPPRAFCLWKNTNMPSAPYEDWIWALPRACKWVFPFDXXNGKTHLQAFGKAQIQSSYGAEGMFVFFHRIQKALGGCQNDWCDMWTLYLLMTLMTLNNAEEDQTYWAYXXIPYSRVEDWWIRYIGPIRLVFFCIAEGHQCHQQVEGPHVAPIVLAATEGLLYSVEKHKHALGPI
24	ORF number 4 in reading frame 1 on the direct strand extends from base 4792 to base 6306 ccaaagccscadmtcgaatttccagagcctctgaaaagatatcagtgggcgagtccttccccaaagacatggcaaatagccccaccttgtgtcagaagttgttagtaaaacaattgataacaccagaaaacagtttcccttctgtgtacattattcattatatggatgacattttattggcttgtaagaaagaaagagattgttagcttctgtgcaaatctgcaaaagaatcttctaaccctcggtcttattattgcacccgaaaaaatacagagaagtgagccttgttcttacttgggatttcagttgtttgctcagta tttcactccacaaaaaaagagccttagaaaagatcatcttaaatctcttaattgattttcaaaagtgttgggagatattaattggctgcacccttcttgggattaactactggagatcttaaccactgtttgaaatttttaaaagagattctgatccgacctccccaggtctcttactgagcctgcacggaaggtctctcttaagttgagaaagccattcagcaacagcatgtttccttttttagattattctaaacctctatatgtgtatatatttttagatacacaacacgcccacggcggtgttatggcaagaaggccacttagatggatacacctccacgtggtgctcaaaagaatcttactccttattgaactgtggccagtttaattcaggagagtcgcttagaagctcgaataattatggaaaggagccagat tctattgttatcccttttacaaaaatgcagattcaaggcctgatgcagtttacaaaacagttttcctatgccttggctcattttgctggggactttggataatcattatcctaagcataaattgcttcaattttttcaacatcatgatccaattttccttcaattgtgtcccatgctcctctcctgctgtacctaattgtttttactgatggatctagcaatggtgttagctgtctatgcactcaatgaaaaagtccaagagagtgagacacctccagcctcagctcaaatgttgagcttcgagcagttcatatggattgtctgattttgcttccagtccttttaatttattctctgacagccattatgtgttgcgtggtcagagaaatttagaaaacagtaccttttattagaccagtaaatcctgttattcaggatctgtttc

	ttcagatacaacaagccattcagctgcgctgtaacaaatatttattgacctatattagagctca ctctaattcttccagggcccttttagcctcaggaaatcaaactgcagattctgccacacagctcatt gttttaactcaaatagaaaaggcacaaaaggctcttagcttccaccatcaaaacaaccagagct taagactgcaatataactataactagagaaacagcacgccagatagtaaaacaatgccagattg ttcgcatttacagcctgtgcctcattatggagtcaccccttga
25	Translation of ORF number 4 in reading frame 1 on the direct strand PKPXXEFPEPLKRYQWRVLPQDMANSPTLCQKFVSKTIDNTRKQFPSVYIIHYMDDILLACKKE GVLLACFANLQKNLLTSGLIIAPEKIQRSEPCSYLGFQLFAQYFTPQKKELRKDHLKSLNDFQK LLGDINWLHPSLGLTTGDLKPLFEILKRSDPTSPRSLTEPARKALSKVEKAIQQQHVSFLDYS KPLYVYILDTKHTPTAVLWQEGPLRWIHLHVAAQKNLTPYYELVASLIQESRLEARKYYGKEPD SIVIPFTKMQIQGLMQFTNSFPIALAHFAGTLDNHYPKHKLQFFQHHDPIFPSIVSHAPLPAV PNVFTDGSSNGVAVYALNEKVTKRVQTPPASAQIVELRAVHMLLDFAQSFNLFSDSHYVVRA VRNLETVPFISTSNPVIQDLFLQIQQAQLRCNKFYIGHIRAHSNLPGPLASGNQTADSATQLI VLTQIEKAQKALSFHHQNNQSLRLQYTTRETARQIVKQCPDCSHLQVPVPHYGVNP
26	ORF number 5 in reading frame 1 on the direct strand extends from base 6307 to base 6987 ggcctacgtcctaataatgatttatggcaaatggatgtaacacatatacctgaatttggaaaattaa aatatgttcatgtctccatagacacatttctggcttctgtcggtaccgctcaaactggaga ggacacatctcatgttatttagacattgtcttgcgtcttttgcctatgattggaacacctaataaaa cttaaaacagataatggctcaggttataccagcaaaaaattctctttattttgccagcaattct cgatcaatcatgttactggcattccttacaatccccaagggcaagggttgttaaacgcactca tggcacattaaaagtcaatttacagaaaaataaaaaagggggagttatatccctgacgccccat aattacctgtctcattctctctttatccaaaattttttgaccttgatgcccatggtaagagt ctgcggagtgtcttttggcatccttctactgccactcaggcttgggtcaaatggaaagaccct tacgggtctcttggcaaggccagatccagtcctcatatggggccgaggacatgtttgtgtttt ccacaggatgcagaaggccctcggtggctgccagaacgattgggtgcgacatgtggaCCCTCTAC CTGCTGATGACATTGATGACscadmgttgggtcaaataa
27	Translation of ORF number 5 in reading frame 1 on the direct strand GLRPNDLWQMDVTHIPEFGKLKYVHVSIDTFSGFVVATAQTGEDTSHVIRHCLAAFAMIGTPKK LKTDNSSGYTSKKFSLFCQQFSINHVTGIPYNPQGQGIVKRTHGTLKVNQLKIKKGELYPLTPH NYLSHSLFIQNFLTLDAGKSAACEFWHPSTATQALVKWKDPLTGSWQGPDPVLWGRGHVCF PQDAEGPRWLPERLVRHVDPLPADDIDDXALVK
28	ORF number 6 in reading frame 1 on the direct strand extends from base 7282 to base 7590 TGGACACATAAAACAACATTTGAAAAGTTTGTAAATCAGGCACTCCCTGCAGTCAGGTGACTG ATTTACAAGATGGGACTAGAGACTGGTCTAAAAAATCTGTTAATGTATCtgccttgtgttccad mgttcatcaatgtcatcagcaggttagaggtccacatatcgccaatcgcttctggcagccac cgaggccctctgtatcctgttgaaaaacacaaacatgccctcgcccccatatgaggactggat ctgggccttgccaagagcctgtaagtgggtctttccatttgaccaaagcctga
29	Translation of ORF number 6 in reading frame 1 on the direct strand WTHKTTFEKFCKSGTPCSQVTDLQDGTDRWSKKS VNVSACVXXGSSMSSAGRGSTYRTNRSGSH RGPSVSCGKTQTCPRPHMRTGSGPCQEPVSGSFHLTKA
30	ORF number 7 in reading frame 1 on the direct strand extends from base 8518 to base 8751 GGCGTGAggtgacataatctggaatctcaggaccatcccatacagcaggtggaggatt ggtggatcaggtacatagggccaatacgtctgttcttttctgcattgttgagggtcatcaatg tcatcagcaggttagaggtccacatgtcgcaccaatcgcttttggcagccaccgagggccctctg tatcctgttgaaaaacacaaacatgccctcgcccccatatga
31	Translation of ORF number 7 in reading frame 1 on the direct strand GVSVIDIWNLRITIPYSRVEDWWIRYIGPIRLVFFCIVEGHQCHQQVEGPHVAPIVLAATEGPL YPVEKHKHALGPI

32	ORF number 8 in reading frame 1 on the direct strand extends from base 14551 to base 14847 aggggtccatatgtcgcaccaatcggttctggcagccaccgagggccctctgcatcctgtggaaaa acacaaacatgccctcggcccatatgaggactggatctggcccttgccaagagcctgtaagtg ggtctttccatttgaccaaagcctgagtggcagtagaaggatgccaaaagcgtctgcagcact cttaccatgggcatccaaggtcaaaaaattttgaataaagagagaatgagacaggtaattatgg ggcgtcaggggatacaactccccctttttattttttgtaa
33	Translation of ORF number 8 in reading frame 1 on the direct strand RVHMSHQSFQPPRALCILWKNTNMPASAPYEDWIWALPRACKWVFPFDQSLSGSRMPKALCST LTMGIQGGQKILNKERMQRQVIMGRQGIQLPLFYFL
34	ORF number 9 in reading frame 1 on the direct strand extends from base 15370 to base 15627 ctttggatgcctatgtaagagtgcagctgaacgttttctggcatccttctgccgtccctgagggc tttggtcagaaagaaggatccacttactggatcatggcaaggcccagaccagtcctcatatgg ggccgagggcatgtttgtgtttttccacaggatgcaascadmAGGAGAAACAAGAATGGTGGT GGCTTTATATCGCAGATAGGAAGGAACAGACATTCTGTATCTATGCCATATCATGTCTGTACATT AA
35	Translation of ORF number 9 in reading frame 1 on the direct strand LWMPMLRVQLNVSGILLPSLRLWSEIRIHLDDHGKAQTQSSYGAEGMFVFFHRMQXXRRNKNGG GFISQIGRNRHSYLCHIMSVH
36	ORF number 10 in reading frame 1 on the direct strand extends from base 17263 to base 17661 cattatacccctcaatacctgaacacgtatcttctaagaacaagggccttttacatcagcacaa tacaattattatattcaggaagttaacattgatattgtattattgtctaatatgcaatccgta ttcaaatcttctcaataactccactaatacccgttacagtccttgtcttgttttaagttcagg atccaatcagggatcacacattgcatttggttgccattcctcgtagcacacttcttgcccttt ttctttttaaatcttcatgccattgatatttttgaggcgccagggaaggtattttgtaaatt agcccttaatttgtaattgtctcattggttactcctgattgtattcatcttaaatatttttggc aaaaatacaacatag
37	Translation of ORF number 10 in reading frame 1 on the direct strand HYTPQYLNTYLLRTRAFYISTIQLLYSGSLTLIWYYCLICNPYSNFLKYSTNTRYSLCLVFKFR IQSGITHCIWLPFLVSTLLGLFLFKFFMPLIFLRRPGKVFCCKLALNLNLSHWLLLVFILNIFG KNTT
38	ORF number 11 in reading frame 1 on the direct strand extends from base 18964 to base 19221 ttcagtgctgacactgtctacctggatctgataatatcagatcccacagggtcaagggtcagtc ccacaggacggctgtcccccccttcagatgccaatcacaagtcgcaggttgccacatatataca ccaaatggctataaatcaggggtaccgcgactccctccttgggttcagtaatttgcgggaatgg ttcacagaactcaggaaaacacattaccagtttattatgaaagactatgataaaggatatatat ga
39	Translation of ORF number 11 in reading frame 1 on the direct strand FSADTVYLDLIISDPTGQGLSPTGRLSPPSDANHKSQVVYIYHQMMAINQGTDRDLLGFSNLPEW FTELKTHYQFIMKDYDKGYI
40	ORF number 12 in reading frame 1 on the direct strand extends from base 19894 to base 20241 aggttagatatagatatatttctattatctcacagcatttatcttagaaaataagaacttgggtta GAATGATTGCCTTTCTGGTGAAGTCTATTTTATTTCAACATTCTTTTCATTATTTTATTTTAAA Ataccaaaattaacatgttgatgccttaaatgtgcacaatgttacatgtcaaatacattttttt tttaaaacttttacttattttaagtgtgttttcccaggaccatcagctccaagtcaagtagttt caatcgagttgtggaggcgagctcacagtgggccatgtggggattgaaccagcaaccttggt gttaagagctcacgctctaaccgactga

41	Translation of ORF number 12 in reading frame 1 on the direct strand RLDIDIFLLSHSIYLRNKNLVRMIAFLVKSILFQHHFFHYFILKYQINMLYALNLHNVTCQIHFF FKLLLLLSVFSQDPSAPSQVVSIELWRAQLTVAHVGI EPATLLLRAHALTD
42	ORF number 13 in reading frame 1 on the direct strand extends from base 21031 to base 21306 CATTTTAGAGTATACTCTTTGTGTATGTATCATTTGAAGCACACTCCCATTAGTGTTTACCATT TTACTTGGGATTTTTATAAAAGTCATTCTATGGTGTTAAAGAGATTGTGCTGCAGTATAGTTTC ACTGTGTACTGCAGTCCCAAAGGAAAGGGAGCCAGTAAAGACGTGCCGCTTTTTTCCACAAGA GTACCATATTTCTTAACGTTGGCTATAAAATTTACTTCATGAGTCCCGAAGCAGCAAAATACC TCTTGAAAGTCACATTGA
43	Translation of ORF number 13 in reading frame 1 on the direct strand HFRVYSLCMYHLKHTPISVYHFTWDFYKSHSMVLKRLCCSIVSLCTAVPKEREPVKTRCFFSTR VPYFLTLAIKFYFMSPEAAKYLFESHI
44	ORF number 14 in reading frame 1 on the direct strand extends from base 21622 to base 21849 TGTCTACATTTAATTCTTTGTAGTTGGAAGTTCACGAGGCTAAGCCCGTGCCAGAAAATCACCC GCAGTGGGATACAGCAGTGGAGGGGGATGAAGACCAGGAGGACAGCGAGGGCTTTGAAGACAGC TTTgaggaagaggaggaagaagaggaagatgacgaCTAAGCAGTACTGCAAACGGACCACAATA CTTTCACATTTTCACTGTTTTGGAAGTGTAATAA
45	Translation of ORF number 14 in reading frame 1 on the direct strand CLHLILCSWKFTRLSPCQKITRSGIQQWRGMKTRRTARALKTALRKRKRKRKMTTKQYCKRTTI LSHFHCFGSVE
46	ORF number 15 in reading frame 1 on the direct strand extends from base 22447 to base 22875 ctttggatgcctatgttaagagtgcagctgaacgtttctggcatccttctgcggtccctgaggg tttggtcagaaagaaggatccacttactggatcatggcaaggcccagaccagctcctcatatgg ggccgagggcatgtttgtgtttttccacaggatgcagaaggccctcggtggctgccagaacgat tgggtgcgacatgtggaccctctacctgctgatgacattgatgaccctcagcaatgcagaagaag accagacgtattgggcctacgtacttgatccacctattctccaccctgctgtgtgggatggctc tgagattccagactatgtcaatgacacTCACGCCCTAGGATTGCCTTCTGATGGACACATAAAA CATTTGGAAAGTTTTGTAAATCAGGCACTCCCTGCAGTCAGGTGA
47	Translation of ORF number 15 in reading frame 1 on the direct strand LWMPMLRVQLNVSGILLRSLRLWSERRIHLLDHGKAQTQSSYGAEGMFVFFHRMQKALGGCQND WCDMWTLYLLMTLMTLSNAEEDQTYWAYVLDPPILHPAVWDGPEIPDYVNDTHALGLPSDGHK HLESFVNQALPAVR
48	ORF number 16 in reading frame 1 on the direct strand extends from base 23074 to base 23310 tacttaaacacaccatcttttggtatgcttcctgttaatatctctggaccttggtataactaaaag aaatttgcatgatgttaatgtgtcttttagatatgtttcagcttcatgagaaaattcaaaatc admtcatcagcaggtagagggccacatgtcgaccaaactgttctggcagccactgagggccct ctgcatcctgtggaaaaacacaaacatgcctcgcccccatatga
49	Translation of ORF number 16 in reading frame 1 on the direct strand YLNHLLLCFLILSLDLGILKEIWHDVNVSLDMFQLHEKIQNXXHQQVEGPHVAPIVLAATEGL LHPVEKHKHALGPI
50	ORF number 17 in reading frame 1 on the direct strand extends from base 23362 to base 23859 ccaaagcctgaggggcagcagaaggatgccagaaacgttcagctgcactcttaccascadmctg gcattccttacaatccacagggacaagggattgttgaaacgactcatggcacattaaaagtcaa tttacaaaaaataaaaaagggggagtcatatcccctgacgccccataattatctgtctcattct ctctttattcaaaattttttgaccttgatgccatggtaagagtgctgcagagcgcttttggc

	atccttccactgccactcaggcttttggtcaaatggaaagaccacttacgggctcttggcaagg cccagatccagtcctcatatggggccgagggcatgtttgtgtttttccacaggatgcagaaggc cctcgggtggctgccagaacgattggtgcgacatgtggaccctctacctgctgatgacattgatg accctaagcaatgcagaagaagaccagacgtattgggcctatgtacctga
51	Translation of ORF number 17 in reading frame 1 on the direct strand PKPEGQQKDARNVQLHSYXXXGIPYNPQQGIVERTHGTLKVNQLQIKKGESYPLTPHNYLSHS LFIQNFLTLDAGKSAERFWHPSTATQALVKWKDPLTGSWQGPDPVLIWGRGHVCFVQDAEG PRWLPERLVRHVDPLPADDIDDPKQCRRRPDVLGLCT
52	ORF number 18 in reading frame 1 on the direct strand extends from base 23947 to base 24384 tggacacatgaaacaacaTTTGAAAGTTTGTAAATCAGGCACTCCCTGCAGTCAGGTGACTG ATTTACAAGACGGGACTAGAGACTGGTCTAAGAAATCTGTAAATATATCTGCTTGTGTTCTTC CCCATATACACTTTTGATTscadmmttggtcaaatggaaagaccacttacaggctcttggcaag gcccagatccagtcctcatatggggccgagggcatgtttgtgtttttccacaggatgcagaagg ccctcgggtggttgccagaacgattggtgcgacatgtggaccctctacctgctgatgacattgat gaccctcagcaatacagaagaagaccagacgtattgggcctatgtacctgatccaccaatcctc cacctgttgtatgggaaggtcctgagattccAGTscadmaataaaaactataa
53	Translation of ORF number 18 in reading frame 1 on the direct strand WTHETTFGKFCCKSGTPCSQVTDLQDGRDWSKKS VNISACVPSPYTLLIXXWSNGKTHLQALGK AQIQSSYGAEGMFVFFHRMQKALGGCQNDWCDMWTLYLLMTLMTLSNTEEDQTYWAYVPDPPIIL HPVVWEGPEIPVXXIKL
54	ORF number 19 in reading frame 1 on the direct strand extends from base 24625 to base 24948 cgccccataattacttgtctttttattcaaaatttttgactttggatgcctatgttaagagtg cagctgaacgttttctggcatccttctgccgacctgaggctttggtcagaaagaaggatccact tactggatcatggcaaggcccagaccagtcctcatatggggccgagggcatgtttgtgttttt ccacaggatgcagatagtcctcgggtggctgccagaacgattggtgcgacatgtggaccctctac ctgctgatgacattgatgaccctcagcaatgcagaagaagaccagacgtattgggcctacgtac ctga
55	Translation of ORF number 19 in reading frame 1 on the direct strand RPIITCLFIQNFLTLDAYVKSAAERFWHPSADPEALVRKKDPLTGSWQGPDPVLIWGRGHVCFV PQDADSPRWLPERLVRHVDPLPADDIDDPQCCRPRPDVLGLRT
56	ORF number 20 in reading frame 1 on the direct strand extends from base 25126 to base 25380 ACCACTGTTGTAAAACTGTTAATATATCtgccttgtgttccttcccccttatatacttttgatta aaaatattaatgtacacscadmagaacaggtctggggatattttccccaggggtcatagatttac ctgtactccacaaaaaactacaaaggcaataatttgaaaacagatacacctgtgtggataga tcagtggcccccttacacagggaaaagatatcgccgcccaggcgcttgtacaggagcagcttga
57	Translation of ORF number 20 in reading frame 1 on the direct strand TTVVKTVNISACVPSPYILLIKNINVHXXEQVWGIFPRGHRFTCTPPKNYKGNLLENRYTCVDR SVAPYTGKDIGRPGACTGAA
58	ORF number 21 in reading frame 1 on the direct strand extends from base 28306 to base 28737 ccttgatgccatgtaagagtgtgcagagcgcttttggtatccttctactgccactcaggc tttgggtcaaatggaaagaccacttacaggctcttggcaaggcccagatccagtcctcatatgg ggccgagggcatgtttgtgtttttccacaggatgcagagggccctcgggtggctgccaagacgat tggtgcgacatgtggaccctctacctgctgatgacattgatgaccctcagcaatgcagaagaag accagacgtattgggcctatgttcctgatccaccaatcctccaccctgctgtatgggaaggtcc tgagattccagactatgtcaatgacactcacgccttaggattgccTTCTGATGGACACATAAAA CAACATTTGGAAGTTTGTAAATCAGGCACTCCCTGCAGTCAGGTGA

59	Translation of ORF number 21 in reading frame 1 on the direct strand PWMPMVRVLQSAFGILLPLRLWSNGKTHLQALGKAQIQSSYGAEGMFVFFHRMQRALGGCQDD WCDMWTLYLLMTLMTLSNAEEDQTYWAYVPDPPIILHPAVWEGPEIPDYVNDTHALGLPSDGHK QHLESFVNQALPAVR
60	ORF number 22 in reading frame 1 on the direct strand extends from base 30907 to base 31191 cttttgatgcccattgtaaaagtgcagctgcacgttttttggcatccttcaactagccctcagg ccttggtcaaattggaaggacccacttacgggtgtctggcaaggcccagatccagtcctcatatg gggcccaggggcatgtttgtgtttttccacaggatgcagatagtcctcggtggctgctagaacga ttggtgcacatgtggaccctctacctgtgatgacattgatgaccctcagcaatgcagaagaa gaccagacgtattgggctacgtaccta
61	Translation of ORF number 22 in reading frame 1 on the direct strand LWMPMVKVQLHVFVWHPSTSPQALVKWKDPLTGVWQGPDVLIWGRGHVCFVFPQDADSPRWLLER LVRHVDPLPADDIDDPQCCRRLPDVGLRT
62	ORF number 23 in reading frame 1 on the direct strand extends from base 31279 to base 32070 TGGACACATGAAACAACATTTGGAAAGTTTGTAAATCAGGCACTCCCTGCAGTCAGGTGACTG ACTTACAAGACGGGACTAGAGACTGGTCTAAGAAATCTGTTAATGTATCTGCTTGTGTTCTTC CCCTTATACACTTTTGATTGAAAATATTAATGTACATTTTGTAGGAGTTCAGTTTAtggaagat gtgattcagagtataaaagttaaattctatttagaatgtcattcagaatatcattggatacgtg ttacttctaaaaggtataataatagtcaatatgattggaatcgggttcggtttacatcttcaagg aatttggcatgatgctaattgtgtctttagatascadmCGAGGAGTGCAGATAGAGCCGGCGGCG GCGGCGCAGCGAGCGAGCAGTGACCGCGCTCTACCCAGTTCTGCCCCACGGCTCTACCTGCT TGCTTCCCTCAGCCCCCTCGCCCGCTGTGACTAACC CGCAGCATGATGTTCTCCAGCTTCAACG CCGACTACGACGCGGCTCTTCCCGCTGCAGCAGCGCTCCCCAGCTGGGGACAGTCTCTCTTA CTACCACTCACCCGCGACTCCTTCTCCAGCATGGGCTCTCTGTCAATGCGCAGGTAAGGCTG GCTTCAACCGAGCCAGGCTCGGGGTCACTGGGGTGGAGGCATCGGGCGGGAAGCTCAGGAAGA CGAGTCGGGTACCCCTTTTGGCGGGGAGGAGCAGCCCTAACTCGCGAGTCCCGGACTTGTGGG GCGCTCACACACGCTTGTCACTAA
63	Translation of ORF number 23 in reading frame 1 on the direct strand WTHETTFGKFKSGTPCSQVTDLQDGRDWSKKS VNVSACVPSPYTLLENIN VHFVGVQFMED VIQSIKVKSYLECHSEYHWIRVTSKRYNNSQYDWNVRVRLHLQGIWHDANVSLDXXRGVQIEPAA AAQRASSDRAPTQFCPTAPTCLPPSAPRPAVTNRDHDVLQLQRRRLRRGLFPLQQLPSWGQSLL LPLTRRLLLQHGLSCQACAGKAGFTEPRARGHWGGGIGREAQEDESGTPFGGEGAALTRESRTC GALHACQ
64	ORF number 24 in reading frame 1 on the direct strand extends from base 34747 to base 35073 CAGACCTCCTGCCCTGGCGGATGCCATGGATTCCAGAGCCCTAGTCTCCACCCCTCACTGTGCG CAGGACAGTCTGGGCATGTTTGCACATGCTCCTGCTGCACAGGGCACTCTCTCGTAATGTATCT CAGAGTTCAGTCCCATAGATGGCCTTATAACGTAAGTACTCTTCTAAGCACTGAAGGACATTAT CATCCACTTTGGGGTCAAACCTTGTGGCCAACAGGTGAGGGTTACGAAGAATCCAGTGCAGGTC CCCAGCCCCATAAATGCAGATACCCCGCTGGTGGGTTCAGAGCAAGGTCCATAAGGTGCCCC TACTGA
65	Translation of ORF number 24 in reading frame 1 on the direct strand QTSCPGGCHGFQSPSLPPLTVAGQSGHVCTCSCCTGHSLVMYLRVQSHRWPYNVSTLLSTEGHY HPLWGQTCWPTGEGYEESAGPQPHKCRYPAGGFQSKVHKVPPY
66	ORF number 25 in reading frame 1 on the direct strand extends from base 36097 to base 36516 GAGAAAGTCTCAGAGCGACAATGGCCAGCAGGAAATAGCAGCCCAGAGCCCACAGGTAGTGCTT CTGGAAGAGTTTCTTCTTCCACCAAATCATCTTCATGGAATGGAAGATCGGTAGAATTTGGGCA CCAGGAAGAAGAAGGATGGGATCCTTscadmACCCTGGCCCGGGGGCGGCGCACCGTCCAC

	CGGTCCGGGGCCAGCGGGGCCGGGGCCGGAGTCGGCATGAATCGCTGCTGGGCGCTCTTCCTG TCTCTCTGCTGCTACCTGCGTCTGGTCAGCGCCGAGGTGAGTTGCGACAGCCGTGGGGCTGGTT CGCTTCATTTCATTGCCCCACCCCATCCCTGTTGCCCCCTCCCCTCCCTGCAGTGAACTTTGG ACCCTTGACAGCCCGTGGGCTTGGCGCCCGCGCTAG
67	Translation of ORF number 25 in reading frame 1 on the direct strand EKVSEKQWPAGNSSPEPTGSASGRVSSSTKSSSWNGRSVEFGHQEEEGWDPXXTLAAGAARTVH ASGAQRGRARSRHESLLGALPVSLLLPASGQRRELRLQPGWGFASFIAPTPIPVAPSPPCSELW TLAARGPGARR
68	ORF number 26 in reading frame 1 on the direct strand extends from base 36649 to base 36957 TCTTATCCCCACCTCCTCAGAAACCCAGAATAAGCCCCTAACTGGCCTAAGGGAGAGGGGGT GGGGTGGTGCCGAGGGTGCAAGGCGGCGCTCCTTCCAAGCCCACTTCAGTTCCAGCTTAGG TTCTGTCCGGGAACCGGCTTGACGGAAGGTGCGAGCTCGCGCACTGGTGGCAGCCACGCCAAC CTACGGCAGGGGTTTTCGTCCACCCCTGGCTCCCGCTCCAGCTCTTGCTTGCTGGCCCCAGAG CGTGGTGACAGGAGCAGCTTGTGTCTTGGGCGCGGGGGGTACAGAGAGATAG
69	Translation of ORF number 26 in reading frame 1 on the direct strand SYPPPPQKQNKPLTGLRERGWGAEGAEGGASFAQHFSSSLGVSREPACTEGASSRTGGSHAN LRQGFASHPGSRSSSCLLPRAWCRSSSLCLGRGGTER
70	ORF number 27 in reading frame 1 on the direct strand extends from base 37270 to base 38031 GGTGAAGAGGCTCAGGGGCTGCGAGGAGCGCTACGCGCCTGGTCCCGTCCCGCCTCAGCTCGGC GGCCCGCGGGAGCCCGCACCGAGCCGGCTCCTGGGAGGGCCGGCCCTCTCGGGCCTCCAACGA GGAGCAGGAAGGAGGCGGCGGCGGCGGAAGGGGTTAAGGTGAAGGGCTTCGAGGCCGCGGCC GGGCTTGGGCGGACGACGCGAGGTTGTTTGACACGAGGAGCGCTCTCCGTCTCCTTTT GTTCTCGGGGCTCCTCGAGGGCGCGCGGCGCTCCGCCCTGGGGCCCCGCCCTTCCGCGGCCGTC CCCCGTGGCCCGCACCCGGGAGGGAGGACGCGGGGATCAGCCTGGCTGCCTGCAGTCCCCTCCC GACGCCCCCTCCTCTCCTGCTGATGCCCCCGGGCCGCGCCAGCTGTTGGGGCGGGGGGCG GCCGGCCGGCCCCAGCTGCCGCTCGCCGCGggggcctggggcctggggcctgtgccagggcgctc ctgggAACGGCGGCGCCCCAGCCGCTGCTCTCCGACGCCACCCCGCCGGCCCCCGACTCGC TCACTACCCACGCATGCACACTCTTGCCGAGGCGATGCTGCGCTCCGGCGGGCGGGCGCG CAGGGCGACGGGCACGCACTGGCGCGGCGGGTcgcgcgcccgccgacgcccgtgcacatgc gggacacacgcgcgcgactacacacacacacgcacatggtccccgcacacacggttga
71	Translation of ORF number 27 in reading frame 1 on the direct strand GEEAQGLRGALRAWSRPASARRPPGARTEPAPGRAGPSRASNEEQEGGGGGGEGVKVKGFEAAA GPWAAAASAGCFDHGGAVSVSFCSSRGSSRAAGRPPWGPALPRPSPVARTREGGRGDQPGCLQSPP DAPSSPPADAPRAAASCWGGRRPAPAAASPRGLGAGPCARASWERRRPSRCSPOPTPPGPPTR SLTPRMHTLGRRRCCAPAGGRAGRRARTGAAGSRARRHARAHAGHTRAHYTHRMVPAHTA
72	ORF number 28 in reading frame 1 on the direct strand extends from base 38401 to base 38718 GCTTCCGAGGGGGTCCCACCCCCACTGTCTGTGCTCTTTGCTGATCCCAGCCAGCACGCTG CAGAGAGGCTGGGTGACAGCTGGATAAGGCTTTCCCGCCTGCCCTTACCATTCCAGCTTCATC CAGCACCTCCTCCTCTTTCCCACTCCCTGGGTGTGTGTTGGGGGGTGAGCCTATGGCAC AGAACTGGTGCCTGTCTCCTCACTTTAATCACAGCATCCTTGGACACATGGCTCTCAGGAACC CACAGTTGTGTGGTGTTCAGTTTACGAAGCACTTTCCTGCTAAGCCTTACTCTGAGTAA
73	Translation of ORF number 28 in reading frame 1 on the direct strand ASEGAPTPHCSVLFADPSQHAAERLGDWSIRLSRLPLFPASSSTSSSFPTTPWVCVWGVSLWH RNWCLSPHFNHSILGHMALRNPQLCGALQFTKHFPAPKYSE
74	ORF number 29 in reading frame 1 on the direct strand extends from base 39607 to base 39849 TCCTCAGCCCAGGGTAGCCTAGAATGGCCACACTGCTCTTACCAGGCATCCTCATTCGAGCCCC CCCCGGCCCCCATCTTGAGAGACAAGCATATCTTTCTTTTCCATGTCTTGGGCTGCCAATATT

	GGACAGGACAGAGGGGAAGAAACAGAAGGAAAATCAGATCGCAAGGCTTCTGTGTATCTTGAGC AGGCCTGGGCCTCAGTTGCCGCCGCTGAGAATATGAGAAGGTTGGATTAG
75	Translation of ORF number 29 in reading frame 1 on the direct strand SSAQGSLEWPHCSSPGILIRAPPAPHLERQAYLSFPCLGLPILDRTEGKKQKENQIARLLCILS RPGPQLPPRENMRRLD
76	ORF number 30 in reading frame 1 on the direct strand extends from base 41215 to base 41634 gCAATCCAGGTTTCCTTGGCAGCTGAAGCTCTACAgtttctctgcctctccactgttgacattt ggggccagacagtttcttgattgtgggggaggtgtcctgtgcatagtaggatgttttagcagcaa ccctggcctctacctactagacaccagtagcaggcctccagttgtgataaccaaaagtgcctcc agacctggccagtgccccctgggggtcaCTCACTCCCTGCTCTATGACCTCCACTGGGTGAAGA GTGGACCTGAACTGAAAACAGTCCATGAAAGAGGGAGGGGCCGCTCTGCTCCTTACCAGTCGT GTTGACCTTTAGCCATTACTTAATTTTCTAAGCCTCAGCTTCCTCATTGGAAGACAGGGAT ACAAACAGTGACAGCCTCTTGATTGTATTGATTGA
77	Translation of ORF number 30 in reading frame 1 on the direct strand AIQVSLAAEALQFLCLSTVDIWGQTVLDCGGGCPVHSRMFSSNPGLYLLDTSSRPPVVITKSAS RPGQCPLGVTHSLLYDLHWKSGPELKTVERGRGPLCSLPVVLTFSHLLNFSKPLPHLED RD TNSDSLIVFD
78	ORF number 31 in reading frame 1 on the direct strand extends from base 41872 to base 42114 GGGAATCATGCAGGAGAGAACCCAGGGAGAAGGGGAGAGTCCCTTCATGCATTTTACCAGTGTT TAGTGAGCACCTACTCTGTGCTTTCCCCAGTCTCTGTCTGGGCTCTTCCCCGTGCAGGCTGG GAGGGTGGGGTCTGGGTTTGTTCCTAAGACATCATCGTCTCTTTTTTATTATAGGCCGGGT CCAGGTGTCTCACTGGGCCAGCTGGGATCTGCCTACTCTGCCATGGCTAG
79	Translation of ORF number 31 in reading frame 1 on the direct strand GNHAGENPREKGRVLHAFYQCLVSTYSVLSPSLCPGLFPVQAGRVGFVWCFHKTSSSLFYRPG PGCPLGPAGICLLCHG
80	ORF number 32 in reading frame 1 on the direct strand extends from base 42115 to base 42393 CAGCTGCAGCCAGCTCTCCAGTGGGCAAGGAGGTCTTGGCATGAGTGTTACGTGCCATTTGGTA CTGGGTCTTCAGTCCGCTCTCCTAAGAGGTTAATTGATTATTATGCCACAAACAGCCTGGGAG ACCTGGCTGGGCACCCCACTTCGGCTTCCTCTGCTGCTGCCTCTCCTGCCAACCCAGACAGA ATTAGAATTAATAATCAAATCAAATGGCTACAACCCCTCAGTTCACAGGTGATAGCCAGGACCC GAGAGGGGCAGCAACCAACCTGA
81	Translation of ORF number 32 in reading frame 1 on the direct strand QLQPALQWARRSWHECYVPFGTGSSVRSRKRLIDSLCHKQPGRPGWAPPLRLPLLLPLLPDPDR IRIKIKSNGYNPLSSQVIARTREGQQPT
82	ORF number 33 in reading frame 1 on the direct strand extends from base 44644 to base 44922 AGGGTGGGGCTGTGGGAGGGGAGGCAGGCAGGGAGAAGGTGCCAGGGCATCTGCACCCTGAGT ATCCAGGTGTGGACTCAGCCAGGGAGGGTGGTGTGGAGGAGCCACCTCCCTGTCTCTGGCC AAAGCCCCGTCTACAAGGTCTCCCGGGACACCTGGCCGGGACCAGTGGGCAGCCCTGCCCGT GCCCAAGAGGGCACTCAGAGAATGGGCACGTGCTTGGTGGCACACACGTGGCAGGGCTGGCGGG CTGTGTCGGGAATGTATTTATAA
83	Translation of ORF number 33 in reading frame 1 on the direct strand RVGLWEGRQAGRRCPGHLHPEYPGVDSAREGGAGGATSLSLWPKARSTRSPGDTWPGPVGSPAR AQEGTQRMGTCLVAHTWQGWRAVSGMYL
84	ORF number 34 in reading frame 1 on the direct strand extends from base 44923 to base 45165 ACGCTGTCTTCAGAGCAAATTCCATTCTATTCTAACCTCTGGCCTGTTCCCTGGAGCCCTGGTC

	AGCACCCCCCTGCACCCCCAGCTCCCCCTCCCTCTGGGGTTTTGTCTCTTTGTCACTTTGTAAT CCTTGCCCAGACTGCTATCTACGGGGGACAGCATTTCTGCCTTTGTTTCCTCTCCCAGTTGGG CCCCTGGCTCCCTCTCAAAAGCATTTCCCGGGCCCTTTCAAACCCGCCTAG T
85	Translation of ORF number 34 in reading frame 1 on the direct strand TLSSEQIPFYSNLWPVPWSPGQHPAPPAPLPSPGVLSLCHFVILAQTAIYGGQHFLPLPLPVG PLAPSQKHSPGPFKPA
86	ORF number 35 in reading frame 1 on the direct strand extends from base 45313 to base 45786 CTTTTCTGCTGTTTCTTTCCAAGTCTTCGCCCCACCCTCATATTGCCCTCCACACCCCGG GTGGGGTTCGGGTCGGAGAAGACGAGGTTTCAATAGCAGGCCTGTTTCGAGGCAACCATGTGG CTATTTTTTCTCAATCAACTTAACCTTTCCACAAAGCACATCTTTCCCCATCTCTCCCAACC AGGGACATTCCAGAAATGGCAGAGAGAAAGGAATGGAGCCAGAGGGACAGACAGACACTGTT CGTGGGACAATAGGCTAGACGGAAGTGCATCAGTTTTAGGAAAGTCTGCTCTAAACAGGGCCCC TTGGGAGCCCACAGGGACGAGCAATAGTTTTGTTCATGGGCAGTGGCAGTGGGATGGGGAGACAG TGTGACCCTGAGATGCTGTGTGGAGGGGACAGAGCTTGTCCCCGACACCCTTCAGTGTATTTG CTGGCTTTCAGCCATCAGAGAGCTAG
87	Translation of ORF number 35 in reading frame 1 on the direct strand LFCCFFPFSFAPTLILPLHTPGGGRVGEDEVFNSRPVSRQPCGYFFLINLTFPQSTSFPHLLPT RDIPEMAERKEWSQRDRQTHCSWDNRLDGSASVLGKSALNRAPWEPTGTSNSFVMGSGSGMGRQ CDPEMLCGGGQSLSPFPFSVFAGFQPSSES
88	ORF number 36 in reading frame 1 on the direct strand extends from base 45787 to base 46023 AAGAGTCTGCCCACCATTCACGTCAGCTCAAAGTTCCCCTGTCCAGCCCTCACTTTCCGCAG CCGGCTTCCGGCTGCCCTCTACCCAGAGGGATGTCTCCAAGGAGTGCTGATGGTGCTGAGATGAG GGCCTCCAGGCTAGAGAAGGGAGCTGTAGTTGTGACCTTAGGAATAAATGTACAGCTTAGGGCA GGCATGGGGCAAAGGTCAGAGGGAGAGAGACAGAAACACAATGA
89	Translation of ORF number 36 in reading frame 1 on the direct strand KSLPTIQRQAQSSPVQPSLSAAGFRLPLPRGMSPRSADGAEMRASRLEKGAVVVTLGINVQLRA GMGQKVRGRETEQ
90	ORF number 37 in reading frame 1 on the direct strand extends from base 46072 to base 46383 GGGCTCCCCTATCCAGCAGTTCCAGctccctacctctctctgccttttagtccccacccccacc caccacccccctctccttcccaccctctctcccgcccaacTGAACCATTTGTCAGGGGCTCCACA GGGGCTGTGTCCAGGGCATGCTGGTCCCCCTGGGGACTATGGGAATTTCTCCATTGAGCACTT CCTATGGGAACGCTGGGTGGAGGGGCACTGGAAAGTGGCCTCAGAGCTCTGGGTCCTTGCCCTG CCCTGGAGGCCGAGGAGGGTTCGCTTACAGTAGCAAAAGGGAACGGTTATTTTTAA
91	Translation of ORF number 37 in reading frame 1 on the direct strand GLPYPSSSSSLPLSAFSPHPTPPHPSPSHPLSRPTEPLSGAPQGLCPGHAGPPWGLWEFLHSAL PMGTLGGGALESGLRALGPCPALEAEEGSLTVAKNGYF
92	ORF number 38 in reading frame 1 on the direct strand extends from base 46576 to base 46890 GGGGCAAGAGGCAATCCTTCCGTCTGTCCAGAGCCCCCACTGGAGTCCCCAGCCCGTGGTATG ACCAGCCAGCACTTGTACAGTGCTTCTGACTGTGCCTTCTCTGCAGATGAAGACGGGGCTGA GTTGGACCTGAATTTGACTCAGTCCCATTTCTGGAGGCAAGCTGGAGAGCTTATCCCGAGGGAGA AGGAGCCTAGGTAAGAATGAGGGTGCAACGGGGGCCCTCAAAGGTGGGGGCCAGGGAAGAAG AACTGAGCACACAGCTGCCGGAGGCTGTGAGGGTGGGCCCTGTTGTCCCACTTAG
93	Translation of ORF number 38 in reading frame 1 on the direct strand GARGNPSVCPRAPTGVPSWPYDQPALVTVLLTVPSLADEDEGAELDLNLTQSHSGGKLESLSRGR RSLGKNEGANGGPSKVGAREEELSTQPAAGCEGGPCLSHT

94	ORF number 39 in reading frame 1 on the direct strand extends from base 47176 to base 47406 GGGCTGTGGCCTGGGACAGCAGGCATGGAGCAAGCCTGGGACCTGCCTCCTGCTGTACTGCAGA AACC AAAAGGAGAATGTAGATCAGGGAAGGCAAGTGCCCACTCCACGCCCCCTCTCCTCTGTGC CCACCTGCAGTCCCCAAAACACTGTAGACAGTGGCTGGGGGGCCTCCAGGTAAGAGTCAGTGGC CTGAGTTCCTCTTTGCTCTGTGAATTTGGGCATCTAA
95	Translation of ORF number 39 in reading frame 1 on the direct strand GLWPGTAGMEQAWDLPPAVLQKPKGECRSKGKASAHSTPLFLCAHLQSPKHCRQWLGGGLQVRVSG LSSTLCSVNLGI
96	ORF number 40 in reading frame 1 on the direct strand extends from base 47863 to base 48297 CAGCCAGTACTTAAACCTCCCCTGACGTGGAAGGAAGAGGGAATCGGCCAACCGTTTTGGAGG TGCTCCACTTCTGCGGCTAGGGGCCCTGAGCAGCCCTCCACCCCACTCCTGACGGAGTTCCC TCTCCCTTCAGATTCCAGCCGGTCGCCGAGCCAGCCATGATCGCGGAGTGTAAACGCGCACT GAGGTGTTTGTAGATCTCCCGGCGCCTGGTTGACCGCACCAACGCCAACTTCTGGTGTGGCCGC CCTGCGTGGAGGTGCAGCGCTGCTCCGGCTGCTGCAACAATCGCAACGTGCAGTGCCGCCCCAC CCAGGTGCAGCTGCGACATGTCCAGGTGTGAGGCCCACTGCCCTCCTGGGCTGGCCCACT GAGAGCAGGGGCTGCCCTCTGGGGCTGGCACTCACGGACCAGGCTCTTGA
97	Translation of ORF number 40 in reading frame 1 on the direct strand QPVLKTS PDVEGRGNRPTVLEVLHFLPARGPEQPSTPLLTEFPLPSDSQPVAEPAMIAECKTRT EVFEISRRLVDRTNANFLVWPPC VEVQRCSGCCNNRNVQCRPTQVQLRHVQVCRPHVPSWAGPA ESRGCPSGAGTHGPGS
98	ORF number 41 in reading frame 1 on the direct strand extends from base 48298 to base 48570 ATGCGTCAAAAGGCATTCTTGGCAGGGTGTGGGCTCAGTCCAGAGAAGGCGCTCTCAGGAAGCT CTCCGGACAGGTGTGCGGAGGCTGCCAAGAATCCTCTATGGCCTCCCAAGCCACTGTGACAAA AAGTCACAGGCAGACCTCCAGACAGGCTGGGTATGGGACATTAAGTAAAAGGCATTGCCTCATT CTTTACAGGGATAAAATCCCAAAATGTCTCTTGAAGAGACATGTCTACAAACATATTGGACCCT CAGGATGTTCTGGGTAG
99	Translation of ORF number 41 in reading frame 1 on the direct strand MRQKAFLAGCGLSPEKALSGSSPDRCAEAAQESSMASQATVTKSHRQTSRQAGYGTLSKRHCLI LYRDKIPKCLLKRHVYKHIGPSGCSG
100	ORF number 42 in reading frame 1 on the direct strand extends from base 49246 to base 49800 AGCCTCCCCTCACTTCTTCCAGACAACCATCTCCCGCCTCTGCCACAGCCCCTGACCTTGGCT GGCGCTCCAGGAATGAGGACACCACAGGCTCCACGCTCCACCCGGAATGCCCTTCTCCCTCTC TGAGAGCACCGAGGGGGGCTGTGGCCAAGCTGGAGGCCAGGTCGGGAGGGCTTGTTTTGTATGGA AAAGCTACAAGAAGGGCAGAGGGCAAGGTCCTGCTATTGTTTTGGCCGCAAGTGTCTGCACTGCT GCTCTTCAGGCTTTCGAGGAAAGATTCCCCACAGAGGACGCTGGGGTGGGAAGAGAAGGCAGGC AGCTACCTCAGCCCCTGCCAAGTGGTCTTACAGAGGCACTTGGGTGGTTCTGTCTCCAGGTG AGGAAGATCGAGATTGTACGGAAGAAGCCAAGCTTTAAGAAGGCCACAGTGACCCTGGAGGACC ACCTGGCGTGCAAGTGTGAGACGGTAGTGGCTGCACGACCTGTGACCCGAAGCCAGGGAGTTC CCAGGAGCAGCGAGGTAACCTTCAGTCCAGGGTTGGTCTCTGA
101	Translation of ORF number 42 in reading frame 1 on the direct strand SLPSLPSRQPSPASATAPDLGWRSRNETTGSTLHPMPFSLSESTEGGCGQAGGQVGRACFDG KATRAEGKVLLLFWPQCLHCCSSGFRGKIPHRGRWGGKRRQAATSAPAQVVLQRHLGGSVLQV RKIEIVRKPSFKKATVTLEDHLACKCETVVAARPVTRSPGSSQEQRGNLQSRVGL
102	ORF number 43 in reading frame 1 on the direct strand extends from base 53419 to base 53697 TATTGTTCCCCTCGTCCGTCTGTCTCGATGCCTGATTCCGACGGCCAATGGTGTCTCCCCGCCC CTCCACGCGTCCGTCCACCCCTCTGCCAGTGGGTCTCCCCTCAGTGGCscadmCTCAGGGGCTG

	CGAGGAGCGCTACGCGCCTGGTCCCCTCCCGCCTCAGCTCGGCGGCCGCCGGGAGCCCCGACCG AGCCGGCTCCTGGGAGGGCCGGCCCCCTCTCGGGCCTCCAACGAGGAGCAGGAAGGAGGCGGCGG CGGCGGCGAAGGGGTTAAGGTGA
103	Translation of ORF number 43 in reading frame 1 on the direct strand YCSPRPSVSMPDSDGQWCFPAPPRVRPPLCQWVSPQWXXLRGCEERYAPGPVPPQLGGRREPAP SRLGGFPAPLGPPTSRKEAAAAAKGLR
104	ORF number 44 in reading frame 1 on the direct strand extends from base 53698 to base 54324 AGGGCTTCGAGGCCGCGGCCGGGCCCTTGGGCCGAGCCAGCGCAGGTTGTTTTGACCACGGAGG AGCCGTCTCCGTCTCCTTTTGTTCGCGGGCTCCTCGAGGGCCGCCGGCCGTCCGCCCTGGGGC CCCGCCCTTCCGCGGCCGTCCCCGTGGCCCGCACCCGGGAGGAGACGCGGGGATCAGCCTG GCTGCCTGCAGTCCCCCTCCGACGCCCCCTCCTCTCCTCCTGCTGATGCCCCCGGGCCGCGGC CAGTGTTGGGGCGGGGGCGCCGGCCGGCCCCAGCTGCCGCTCGCCGCGggggcctgggggct ggggcctgtgccaggcgctcctgggAACGCGGCGCCCCAGCCGCTGCTCTCCGCAGCCACCC CGCCCGGCCCCCGACTCGCTCACTACCCACGCATGCACACTCTTGGCCGGAGGCGATGCTG CGCTCCGGCGGGCGGGCGCGCAGGGCGACGGGCACGCACTGGCGCGGCCGGGTcgcgcgcccg cgccacgcccgtgcacatgcgggacacacgcgcgcactacacacacacacacatggtccccg cacacacggccttgagcacacgtgcgcgcacacccacgcacgcacAGCCTAG
105	Translation of ORF number 44 in reading frame 1 on the direct strand RASRPRPLGPQPAQVVLTTTEEPSPSPFVLGAPRGPPAVRPGAPPFRGRPPWPAPGREDAGISL AACSPLPTPPPLLLMPPGPRPAVGAGGAGRPQLPPRRGAWGLGPVPGRPGNGGAPAAALRSPP RPAPRLAHSFHACTLLAGGDAALRRAGAQGDGHALARPGRAPAATPVHMRDTRARTTHTHAWSP HTRLEHTCAHTHARTA
106	ORF number 45 in reading frame 1 on the direct strand extends from base 54394 to base 54621 CTCTGTTTCCTTCTTGGGTGTTCTGAGGGAGGGGAAACAGGAACCCCTCCCTCGGGTCCCTCTCC ACAGACCCCATGGGTGTGTTTTTTTTTTTTTTTGGTCAGGTCAGTTCACACCCCTTTGCGCATT ACCTTCTATGATTGCTTTCTTTTCAGCCACTCCCATGTGGCTGAAAATAGTGTGCATGTGCTTG GGGGTACTGTTCAGAGCATTTCTCCCTTCAAGTAA
107	Translation of ORF number 45 in reading frame 1 on the direct strand LCFLLGCSEGGGTGTLPRVPLHSTHGCVFVFFFWSGQFHTLCALPFYDCFLSATPMWLKIVSMCL GGTVQSI SPFK
108	ORF number 46 in reading frame 1 on the direct strand extends from base 54838 to base 55116 GCCTATGGCACAGAACTGGTGCCTGTCTCCTCACTTTAATCACAGCATCCTTGGACACATGGC TCTCAGGAACCCACAGTTGTGTGGTGCTTTGCAGTTTACGAAGCACTTTCCTGCTAAGCCTTAC TCTGAGTAAGCAAGCCTCAGGCAGCTCTTGGGGAAGAGACCTAAAGGAAAAACCTATCGACATG GGGACCAGTCCAGGAAGGTGGACTTCAGGAGATCTTACTGGCAGAGGTGGCCTTGGGGCTGGCC ACGTCTCAGGCCTGTGTGGCTGA
109	Translation of ORF number 46 in reading frame 1 on the direct strand AYGTETGACLLTLITASLDTWLSGTHSCVVLCSLRSTFLLSLTSLKQASGSSWGRDLKGPIDM GTSPGRWTSGLTGRGGLGAGHVSLCG
110	ORF number 47 in reading frame 1 on the direct strand extends from base 56464 to base 56892 ATTAGGCCTGAGACTCTCAGGCAGGCTGGCGCTTGAGGTATGGTCGGCTTCTGCCTCTGCCAA CCATAGACAACGCCCTGGGTGCTGGGGCCAAGAGCGACGTCCTCTCTCAGTGAACGGCGCAC TGGGGAGtggtgtatctgtgtgcagagtgtgtgtctgtgtgCGTGGGGCCAGGTGGAGGGTGG GGTCCAAGCCCCCTTTGATCTGCCAGCATGGTTGGGAGCAGGTAATTACCTGGCCTCACGCTTC CTACCTTCTGCAGCTGGTGTTGGGGGTGGGGTGGGGAAGAGACTGTTTGCTTGGCTCC CAAGGCTGGCTGTGCCCCAGCTGCCTTCTGCCACGCCCCACCCCTGCTAGGAACCCAGGCCT GAGATCTGGGACAGTTTCTCATAGTACCAAGCCTCCTTTCCCTAG

111	Translation of ORF number 47 in reading frame 1 on the direct strand IRPETLRQAGAWRYGRLLPLPTIDNAPGCWGQERRPLSAERRTGECVSVCRVCVVRWGPGGGW GPSPFDLPAWLGAAGNSPGLTLPTFCWCWGGVVKRLFALAPKAGCAPAAFSRPHPARNPRP EIWDSFLIVPSLLS
112	ORF number 48 in reading frame 1 on the direct strand extends from base 57937 to base 58194 GAGTTAGTTGTGGTATTATCAAACCCAGGGCCTCTTAGTGAGTTCTGGGCACCCAGTGGTCAAA TTGCTAGAAGCATGTGCAGGAATGACCTCTCTGCTAAGAATAAAGTGGACTCTATAGGAAACAA TTTGCATGTGTGGGGGTGGTATGGGAGACTATCCCAGGTGGTCCTCCTGGTGGAGGAGGTGAG GGAATCATGCAGGAGAGAACCCCAGGGAGAAGGGGAGAGTCCTTCATGCATTTTACCAGTGTTT AG
113	Translation of ORF number 48 in reading frame 1 on the direct strand ELVVVLSNPGPLSEFWAPSGQIARSMCRNDLSAKNKVDSIGNNLHVWGVVWETIPGGPPGGGGGE GIMQERTPGRRGESFMHFTSV
114	ORF number 49 in reading frame 1 on the direct strand extends from base 58198 to base 58467 GCACCTACTCTGTGCTTTCCCCAGTCTCTGTCTGGGCTCTTCCCCGTGCAGGCTGGGAGGGT GGGGTTCTGGGTTTGTTCCTAAGACATCATCGTCTCTTTTTTATTATAGGCCGGGTCCAGGG TGTCCACTGGGCCCAGCTGGGATCTGCCTACTCTGCCATGGCTAGCAGCTGCAGCCAGCTCTCC AGTGGGCAAGGAGGTCTTGGCATGAGTGTTACGTGCCATTTGGTACTGGGTCTTCAGTCCGCTC TCCTAAGAGGTTAA
115	Translation of ORF number 49 in reading frame 1 on the direct strand APTLCFPPVSVLGSSPCRLGGWGSFVSIRHHRLFFIIGRVQGVHWAQLGSAYSAMASSCSQLS SGQGGLGMSVTCHLVGLQSAALLRG
116	ORF number 50 in reading frame 1 on the direct strand extends from base 59461 to base 59850 GGCACTGAGTTGTTAGACCCAAGGTTAAACAGTGGTAAGTCAAGTCAGCTGACACCCTCCCAGG GCTCCTCCCACGAGACCATGCCGTCCTGTGTGTTGTGCACACACGTGTGTGTTGTGCACACA CGTGTGTGTTTGCCTGGGAGTGAGTGCGGAGGTACAGCAGCATCTTATGCATTTTCTTTGCCCT GAGGCGCTGCGTGTGTCAGCTTTGTGTATCTCAGATTCTCATCTGCCCTCACTTCTTTCTCTAGAC CTCTGGCTTCAGCCCCTTGGGTCTCCCTGGACAGGGGGGATGTGGCTGCGTCTCTTCTATCGG GCTGCTCTCATGTTCATTGTGGGTCTGTGGTTTCCCTGGAGGAAGCCCAGCTCCGAGTGGGGCC TGTTAA
117	Translation of ORF number 50 in reading frame 1 on the direct strand GTELLDPRNLNSGKSSQLTPSQGSSHETMPSCVFVHTRVCLCTHVCLGVSAEVQQHLMHFLCP EALRVSFVYLRFSALTSFSRPLASAPWVSLDRGGCGCVLPIGLLSCHCGSCGFPGGSPAPSGA C
118	ORF number 51 in reading frame 1 on the direct strand extends from base 60442 to base 60786 CCCGGCTGTCCACCTGTCCATGTCCAAGAGGCCCGTGGGAACCTTTCTGTAGGGGATAGTGTCT GTTGGGGCGAAGAGGGCTGTGGCTGGAAAGTCCTTACTCCCAGCGTGTGCTGGCAGGGGGA CCCCATTCCTGAGGAACCTCTATGAGATGCTGAGTGACCACTCGATCCGCTCCTTCGATGACCTC CAGCGCCTGTGCACGGAGACTCCGTAGGTAAATTGAATCCTCGCCAGGGCTCTGGCCCTCCA CTGAGTCCTCGCGTGCCAGGGGGTGGGGAGTGGGTGCCGGGCAAGGGCCATCTCTTTTGTG CCATCCAGAGACCTGTGGCAGCTGA
119	Translation of ORF number 51 in reading frame 1 on the direct strand PGCPPVHVQEAPWELSVGDSVCWGEEGCGWKVLTPSVFAWQGDPIPEELYEMLSDHSIRSFDLL QRLHLHGDSVGKLNPRPGLWPSTESSRARGWVGAGQGPPSSLLCHPETCGS
120	ORF number 52 in reading frame 1 on the direct strand extends from base 60787 to base 61305

	GGGAGGACTTGGCCACACCTGTCTGGGGCAGGGCTGAGTAGGCGGACGGGCTGGTACCTAGGGT GTGAGGTGTGGCAGGAGAAGCATCCACATGTGGCTCTGGCTTGGGGTAGAGGGTGGGGCTGTGG GAGGGGAGGCAGGCAGGAGAGAAGGTGCCAGGGCATCTGCACCCTGAGTATCCAGGTGTGGACT CAGCCAGGGAGGGTGGTGTCTGGAGGAGCCACCTCCCTGTCTCTTGGCCAAAAGGCCCGCTCTAC AAGGTCTCCCGGGACACCTGGCCGGGACCAGTGGGCAGCCCTGCCCCTGCCCAAGAGGGCACT CAGAGAATGGGCACGTGCTTGGTGGCACACACGTGGCAGGGCTGGCGGGCTGTGTCCGGAATGT ATTTATAAACGCTGTCTTCAGAGCAAATTCATTCTATTCTAACCTCTGGCCTGTTCCCTGGAG CCCTGGTCAGCACCCCCCTGCACCCCCAGCTCCCCCTCCCTCTGGGGTTTTGTCTCTTTGTAC TTTGTA
121	Translation of ORF number 52 in reading frame 1 on the direct strand GRTWPHLSGAGLSRRTGWYLGCEVWQEKHPHVALAWGRGWGCGRGRQGEQAQIGICTLSIQVWT QPGRVVLEPPPPCLSGQRPALQGLPGTPGRDQWAALPVPKRALREWARAWWHTRGRAGGLCREC IYKRCLQSKFHSILTSGLFPGALVSTPLHPQLPFPLGFCFLFVTL
122	ORF number 53 in reading frame 1 on the direct strand extends from base 61306 to base 61710 TCCTTGCCAGACTGCTATCTACGGGGACAGCATTTCCCTGCCTTTGTTTCCTCTCCAGTTGG GCCCTGGCTCCCTCTCAAAGCATTTCCCGGGCCCTTTCAAACCCGCCTAGGCCGGGGGCTGA TGATGCAGGCAGGAGGGGGCCCCAGCTGGGCCACCTATTGTTCAACAGGCCCCCCACCCGATG TCTCCACACCCCCACCCCATGCCGACTGGCCAGCCCTGGCCAACACAATGGGGCAACTTCCA AATTTAGCTTTTCTGCTGTTTCTTTCCAAGTCTTCGCCCCACCCCTCATATTGCCCTCCAC ACCCGGGTGGGGTTCGGTTCGAGAGAAGACGAGGTTTTCAATAGCAGGCCTGTTTCGAGGCAAC CATGTGGCTATTTTTTCCTAA
123	Translation of ORF number 53 in reading frame 1 on the direct strand SLPRLSTGDSISCLCFLSQLGPWLPLKSIPRALSNPPRPGADDAGRRGPQLGPPIVHQAPHPM SPTPPPHARLASPGQHNGATSKFSFSAVSFQGPSPPPSYCPSTPRVGVGSEKTRFSIAGLFRGN HVAIFS
124	ORF number 54 in reading frame 1 on the direct strand extends from base 61879 to base 62169 ACAGGGCCCTTGGGAGCCACAGGGACGAGCAATAGTTTTGTCATGGGCAGTGGCAGTGGGAT GGGAGACAGTGTGACCCTGAGATGCTGTGTGGAGGGGACAGAGCTTGTCCCCGACACCCTTC AGTGATTTTGTGGCTTTCAGCCATCAGAGAGCTAGAAGAGTCTGCCACCATTCACAGTCAAG CTCAAAGTTCCCCTGTCCAGCCCTCACTTTCCGCAGCCGGCTTCCGGCTGCCTCTACCCAGAGG GATGTCTCCAAGGAGTGCTGATGGTGTGAGATGA
125	Translation of ORF number 54 in reading frame 1 on the direct strand TGPLGSPQGRAIVLSWAVAVGWGDSVTLRCCVEGDRACPRHPSVYLLAFSHQRARRVCPPFNVK LKVPLSSPHFPQPASGCLYPEGCLQGVLMLR
126	ORF number 55 in reading frame 1 on the direct strand extends from base 62218 to base 62616 ATGTACAGCTTAGGGCAGGCATGGGGCAAAAGGTGAGAGGAGAGAGACAGAAACACAATGAGG GACTGGGAGATGGAGAGAGACCAAGACCTAGAAGGACGCTGGGTGAGGGCTCCCTATCCAGC AGTTCCAGctccctacctctctctgccttagtccccacccacccacccacccctctcctt cccccctctctcccgcccaacTGAACCATTGTCAGGGGCTCCACAGGGGCTGTGTCCAGGGCA TGCTGGTCCCCCTGGGGACTATGGGAATTTCTCCATTGAGCACTTCCTATGGGAACGCTGGGT GGAGGGGCACTGGAAAGTGGCCTCAGAGCTCTGGGTCTTGCCTGCCCTGGAGGCCGAGGAGG GTTTCGCTTACAGTAG
127	Translation of ORF number 55 in reading frame 1 on the direct strand MYSLGQAWGKRSEGERQKHNEGLDGERPRPRRTLGEGPSIPAVPAPYLSLPLVPTPPHPTPLL PTLSPAQLNHCQGLHRGCVQGMVLPPGDYGNFSIQHFLWERWVEGHWKVASELWVLALPWRPRR VRLQ
128	ORF number 56 in reading frame 1 on the direct strand extends from base 62677 to base 62925

	AGAGCCCAGAGTGGGGCTGAAGGCCCTCCGAGGGTACAGTCTGGGCCCCATCACCTCCTGAACC CCATGGCCACCCTGGGGTTTGCCTGGAGGGCGCCTCCTCAGAGGCAGGGAGCCAGAAGGGGAGT ATGTTCTCTGGAGTGGGGTCCCACTGAGGGGCAAGAGGCAATCCTTCCGTCTGTCCCAGAGCCC CCACTGGAGTCCCCAGCCCGTGGTATGACCAGCCAGCACTTGTCACAGTGCTTCTGA
129	Translation of ORF number 56 in reading frame 1 on the direct strand RAQSGAEGPPRVQSGPHLLNPMATLGFARWRAPPQRQGARRGVCSLEWGPSEGQEAILPSVPEP PLESPARGMTSQHLSQCF
130	ORF number 57 in reading frame 1 on the direct strand extends from base 63295 to base 63612 ccctattttataaaattggagactggagcccagagaagggaaagaagtggctgtggtgacacag ctagcatgtgttacgctgggatcccaaTAGCTCTTCTCAGTGCCGCTGCTGTGTCTCTGCTG TGTGGCTAAGGGCTGTGGCCTGGGACAGCAGGCATGGAGCAAGCCTGGGACCTGCCTCCTGCTG TACTGCAGAAACCAAAAGGAGAATGTAGATCAGGGAAGGCAAGTGCCCACTCCACGCCCCCTCTT CCTCTGTGCCCACCTGCAGTCCCCAAAACACTGTAGACAGTGGCTGGGGGGCTCCAGGTAA
131	Translation of ORF number 57 in reading frame 1 on the direct strand PYFIKLETGAQRRERSGCGDTASMWYGDWPNSSSQRLLCVSAVAKGCGLGQQAWSKPGTCLLL YCRNQKENVDQGRQVPTPRPSSSVPTCSPQNTVDSGWGASR
132	ORF number 58 in reading frame 1 on the direct strand extends from base 63946 to base 64236 AATGGATGGGGGCTGGCGGAAGGAACTGGCATTACAAACATGCAGCAGCCTCTGAATTACCTC ACTTGATCCTGACAGTGGTTCTTGGGTGTAGACCTCATCCCCCACTTGACACAGGGGGAAACA GATTCAGAACCCATCAGCGACCTGCCAAAATACCATGGCTGATAACAGCCAGTACTTAAACCT CCCCTGACGTGGAAGGAAGAGGGAATCGGCCAACCGTTTTGGAGGTGCTCCACTTCCTGCCGGC TAGGGGCCCTGAGCAGCCCTCCACCCCACTCCTGA
133	Translation of ORF number 58 in reading frame 1 on the direct strand NGWGLAEGNWHLQHAAASELPHLILTVVLGCRPHPHLHRGKQIQNPSTCPNTMADNSQYLKP PLTWKEEGIGQPFWRCTSCRLGALSSPPPHS
134	ORF number 59 in reading frame 1 on the direct strand extends from base 64288 to base 64677 TCGCGGAGTGTAACGCGCACTGAGGTGTTTGTAGATCTCCCGGCGCCTGGTTGACCGCACCAA CGCCAACTTCTGGTGTGGCCGCCCTGCGTGGAGGTGCAGCGCTGCTCCGGGTGCTGCAACAAT CGCAACGTGCAGTGCCGCCCCACCCAGGTGCAGCTGCGACATGTCCAGGTGTGACGGCCCCACG TCCCCTCCTGGGCTGGCCAGCTGAGAGCAGGGGCTGCCCTCTGGGGCTGGCACTCACGGACC AGGCTCTTGAATGCGTCAAAAGGCATTCTTGGCAGGGTGTGGGCTCAGTCCAGAGAAGGCGCTC TCAGGAAGCTCTCCGGACAGGTGTGCGGAGGCTGCCAAGAATCTCTATGGCCTCCCAAGCCA CTGTGA
135	Translation of ORF number 59 in reading frame 1 on the direct strand SRSVKRALRCLRSPGAWLTAPTPTSWCGRPAWRCSAAPAAATIATCSAAPPSCDMSRCAGPT SPPGLAQLRAGAAPLGLALTDQALECVKRHSWQGVGSVQRRRSQEALRTGVRRLPKNPLWPPKP L
136	ORF number 60 in reading frame 1 on the direct strand extends from base 65287 to base 65886 TCTGGTGACTTACCACGCCCCCTCCCCTGCGGTCAGCTGTGGCCCTTCTCTTGGCCACCTTC CATCCCAGGGCTGGGCCCTGAGCCCGAGATTACGAGTGTCACTCTCCACCCCACTCCCACTGC CATGGTATCTCCTGTCCCCAATGCTTCCAGCTCTATGGATGGACACCTGACAGCTGACCTCCCC CTTCCCGCCTCCCTCCTGGATAAAGCCTCCCCTCACTTCTTCCAGACAACCATCTCCCGCCTC TGCCACAGCCCCCTGACCTTGGCTGGCGCTCCAGGAATGAGGACACCACAGGCTCCACGCTCCAC CCGGAATGCCTTTCTCCCTCTCTGAGAGCACCGAGGGGGGCTGTGGCCAAGCTGGAGGCCAGG TCGGGAGGGCTTGTTTTGTATGGAAGCTACAAGAAGGGCAGAGGGCAAGTCTGTATTGTT TTGGCCGCACTGTCTGCACTGCTGCTTTCAGGCTTTCGAGGAAAGATTCCCCACAGAGGACGC

	TGGGGTGGGAAGAGAAGGCAGGCAGCTACCTCAGCCCCCTGCCCAAGTGGTCTTACAGAGGCACT TGGGTGGTTCTGTCTCCAGGTGA
137	Translation of ORF number 60 in reading frame 1 on the direct strand SGDFTTPPPLRSAVALPLAHLPSQGWALSPRLRVSLSTPPPTAMVSPVPNASSSMDGHLTADLP LPASLLDKASPHFLPDNHLPLPQPLTLGAPGMRTQPAPRSTRKCLSPSLRAPRGAVAKLEAR SGGLVLMEKLEQEQRARSCYCFGRSVCTAALQAFEERFPTEDAGVGREGRQLPQLPKWSYRGT WVVLSSR
138	ORF number 61 in reading frame 1 on the direct strand extends from base 65995 to base 66225 CCCGAAGCCCAGGGAGTTCCCAGGAGCAGGTAACCTTCAGTCCAGGGTTGGTCTCTGATT GCTCAGCCTGGCCAGCCCCCTTCTCCTGTGGCAGCTGCCGGGTGGGGGAAATTGGATCAGGCAT GCGCCCCACCCCCcactcctggttaaattcatctgaagctttccatctcacagaacaatccaga ttcatcccccgactgcaaggccctatatgaggatgtag
139	Translation of ORF number 61 in reading frame 1 on the direct strand PEAQGVPRSEVTFSPGLVSDCSAWPAPSPVAAAGWGEIGSGMRPTPHSWLNSSEAFHLTEQSR FIPPTARPYMRM
140	ORF number 62 in reading frame 1 on the direct strand extends from base 67639 to base 67965 TCAGAACCTTGGGCTAAAATTTCTGCTCTGTCACTTGTGAGTTGTACGACAACCTTGAGCTGGC TCGGGCTTTGCCAGTCCAGTGTCTGCTGGTGGCCTGGTCTCTGGATCAGAACTCCAGGCCCT CAATGGTTCTTCTGGGTACAAAGGTCCCAAGTCCCTGAATTGCAGAGATAGGGTAACACTTTTA TGGGAGCTTGTGTCTGCAAGGTGGGAGGTCAAGTGTTTAACCCAAAGAGTGGGTGGGCCTTGA GCTTGGCAGAGAAAGCTTTCATTTTCTACTTGGGGGCCAGGAGGAAGAGAGATGTAAGCGCAA ACCTTGA
141	Translation of ORF number 62 in reading frame 1 on the direct strand SEPWAKISALSLVSCTTTLWLGLCQSSVLLVAWSLDQKLQALNGSSGYKGPKSLNCRDRVTTL WELVSARWEVKCLTQRVGWALSLAEKAFIFYLGAQEEERCKRKP
142	ORF number 63 in reading frame 1 on the direct strand extends from base 68611 to base 68883 gtctgtgggcgatggggtcagctgggtggttctactgtctgtctcatagtttcggtcagtc atctggaggccacactgggacagctgggcctctgtcattcagggcctctcttttccataggtc tccccagcagggtaccagacttcttatgtggcgacagggctccacaaagtgcaaaggtggg acctaccaggccttttttaggcttatgcctggacctggcacagcactgctctgcctccttttatt gTTTAACAGatagatag
143	Translation of ORF number 63 in reading frame 1 on the direct strand VCGRMGLSWVVLVLLSLIVSVSHLEATLGQLGLCHSGPLFSIWSPPQQGNQTSYVAAQGSTKCKGG TYQAFGLGLCLDLAQHCSASFYCLTDR
144	ORF number 64 in reading frame 1 on the direct strand extends from base 69562 to base 69948 GCGTGGCATGGAGTTCCTAGGCTGCTTCTGACCCCGTGTTCCTCTGCTTACCTTACAGGGTTAT TTAATATGGTATTTGCTGTATTGCCCCATGGGGTCTTGAGTGATAATATTGTTCCCTCGT CCGTCTGTCTCGATGCCTGATTCCGACGGCCAATGGTGCTTCCCCGCCCCCTCACGCGTCCGTC CACCCCTCTGCCAGTGGGTCTCCCCCTCAGTGGCscadmCTCAGGGGCTGCGAGGAGCGCTACGC GCCTGGTCCCGTCCCGCTCAGCTCGGCGGCCCGGGAGCCCGCACCGAGCCGGCTCCTGGGA GGGCCGCCCCCTCTCGGCCTCCAACGAGGAGCAGGAAGGAGGCGGCGGCGGCGGAAGGGGT TAA
145	Translation of ORF number 64 in reading frame 1 on the direct strand AWHGVPRLLLLTPCSSAYLTGLFNMVFAVLPPWGPWSDNIVPLVRLSRCLIRTANGASPLHASV HPSASGSPLSGXXSGAARSATRLVPSRLSSAAAGSPHRAGSWEGRPLSGLQRGAGRRRRRRRG

[illegible]

	ACAACGCCCCCTGGGTGCTGGGGCCAAGAGCGACGTCCTCTCTCAGCTGAACGGCGCACTGGGGA gtgtgtatctgtgtgtgcagagtgtgtgtctgtgtgCGCTGGGGCCCAGGTGGAGGGTGGGGTCCA AGCCCCCTTTGATCTGCCAGCATGGTTGGGAGCAGGTAATTCACCTGGCCTCACGCTTCCTACCT TCTGCAGCTGGTGTGGGGGTGGGGTGGGGTGGGGAAGAGACTGTTTGCCCTTGGCTCCCAAGGC TGGCTGTGCCCCAGCTGCCTTCTCGCCACGCCCTCACCCCTGCTAGGAACCCAGGCCTGAGATC TGGGACAGTTTCCTCATAG
155	Translation of ORF number 69 in reading frame 1 on the direct strand TTPLGAGAKSDVLSQLNGALGSVYLCAECVSVCAQAQVEGGVQAPLICQHGWEQVIHLASRFLP SAAGVGGGVGWGRDCLPWLPRPRLAVPQLPSRHALTLLGTPGLRSGTVSS
156	ORF number 70 in reading frame 1 on the direct strand extends from base 74314 to base 74541 GAAACAATTTGCATGTGTGGGGGTGGTATGGGAGACTATCCCAGGTGGTCCTCCTGGTGGAGG AGGTGAGGGAATCATGCAGGAGAGAACCCAGGGAGAAGGGGAGAGTCCCTTCATGCATTTTACC AGTGTTTAGTGAGCACCTACTCTGTGCTTCCCCAGTCTCTGTCTGGGCTCTTCCCCGTGCA GGCTGGGAGGGTGGGGTTCTGGGTTTGTTCATAA
157	Translation of ORF number 70 in reading frame 1 on the direct strand ETICMCGGWYGRLSQVLLVEEVRESRREPQGESESPSCILPVFSEHLLCAFPQSLSWALPRA GWEGGVLGLFP
158	ORF number 71 in reading frame 1 on the direct strand extends from base 75868 to base 76191 GTGCGGAGGTACAGCAGCATCTTATGCATTTTCTTTGCCCTGAGGCGCTGCGTGTCAGCTTTGT GTATCTCAGATTCTCATCTGCCCTCACTTCTTTCTCTAGACCTCTGGCTTCAGCCCCTTGGGTC TCCCTGGACAGGGGGGATGTGGCTGCGTCCTTCCTATCGGGCTGCTCTCATGTGCTATTGTGGGT CCTGTGGTTTCCCTGGAGGAAGCCCAGTCCGAGTGGGGCCTGTTAAAGTGCTTATTAAGTTTC AAGTGTTTTTGGTAACAGGCCAGAGAGGCTCTAAAAATAGGGTTTGCCTGGGCACCGGGCATGG GTAA
159	Translation of ORF number 71 in reading frame 1 on the direct strand VRRYSILCIIFFALRRCVSALCISDSHLPSLLSLDLWLQPLGSPWTGGDVAASFLSGCSHVIVG PVVSLLEEAQLRVGPVKVLIKQVFLVTGQRGSKNRVCLGTGHG
160	ORF number 72 in reading frame 1 on the direct strand extends from base 76456 to base 76749 CAGACGCTGGCTGTCATCTGTCAGGTGTGGAGGAGAAGCATAAAGATTGTGGGGTTTCCCGGAA CCTGTAGTGTGATGAGGAGATGGATGTATAAATCAATCAGAGCAAACCTGGGGTCTCTTTG GAGGCGAGGGATACAGCATCCTCTCTGGGTCTTCAAGGCTTCGGCAGATTCTGGCCCTTGGGCC TTTGTGTTCTCTGTTCTCAGGCCTGGAATCTACCTCCTGCCACCCCTAGCCCGGCTGTCCACC TGCCATGTCCAAGAGGCCCGTGGGAACCTTTCTGTAG
161	Translation of ORF number 72 in reading frame 1 on the direct strand QTLAVICQVWRRSIKIVGFPGTCSVMREMDVYNQSEQTGGPLWRRGIQHPLWVFKASADSGPWA FVFLVLRPGIYLLPTSPAVHLMSKRPRGNFL
162	ORF number 73 in reading frame 1 on the direct strand extends from base 77218 to base 77469 GTATCCAGGTGTGACTCAGCCAGGAGGGTGGTGCTGGAGGAGCCACCTCCCTGTCTCTCTGG CCAAAGGCCCGCTCTACAAGGTCTCCCGGGACACCTGGCCGGGACAGTGGGCAGCCCTGCCC GTGCCCCAAGAGGGCACTCAGAGAATGGGCACGTGCTTGGTGGCACACACGTGGCAGGGCTGGCG GGCTGTGTCGGGAATGTATTTATAAACGCTGTCTTCAGAGCAAATTCCATTCTATTCTAA
163	Translation of ORF number 73 in reading frame 1 on the direct strand VSRCGLSQGGWCWRSHLPVSLAKGPLYKVSRLAGTSGQPCPCPRGHSENGHVLGGTHVAGLA GCVGNVFINAVFRANSILF
164	ORF number 74 in reading frame 1 on the direct strand extends from base 77470 to base 77925

	CCTCTGGCCTGTTCCCTGGAGCCCTGGTCAGCACCCCCCTGCACCCCAGCTCCCCTTCCCTCT GGGGTTTTGTCTCTTTGTCACTTTGTAATCCTTGCCCAGACTGCTATCTACGGGGGACAGCATT TCCTGCCTTTGTTTCCCTCTCCAGTTGGGCCCCCTGGCTCCCTCTCAAAAGCATTCCCCGGGCCC TTTCAAACCCGCCTAGGCCGGGGGCTGATGATGCAGGCAGGAGGGGGCCCCAGCTGGGCCCCACC TATTGTTACACAGGCCCCCACCCGATGTCTCCACACCCCCACCCCATGCCGACTGGCCAGC CCTGGCCAACACAATGGGGCAACTTCCAAATTTAGCTTTTCTGCTGTTTCTTTCCAAGGTCCTT CGCCCCACCCCTCATATTGCCCTCCACACCCCGGTGGGGGTCGGGTGCGAGAAGACGAGGTT TTCAATAG
165	Translation of ORF number 74 in reading frame 1 on the direct strand PLACSLPWSAPPCTPSSPSLWGFVSLCNPCDCYLRGTAFFPAFVSSPSWAPGSLSKAFPGP FQTRLRGRLLMQAGGGPSWAHLLETRPPTRCLPHPHMPDWPALANTMGQLPNLAFLFLSKVL RPHPHIAPPHPGWGSRRRRGFQ
166	ORF number 75 in reading frame 1 on the direct strand extends from base 78691 to base 78993 ACCATTTGTCTAGGGGCTCCACAGGGGCTGTGTCCAGGGCATGCTGGTCCCCCTGGGGACTATGG GAATTTCTCCATTACAGCACTTCTATGGGAACGCTGGGTGGAGGGGCACTGGAAAGTGGCCTCA GAGCTCTGGGTCTTGGCCTGCCCTGGAGGCCGAGGAGGGTTCGCTTACAGTAGCAAAAGGGAA CGGTTATTTTTTAACCTCATTGACATGGGTTCTGTCCAAAAATGTGGCTGAAGAGCCAGAGTGG GGCTGAAGGCCCTCCGAGGGTACAGTCTGGGCCCCATCACCTCCTGA
167	Translation of ORF number 75 in reading frame 1 on the direct strand TIVRGSTGAVSRACWSPLGTMGISPFSTSYGNAGWRGTGKWPQSSGSLPCPGGRGGFAYSSKRE RLFLTPLTWVLSKNVAEEPRVGLKALRGYSLGPITS
168	ORF number 76 in reading frame 1 on the direct strand extends from base 80761 to base 80985 GAGCAGGGGCTGCCCCCTCTGGGGCTGGCACTCACGGACCAGGCTCTTGAATGCGTCAAAAGGCA TTCTGGCAGGGTGTGGGCTCAGTCCAGAGAAGGCGCTCTCAGGAAGCTCTCCGGACAGGTGTG CGGAGGCTGCCCAAGATCCTCTATGGCCTCCCAAGCCACTGTGACAAAAAGTCACAGGCAGAC CTCCAGACAGGCTGGGTATGGGACATTAAGTAA
169	Translation of ORF number 76 in reading frame 1 on the direct strand EQGLPLWGWHSRTRLLNASKGIPGRVWQSRREGALRKLSGQVCGGCPRIYGLPSHCDKKSQAD LQTGWVWDIK
170	ORF number 77 in reading frame 1 on the direct strand extends from base 81946 to base 82179 TGGAAAAGCTACAAGAAGGGCAGAGGGCAAGGTCCTGCTATTGTTTTGGCCGAGTGTCTGCAC TGCTGCTCTTCAGGCTTTCGAGGAAAGATTCCCCACAGAGGACGCTGGGGTGGGAAGAGAAGGC AGGCAGCTACCTCAGCCCCTGCCCAAGTGGTCTTACAGAGGCACTTGGGTGGTCTGTCTCTCA GGTGAGGAAGATCGAGATTGTACGGAAGAAGCCAAGCTTTAA
171	Translation of ORF number 77 in reading frame 1 on the direct strand WKSYYKGRGQGPVIAVLAASALLFRLSRKDSQPRTLGWEEKAGSYLSPCPSGLTEALGWFCPP GEEDRDCTEEAKL
172	ORF number 78 in reading frame 1 on the direct strand extends from base 82474 to base 82701 ggatgtagccccagttggccctttgtcttgcctgccaaccaatccccctcactgtgacaccc cagccagcctggcctttttgaatggcagctacatttctgcctcagggcctttgcacatgccac tctgtctgaaactcacttctctcagctcttcacaagcctactccttctcttcatcttgatctta gctcagaagtcatctcctcctagaagtctgccctga
173	Translation of ORF number 78 in reading frame 1 on the direct strand GCSPQLALWSCQPIPPHCDTPASLAFLNGQLHFCLRAFAHATLSETHFSQLFTSLLLLFIWIL AQKSSPPRSLP

174	ORF number 79 in reading frame 1 on the direct strand extends from base 84400 to base 84645 gggtttctggctattttcatatactatctcctaatacctaggaggccagggtgctggcatctcc attttagagatgtggaattgaggcacagggagtttatatgacttgcccaaaccacatgactaa cacgtgggagagcccagatttgaaccaggtGGTCTGGCCACCATCTGAGCTCTGGACTGCC CACTGTGCCGTTACTCTAAGTGGCGAGGTAAGGCAGACGTCAGGCGCAACTGA
175	Translation of ORF number 79 in reading frame 1 on the direct strand GFLAIFIYYLLILGGQGCWHLHFRDVEIEAQGVYMTCPNHMTNTWESPDLPNGGLAHLSSGLP HCAVTLSGEGKADVRN
176	ORF number 80 in reading frame 1 on the direct strand extends from base 85966 to base 86799 TTCGACGGCCAATGGTGCTTCCCCGCCCTCCACGCGTCCGTCCACCCCTCTGCCAGTGGGTC TCCCCTCAGTGGCscadmCTCAGGGGCTGCGAGGAGCGCTACGCGCCTGGTCCCGTCCCGCCTC AGCTCGGCGGCCCGGAGCCCGCACCAGCCGCTCCTGGGAGGGCCGGCCCTCTCGGGCC TCCAACGAGGAGCAGGAAGGAGGCGGCGGCGGCGGAAGGGTTAAGGTGAAGGGCTTCGAGG CCGCGGCGGGCCTTGGGCGCAGCCAGCGCAGGTTGTTTTGACCACGGAGGAGCCGTCTCCGT CTCCTTTTGTCTCGGGGCTCCTCGAGGGCCGCGGCCGTCCGCCCTGGGGCCCCGCCCTTCG CGGCCGTCCCCGTGGCCCGCACCCGGGAGGAGACGCGGGGATCAGCCTGGCTGCCTGCAGT CCCCTCCGACGCCCCCTCCTCTCCTGCTGATGCCCCCGGGCCGCGGCCAGCTGTTGGGG CGGGGGCGCGGCCCGGCCAGCTGCCGCTCGCCGCGggggcctgggggctgggcctgtgcc agggcgctcctgggAACGCGGCGCCCCAGCCGCTGCTCTCCGACGCCACCCCGCCCGCCCC CGACTCGCTCACTACCCCCACGCATGCACACTCTTGCCCGGAGGCGATGCTGCGCTCCGGCGGG CGGGCGCGCAGGGCGACGGGCACGCACTGGCGCGGCCGGGTcgcgcgcccgccacgcccgt gcacatgcgggacacacgcgcgcgcactacacacacacacgcacatggtccccgcacacacggctt ga
177	Translation of ORF number 80 in reading frame 1 on the direct strand FGRPMVLRPSTRPSTPLPVGLPSVAXXQGLRGALRAWSRPASARRPPGARTEPAPGRAGPSRA SNEEQEGGGGGEGVKVGFEEAAGPWAAASAGCFDHGGAVSVSFCSRGSSRAAGRPPWGPALP RPSPVARTREGGRGDQPGCLQSPDPAPSSPPADAPRAAASCWGGGRRPAPAAASPRGLGAGPCA RASWERRRPSRCSPPQTPPGPPTRSLTPRMHTLGRRRCCAPAGGRAGRRARTGAAGSRARRHAR AHAGHTRAHYTHTRMVPAHTA
178	ORF number 81 in reading frame 1 on the direct strand extends from base 87169 to base 87486 GCTTCGAGGGGGCTCCACCCCCACTGTCTGTGCTCTTTGCTGATCCCAGCCAGCACGCTG CAGAGAGGCTGGGTGACAGCTGGATAAGGCTTTCCCGCCTGCCCTTACCATTCCCAGCTTCATC CAGCACCTCCTCCTCTTTCCCACTCCCTGGGTGTGTGTTGGGGGTGAGCCTATGGCAC AGAACTGGTGCCTGTCTCCTCACTTTAATCACAGCATCCTTGACACATGGCTCTCAGGAACC CACAGTTGTGGTGTCTTGCAGTTTACGAAGCACTTTCTGCTAAGCCTTACTCTGAGTAA
179	Translation of ORF number 81 in reading frame 1 on the direct strand ASEGAPTPHCSVLFADPSQHAAERLGDWIRLSRLPLPFPASSSTSSSFPTTPWVCVWGVSLWH RNWCLSPHFNHSILGHMALRNPQLCGALQFTKHFPKPYSE
180	ORF number 82 in reading frame 1 on the direct strand extends from base 88375 to base 88617 TCCTCAGCCCAGGGTAGCCTAGAATGGCCACACTGCTCTTACCAGGCATCCTCATTCGAGCCC CCCCGCCCCCATCTTGAGAGACAAGCATATCTTTCTTTTCCATGTCTTGGGCTGCCAATATT GGACAGGACAGAGGGGAAGAAACAGAAGGAAAATCAGATCGCAAGGCTTCTGTGTATCTTGAGC AGGCTGGGCCTCAGTTGCCGCCGCGTGAGAATATGAGAAGGTTGGATTAG
181	Translation of ORF number 82 in reading frame 1 on the direct strand SSAQGSLEWPHCSPPGILIRAPPAPHLERQAYLSFPCLGLPILDRTEGKKQKENQIARLLCILS RPGPQLPPRENMRRLD

182	ORF number 83 in reading frame 1 on the direct strand extends from base 89983 to base 90402 gCAATCCAGGTTTCCTTGGCAGCTGAAGCTCTACAgtttctctgcctctccactgttgacattt ggggccagacagttcttgattgtgggggagggctgtcctgtgcatagtaggatgttttagcagcaa ccctggcctctacctaactagacaccagtagcaggcctccagttgtgataaccaaagtgcctcc agacctggccagtgtcccctgggggtcaCTCACTCCCTGCTCTATGACCTCCACTGGGTGAAGA GTGGACCTGAACTGAAAACAGTCCATGAAAGAGGGAGGGGCCGCTCTGCTCCTTACCAGTCGT GTTGACCTTTAGCCATTACTTAATTTTTCTAAGCCTCAGCTTCCTCATTGGAAGACAGGGAT ACAAACAGTGACAGCCTCTTGATTGTATTTGATTGA
183	Translation of ORF number 83 in reading frame 1 on the direct strand AIQVSLAAEALQFLCLSTVDIWGQTVLDCGGGCPVHSRMFSSNPGLYLLDTSRPPVVITKSAS RPGQCPLGVTHSLLYDLHWVKSGPELKTVHERGRGPLCSLPVVLTFSHLLNFSKPQLPHLED RD TNSDSLIVFD
184	ORF number 84 in reading frame 1 on the direct strand extends from base 90640 to base 90882 GGGAATCATGCAGGAGAGAACCCAGGGAGAAGGGGAGAGTCCCTTCATGCATTTTACCAGTGTT TAGTGAGCACCTACTCTGTGCTTTCCCCCAGTCTCTGTCTGGGCTCTTCCCGTGACAGGCTGG GAGGTGGGGTTCTGGGTTTGTTCATTAAGACATCATCGTCTCTTTTTTATTATAGGCCGGGT CCAGGGTGTCCACTGGGCCAGCTGGGATCTGCCTACTCTGCCATGGCTAG
185	Translation of ORF number 84 in reading frame 1 on the direct strand GNHAGENPREKGRVLHAFYQCLVSTYSVLSPSLCPGLFPVQAGRVGFVWCFHKTSSSLFYRPG PGCPLGPAGICLLCHG
186	ORF number 85 in reading frame 1 on the direct strand extends from base 90883 to base 91161 CAGTGCAGCCAGCTCTCCAGTGGGCAAGGAGGTCTTGGCATGAGTGTTACGTGCCATTTGGTA CTGGGTCTTCAGTCCGCTCTCCTAAGAGGTTAATTGATTCAATTATGCCACAAACAGCCTGGGAG ACCTGCTGGGCACCCCACTTCGGCTTCTCTGCTGCTGCTCTCCTGCCAACCCAGACAGA ATTAGAATTAATAATCAAATCAAATGGGTACAACCCCTCAGTTCACAGGTGATGCCAGGACCC GAGAGGGGCAGCAACCAACCTGA
187	Translation of ORF number 85 in reading frame 1 on the direct strand QLQPALQWARRSWHECYVPFGTGSSVRSRKRLIDSLCHKQPGRPGWAPPLRLPLLLPLLPDR IRIKISNGYNPLSSQVIARTREGQQPT
188	ORF number 86 in reading frame 1 on the direct strand extends from base 93412 to base 93690 AGGGTGGGGCTGTGGGAGGGGAGGCAGGCAGGGAGAAGGTGCCCAGGGCATCTGCACCCTGAGT ATCCAGGTGTGGACTCAGCCAGGGAGGGTGGTGTGGAGGAGCCACCTCCCTGTCTCTTGGCC AAAGGCCCGCTCTACAAGGTCTCCCGGGACACCTGGCCGGGACCACTGGGGCAGCCCTGCCCGT GCCCAAGAGGGCACTCAGAGAATGGGCACGTGCTTGGTGGCACACACGTGGCAGGGCTGGCGGG CTGTGTCCGGAATGTATTTATAA
189	Translation of ORF number 86 in reading frame 1 on the direct strand RVGLWEGRQAGRRCPGHLHPEYPGVDSAREGGAGGATSLSLWPKARSTRSPGDTWPGPVGSPAR AQEGTQRMGTCLVAHTWQGWRAVSGMYL
190	ORF number 87 in reading frame 1 on the direct strand extends from base 93691 to base 93933 ACGCTGTCTTCAGAGCAAATTCCATTCTATTCTAACCTCTGGCCTGTTCCCTGGAGCCCTGGTC AGCACCCCCCTGCACCCCCAGCTCCCTTCCCTCTGGGGTTTTGTCTCTTTGTAAT CCTTGCCCACTGCTATCTACGGGGACAGCATTTCTGCTTTGTTTCTCTCCAGTTGGG CCCTTGCTCCCTCTCAAAGCATTCCTCGGGCCCTTTCAAACCCGCCTAG
191	Translation of ORF number 87 in reading frame 1 on the direct strand

	TLSSEQIPFYSNLWPVPWSPGQHPAPPAPLPSGVLSSLCHFVILAQTAIYGGQHFLPLFPLPVG PLAPSQKHSPGPFKPA
192	ORF number 88 in reading frame 1 on the direct strand extends from base 94081 to base 94554 CTTTTCTGCTGTTTCTTTCCAAGGTCCTTCGCCCCACCCTCATATTGCCCCCTCCACACCCCGG GTGGGGGTGCGGTTCGGAGAAGACGAGGTTTCAATAGCAGGCCTGTTTCGAGGCAACCATGTGG CTATTTTTTCTTAATCAACTTAACCTTTCCACAAAGCACATCTTTCCCCATCTCTCCCAACC AGGGACATTCCAGAAATGGCAGAGAGAAAGGAATGGAGCCAGAGGGACAGACAGACACACTGTT CGTGGGACAATAGGCTAGACGGAAGTGCATCAGTTTTAGGAAAGTCTGCTCTAAACAGGGCCCC TTGGGAGCCCCACAGGGACGAGCAATAGTTTTGTTCATGGGCAGTGGCAGTGGGATGGGGAGACAG TGTGACCCCTGAGATGCTGTGTGGAGGGGACAGAGCTTGTCCCCGACACCCCTCAGTGTATTG CTGGCTTTTCAGCCATCAGAGAGCTAG
193	Translation of ORF number 88 in reading frame 1 on the direct strand LFCCFFPRSFAPTLILPLHTPGGGRVGEDEVFNSRPVSRQPCGYFFLINLTFPQSTSFPHLLPT RDIPEMAERKEWSQRDRQTHCSWDNRLDGSASVLGKSALNRAPWEPTGTSNSFVMGSGSGMGRQ CDPEMLCGGGQSLSPFPFSVFAGFQPSSES
194	ORF number 89 in reading frame 1 on the direct strand extends from base 94555 to base 94791 AAGAGTCTGCCCACCATTCAACGTCAAGCTCAAAGTTCCCCTGTCCAGCCCTCACTTTCCGCAG CCGGCTTCCGGCTGCCTCTACCCAGAGGGATGTCTCCAAGGAGTGCTGATGGTGCTGAGATGAG GGCTTCCAGGCTAGAGAAGGGAGCTGTAGTTGTGACCTTAGGAATAAATGTACAGCTTAGGGCA GGCATGGGGCAAAAGGTCAGAGGGAGAGAGACAGAAACACAATGA
195	Translation of ORF number 89 in reading frame 1 on the direct strand KSLPTIQRQAQSSPVQPSSLAAGFRLPLPRGMSPRSADGAEMRASRLEKGAVVVTLGINVQLRA GMGQKVRGRETETQ
196	ORF number 90 in reading frame 1 on the direct strand extends from base 94840 to base 95151 GGGTCCCCCTATCCCAGCAGTTCCAGctccctacctctctctgccttttagtccccacccccacc cacccccacccctctccttccccaccctctctccccgccaacTGAACCATTGTCCAGGGGCTCCACA GGGGCTGTGTCCAGGGCATGCTGGTCCCCCTGGGGACTATGGGAATTTCTCCATTACGACTT CCTATGGGAACGCTGGGTGGAGGGGCACCTGGAAGTGGCCTCAGAGCTCTGGGTCTTGCCCTG CCCTGGAGGCCGAGGAGGGTTCGCTTACAGTAGCAAAAGGGAACGGTTATTTTAA
197	Translation of ORF number 90 in reading frame 1 on the direct strand GLPYPSSSSSLPLSAFSPHPTPPHPSPSHPLSRPTEPLSGAPQGLCPGHAGPPWGLWEFLHSAL PMGTLLGGGALESGLRALGPCPALEAEESLTVAKNGYF
198	ORF number 91 in reading frame 1 on the direct strand extends from base 95344 to base 95658 GGGGCAAGAGGCAATCCTTCCGTCTGTCCCAGAGCCCCACTGGAGTCCCCAGCCCGTGGTATG ACCAGCCAGCACTTGTACAGTGCTTCTGACTGTGCCTTCTCTGCAGATGAAGACGGGGCTGA GTTGGACCTGAATTTGACTCAGTCCCATTTCTGGAGGCAAGCTGGAGAGCTTATCCCAGGGGAGA AGGAGCCTAGGTAAGAATGAGGGTGCAAACGGGGGCCCTCAAAGGTGGGGGCCAGGGAAGAAG AACTGAGCACACAGCCTGCCGGAGGCTGTGAGGGTGGGCCCTGTTTGTCCCACTTAG
199	Translation of ORF number 91 in reading frame 1 on the direct strand GARGNPSVCPRAPTVGPSPWYDQPALVTVLLTVPSLADEDGAELDLNLTQSHSGGKLESLSRGR RSLGKNEGANGGPSKVGAREEELSTQPAGEGEGPCLSHT
200	ORF number 92 in reading frame 1 on the direct strand extends from base 95944 to base 96174 GGGTGTGGCCTGGGACAGCAGGCATGGAGCAAGCCTGGGACCTGCCTCCTGCTGTACTGCAGA AACCAAAAGGAGAATGTAGATCAGGGAAGGCAAGTGCCCACTCCACGCCCCCTCTCCTCTGTGC CCACCTGCAGTCCCCAAAACACTGTAGACAGTGGCTGGGGGGCCTCCAGGTAAGAGTCAGTGGC CTGAGTTCCACTCTTTGCTCTGTGAATTTGGGCATCTAA

201	Translation of ORF number 92 in reading frame 1 on the direct strand GLWPGTAGMEQAWDLPPAVLQKPKGECRSKASAHSTPLFLCAHLQSPKHCRQWLGGGLQVRVSG LSSTLCSVNLGI
202	ORF number 93 in reading frame 1 on the direct strand extends from base 96631 to base 97065 CAGCCAGTACTTAAACCTCCCCTGACGTGGAAGGAAGAGGGAATCGGCCAACCGTTTGGAGG TGCTCCACTTCTGCCGCTAGGGGCCCTGAGCAGCCCTCCACCCCACTCCTGACGGAGTTCCC TCTCCCTTCAGATTCCCAGCCGGTCGCCGAGCCAGCCATGATCGCGGAGTGTAACGCGCACT GAGGTGTTTGAGATCTCCCGGCGCCTGGTTGACCGCACCAACGCCAACTTCTGGTGTGGCCGC CCTGCGTGGAGGTGCAGCGCTGCTCCGGCTGCTGCAACAATCGCAACGTGCAGTGCCGCCCCAC CCAGGTGCAGCTGCGACATGTCCAGGTGTGCAGGCCCCACGTCCCCTCCTGGGCTGGCCAGCT GAGAGCAGGGGCTGCCCTCTGGGGCTGGCACTCACGGACCAGGCTCTTGA
203	Translation of ORF number 93 in reading frame 1 on the direct strand QPVLKTS PDVEGRGNRPTVLEVLHFLPARGPEQPSTPLLTEFPLPSDSQPVAEPAMIAECKTRT EVFEISRRLVDRTNANFLVWPPCQVEVQRCSGCCNNRNQCRPTQVQLRHVQVCRPHVPSWAGPA ESRGCPSGAGTHGPGS
204	ORF number 94 in reading frame 1 on the direct strand extends from base 97066 to base 97338 ATGCGTCAAAGGCATTCTGGCAGGGTGTGGGCTCAGTCCAGAGAAGGCGCTCTCAGGAAGCT CTCCGACAGGTGTGCGGAGGCTGCCAAGAATCCTCTATGGCTCCCAAGCCACTGTGACAAA AAGTCACAGGCAGACCTCCAGACAGGCTGGGTATGGGACATTAAGTAAAAGGCATTGCCTCATT CTTTACAGGGATAAAATCCCAAAATGTCTCTTGAAGAGACATGTCTACAAACATATTGGACCCT CAGGATGTTCTGGGTAG
205	Translation of ORF number 94 in reading frame 1 on the direct strand MRQKAFLAGCGLSPEKALSGSSPDRCAEAAQESSMASQATVTKSHRQTSRQAGYGTLSKRHCLI LYRDKIPKCLLKRHVYKHIGPSGCSG
206	ORF number 95 in reading frame 1 on the direct strand extends from base 98014 to base 98568 AGCCTCCCCTCACTTCCCTCCAGACAACCATCTCCCGCCTCTGCCACAGCCCCTGACCTTGGCT GGCGCTCCAGGAATGAGGACACCACAGGCTCCACGCTCCACCCGAAATGCCTTTCTCCCTCTC TGAGAGCACCGAGGGGGGCTGTGGCCAAGCTGGAGGCCAGGTGCGGAGGGCTTGTTTTGATGGA AAAGCTACAAGAAGGGCAGAGGGCAAGGTCCTGCTATTGTTTTGGCCGAGTGTCTGCACTGCT GCTCTTCAGGCTTTTCGAGGAAAGATTCCCCACAGAGGACGCTGGGGTGGGAAGAGAAGGCAGGC AGCTACCTCAGCCCCCTGCCAAGTGGTCTTACAGAGGCACTTGGGTGGTTCTGTCTCCAGGTG AGGAAGATCGAGATTGTACGGAAGAAGCCAAGCTTTAAGAAGGCCACAGTGACCCTGGAGGACC ACCTGGCGTGCAAGTGTGAGACGGTAGTGGCTGCACGACCTGTGACCCGAAGCCAGGGAGTTC CCAGGAGCAGCGAGGTAACCTTCAGTCCAGGGTTGGTCTCTGA
207	Translation of ORF number 95 in reading frame 1 on the direct strand SLPSLPSRQSPASATAPDLGWRSRNETTGSTLHPMPFSLSESTEGGCGQAGGQVGRACFDG KATRRAEQKVLFLFWPQCLHCCSSGFRGKIPHRGRWGGKRRQAATSAPAQVVLQRHLGGSVLQV RKIEIVRKPSFKKATVTLEDHLACKCETVVAARPVTRSPGSSQEQRGNLQSRVGL
208	ORF number 96 in reading frame 1 on the direct strand extends from base 102187 to base 103830 TATTGTTCCCCTCGTCCGTCTGTCTCGATGCCTGATTTCGGACGGCCAATGGTGTCTCCCCGCC CTCCACGCGTCCGTCCACCCCTCTGCCAGTGGGTCTCCCCTCAGTGGCscadmatctgtggcca gcctaattcaagaaagtcgtttggaagctcgaaaatattacggaaggagccagatttgattgt tgttccttttacaaaaatgcagattcaaggcttgatgcagtttacagttttcccatgccttg ctcattttacaggaacttttagataatcattatcctaagcataaattgcttcagttttttcaaca tcatgatccaatttttccttcaattgtgtcacatgctcctcttctgctgttccaaatggtttt actgatggatctaataatggagtagctgtttatgcactcaataaaaaagtcaccaagagagtac agacacctccagcttcagctcaaatagttgagcttcgagcagttacataaggtgctgcttgattt

	tgcttctcagtccttttaattttattctctgacagccattatgtggttcgtgcagtcagaaattta gaaacagtacctttttatttagcactagtaatcctgttattcaggatttggtttcttcagatacaac aggccattcagctgcgctgtaaaaaattttatattggccatatttagagctcactctaattctcc aggtccttttagcagcagggcaatcaaattgcagattctgccacgcagcttattgccttaactcaa atagaaaaagcacaaaaggctcatagcctccaccatcaaaatagccagagcctaagattacagt ataagatcctcagagaagcagcagcagcagattataaaacaatgtccagattgctcgcatttaca acctgtgcctcattatggcattaaccctcgaggcttgctgcccaatgatctgtggcaaatggat gttactcatatacctgaatttggaataaaatacgtccatgtctctatagacacgttttctg gctttgtaatagcttctgctcaatcaggagaagctacatctcatgttatttagacattgtcttg tgcttttgccatgattggcactcctaaaaacttaaaacagataatggctccggctacaccagt aaaaaatttgctttattttgtcaacaatttttaattaatcatgttactggcatccttacaatc cccagcgacaagggattgttgaacgtactcatggcacattaaaagtcattttacaaaaataaa aaagggggagttatatcccctaacgcccataattacttgtctcattctctttttattcaaat ttttgaccttgatgcccattggaagagtgtgcagagcgcttttgatccttctactgcc ctcaggttttggtcaaatggaaagatccacttactggatcttggaaggccagatccagtcct catatggggcgagggcatgtttgtgtttttccacaggatgcagaaggccctcggtggctgcc gaacgattgggtgcacatgtggaccctctacctgtgatgacattgatgascadmATGTTGTTT TATGTGTCCATCAGAAGGCAATCCTAGGGCGTGAGTGTTCATTGA
209	Translation of ORF number 96 in reading frame 1 on the direct strand YCSPRPSVSMPSDSDGQWCFPAPPRVRPPLCQWVSPQWXXICGQPNRKSFGSSKILRKGARFDC CSFYKNADSRDLDAVYSFPIALAHFTGTLDNHYPKHKLQFFQHHDPIFPSIVSHAPLPVNVF TDGSNNGVAVYALNKKVTKRVQTPPASAQIVELRAVHKVLLDFASQSFNLFSDSHYVVRVRNL ETVPFISTSNPVIQDLFLQIQAIQLRCKKFYIGHIRAHSNLPGPLAAGNQIADSATQLIALTQ IEKAQKAHSLHHQNSQSLRLQYKILREAAEQI IKQCPDCSHLQPVPHYGINPRGLRPNDLWQMD VTHIPEFGKLKYVHVSIDTFSGFVIASQSGEATSHVIRHCLAAAFAMIGTPKKLKTDNGSGYTS KKFALFCQQFLINHVTGIPYNPQRQGIVERTHGTLKVILQIKKGELYPLTPHNYLSHSLFIQN FLTLDAHGKSAERFWHPSTATQALVKWKDPLTGSWQGPDPVLIWGRGHVCFVQDAEGPRWLP ERLVRHVDPLPADDIDXXXVVLVHQAAILGRECH
210	ORF number 97 in reading frame 1 on the direct strand extends from base 107215 to base 107613 tgtgggacacagctccctagcccatgctgttattatgagccttgcgctccccctaattcctccc ccatcacttggtcggttgctgactgctcacagcagctcacgcagctccttgccggtgccagcca ctcatgctggttgccggccactcacgctgaccatcgccgctcctggcagcacacggcagcaca cagcagcccgccgagctcacgccgacctctggctgctgcagcccagctccaggagccggttg ttcaaatcttagctgtagaggtgagctcactggcccatgtgggaatcgaaccggtgacctc gttgtaggcgcagcgctccaaccacctgagccaccagcgcccccTGATTGTGTTTCTATAT ACTGtgtttccctga
211	Translation of ORF number 97 in reading frame 1 on the direct strand CGTQLPSPCWYYEPCAPPNSSPITGRWSTAHSRRQLLPAASHSCWLPATHADHRPLLAHGST QQPAAAHADLWLLAAQLQGAUVHNLSCRGCSLAHVGIPEVTSLLGARRSNHLSHQALIVFLY TVFP
212	ORF number 98 in reading frame 1 on the direct strand extends from base 107752 to base 107997 aataagaccgggttttatattaagttttgctccaaaagacgcattagagctgattgtccagcta ggtcttattttcggggaacatggTAGAGAATCATACAGATTCTCTGCATATAAGGAATTTTGT AAAGGAGAAGGGTACTGAGCAGAGATTATATCTCTCAAATAACACTATTCTCTCTCTCTTTTGT ATTTTACAGTGGAGGAAAGGAGGACAAAGTACTAAAGTAAAAAGTAGATCTTGA
213	Translation of ORF number 98 in reading frame 1 on the direct strand NKTGFYIKFCSKRRIRADCPARSYFRGNMVENHTDSLHIRNFVKEKGTEQRLYLSNNTILSSFL ILQWRKGGQSTKVKRSR
214	ORF number 99 in reading frame 1 on the direct strand extends from base 113266 to base 113505

	AGAGAACTGAGGTTGCTTGTCTTTATAGCTACTAGTGGCCTCAAAAGGCCAATACATCTGTCTC CATTGTGCCCTTGTCTCAATACCCTCTGATTACAAAGCCTTTCTTCTCTTAGGAAACGAATGGC AGAGAATGAACTGAGCCGGTCGGTGAATGAGTTTCTGTCCAAGCTGCAGGATGACCTCAAAGAG GCAATGAATACCATGATGTGCAGCCGATGCCAGGGAAGCATAGGTAG
215	Translation of ORF number 99 in reading frame 1 on the direct strand RELRLLVFIATSGLRPIHLSFVPCSIPLSDLSLRLRMAENELSRVNEFLSKLQDDLKE AMNTMCSRQCKHR
216	ORF number 100 in reading frame 1 on the direct strand extends from base 113818 to base 114210 GGAGTTGTCCTTTTGTGGGTTGTAGGAGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGAT ACTGTGCTGAGTGTAATAGGCTGCATCCCGCTGAAGAAGGAGACTTTTGGGCAGAGTCTAGCAT GTTGGGCCTGAAAATCACCTACTTTGCGCTGATGGATGGAAAGGTGTATGACATCACAGGTACA CTTCTGTCTCTAGAAATCCAGACTCATGTATGCTCAAACTGTTATGTATTGGCTAATTATTT CTCATGCTTGCAGAGTGGGCTGGATGCCAGCGTGTGGGAATCTCCCAGATACCCACAGAGTCC CCTATCACATCTCATTTGGTTCTCGGATCCCAGGCACCACTGGGCGACAGAGGTGGGTGATATT TTCCAATAA
217	Translation of ORF number 100 in reading frame 1 on the direct strand GVVLLLGCRRFEMDREPKSARYCAECNRLHPAEEGDFWAESSMLGLKITYFALMDGKVYDITGT LLSSRIPDSCMLKTVMYWLIISHACRVGWMPACGNLPRYPQSPLSHLIWFSDPRHQWATEVGDI FQ
218	ORF number 101 in reading frame 1 on the direct strand extends from base 114376 to base 114630 CTCTTAATTTCTTTTGCCTCATTATTCTTTTGTTTTCCACCCAGAGCCACCCAGATGCCCTC CTGCTGACCTTCAGGATTTCTTGAGCCGATCTTTCAAGTACCCACAGGACAGATGTCTAATGG GAACTTCTTTGCAGCTCCTCAGCCTGGCCCTGGGGGCACCGCAGCCTCCAAGCCTAACAGCACA GTACCCAAGGGAGAAGCCAAACCGAAGAGGCGGAAGAAAGTGAGGAGGCCCTTCCAACGTTGA
219	Translation of ORF number 101 in reading frame 1 on the direct strand LLISFASLFFCFPPRATPDAPPADLQDFLSRIFQVPPGQMSNGNFFAAPQPGPGGTAASKPNST VPKGEAKPKRRKKVRRPFQR
220	ORF number 102 in reading frame 1 on the direct strand extends from base 114631 to base 114945 CACCCCTTCTCTCTCTCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTGTACAGCACAGGA TGGAGTTTGATTGTTTATTTTAAATATTTAAAAAGGAAAATTTAAGCTCAAATTGTTCACT CAGTACTTGTAGScadmgagaacaggtctgggggtattttcccaggggtcatagatttacctgt actccacaaaaaaactgcaaaggcaataatttggaacacagatacacctgtgtgaatagatcag tggccccttacacagaaaaagatatcgccgcccagcgcttgtagcaggagcagcttga
221	Translation of ORF number 102 in reading frame 1 on the direct strand HPFSSLLKSMGSGQKGCVQHRMEFDLFIFKYLRKILSSNCSLSTCXXREQVWGI FPRGHRFTC TPPKNCKGNNLENRYTCVNRSVAPYTEKDIGRPGACTGAA
222	ORF number 103 in reading frame 1 on the direct strand extends from base 119038 to base 119274 gtgatagctccacgacctcgtgttacggagcttgagtgaggctcgtaactgcgtttccggcactg tcttacggctaaacggcgatcaaaacttcggttttgccaggcgggggtttataccgccacgct taattgccacgatagctcttggtcccgcgaggggcacggccagccgagcatctgtgtgTTTTACT TGTGTGAAAGAAGGGCCGAGGATAAAGGGAATGGGTACACGCTAA
223	Translation of ORF number 103 in reading frame 1 on the direct strand VIAPRPRVTELEWARNCVSGTVLRLNGDQNFGEFARAGVYTATLNCHDSLGPARGTASRASVCFT CVKEGPRIKNGSR
224	ORF number 104 in reading frame 1 on the direct strand extends from base 121210 to base 122190

	caaagacggcgaacccttacagggaaactgggtgaggggcccagccccaggccccgactcagcaa tgttatggggcactgcaggttcaggaacagaccaggagccgaaaaagaacgaacccctgctag gaagcatgtcacagacttattcagggccaccacaggcagcgcaggattggacttgtgtccacc tccgacatcatattaactcctgaaatgggaatgcaagttttgccactggagtttttgggcccc tgccacctaaaacgggtgggtttactgttaggaagaagcagctccgttataaaagggaattcatgt ttctccagggtatttattgataggattttacaggagaaaataaaaaattatggctcattctcctctt aatatttctgccattcctgctggaaccggtattgcacaactgtttattttgctcgtcttaata ttggaaaaaacaggcaaaatcaagagcgggggaaccaaggatttggctcttctgatgtatattg gattcaagaaataaaaaaggatcgaccctgattgttactcaaaataaatggaaaagattttcaa ggacttctggacactggagccgatgtctcgtgcataatctgctgaacattggccctccagttggc cgacgcgctttactaatccaattttacaaggcataggccaatcgcaatccccctccaagtag tgatcttttgtcttggcaagatccggagggtcatcaggggacgtttcagccatataattatccct ggtcttccagtttaatttatggggaagagatgttatgagtaaaatgggagtttatctttacagtc ctagttcacaagtaactcaacagatgtttgatcaaggctttctccctggtcagggttaggctc ggtgggacaagggcgccgagagcctatttcaactaatcctaacttacagagaacagggtctggg tattttccccagggtcatag
225	Translation of ORF number 104 in reading frame 1 on the direct strand QRRQLTLGKLGEPPAPGPDSSAMLWGTSAGTDPGAEKERTPARKHVTDLFRATTGSAGLDLCST SDIILTPMGMQVLPFGVGLPPKTVGLLLGRSSSVIKGIHVSPGIIDEDFTGEIKIMAHSP NISAI PAGTRIAQLFILPRLNIGKNRQNERGNQGGSSDVYWIQEIKKDRPVLLKINGKDFQ GLLDTGADVSCISAEHWPSWPTRFTNTNLQGGIGQSQSPLQSSDLLSWQDPEGHQGTQFPYIIP GLPVNLWGRDVMKMGVYLYSPSSQVTQQMFDQGLPGQGLGSVGQGRREPISTNPNLQRTGLG YFPQGS
226	ORF number 105 in reading frame 1 on the direct strand extends from base 122728 to base 123048 ctttgatgcctatgttaagagtgcagctgaacgtttctggcatccttctgccgtccctgaggg tttggtcagaaagaaggatccacttactggatcatggcaaggcccagaccagctcctcatatgg ggccgagggcatgtttgtgtttttccacaggatgcagatagtcctcggtggctgccagaacgat tggtcgcacatgtggaccctctacctgctgatgacattgatgaccctcagcaatgcagaagaag accagacgtatttgggcctacgtacctgatccacctattctccaccctgctgtatscadmatgta a
227	Translation of ORF number 105 in reading frame 1 on the direct strand LWMPMLRVQLNVSGILLPSLRLWSEIRIHLDDHGKAQTQSSYGAEGMFVFFHRMQIVLGGCQND WCDMWTLTYLLMTLMTLSNAEEDQTYWAYVPDPPIILHPAVXXM
228	ORF number 106 in reading frame 1 on the direct strand extends from base 123565 to base 123798 ggctgtgagtgtcactgacataatctggaatctcaggaccatcccatcacagcagggtggagaata ggtgatcaggtacgtaggcccaatacgtctggtcttcttctgcattgctgaggtcatcaatg tcatacagcaggtagagggtccacatgtcgcaccaatcgttctggcagccaccaggactatctg catcctgtggaaaaacacaaacatgccctcgcccccatatga
229	Translation of ORF number 106 in reading frame 1 on the direct strand GVSVDIIWNLRTPYSRVENRWIRYVGPIRLVFFCIAEGHQCHQQVEGPHVAPIVLAATEDYL HPVEKHKHALGPI
230	ORF number 107 in reading frame 1 on the direct strand extends from base 125896 to base 126126 GCGGTGAGGACGTGTGCGCCCTTCCTCCTTCTCTTCTCGACTCCATCTTCGCGGTAGCGGTA GCGGCCGACGTTAGGTAAGATTTGGGCCACGGCTGGATCCGGACGACTTAATAGGTTAGCCGC GAGGTCTGACGGCTTGGGAAAAATAGAGGAAGAGGGGCTGCTCTGTGGGCCGGGTCTTGTGTCAC CACCCGACCTCCCTGGCTGGCCTGGCCTTAGGCACGTGA
231	Translation of ORF number 107 in reading frame 1 on the direct strand

	AVRTCAPFLLPLSRLHLRSGSGRSSGKIWATAGSRLNRLAARSDGLGKIEEEGLLCGPGSCH HPTSLAGLALGT
232	ORF number 108 in reading frame 1 on the direct strand extends from base 126127 to base 126387 GACCCGCGATCGTCCCCGGCCGCCACCCACTCCCCGACTCCCTTACTCCCAGAGCATTCTTC TCTTACAAGCATTCTTCTCCTCAGTCGCCGACATGCAGCTCTTGTTCGCGCCCAAGATCTACA CACCTCGAGGTGACCGGCCAGGAGACTGTCTCCAGATCAAGGTAAGGCTGCGTGGTGTCTCCT GGTCTGCATCCTCTTGTGTTCTTTAACCTCGCTCCCCACGGGAGCGCTGAGCCTCACTTTCCCC TGTAG
233	Translation of ORF number 108 in reading frame 1 on the direct strand DPRSSPARHPLPDSLTPRAFLLLQAFLLSSVADMQLFVRAQDLHTLEVTGQETVSQIKVRLRGAP GLHPLVFFNLAPHGSAEPHFPL
234	ORF number 109 in reading frame 1 on the direct strand extends from base 126961 to base 127260 AGTCCATGGTTCCTTGCCCCGTGCTGGGAAAGTAAGAGGTCAGACTCCCAAGGTAAGAGAGTAT TAGTGGTGCCCTTTGGACTTTTGTTCCTGTACCTTCCTCATGAAATGAGCCTGAGGGAAGG CACGGAAGAGATGAACCAGGGTCTGATTAGCCCTCCTTTTCCAGGTGGCCAAACaggagaag aagaagaagaagaCTGGCCGAGCCAAGCGCGGATGCAGTACAACCGCGTTTGTCAATGTTG TGCCACCTTTGGCAAGAAGAAGGGCCCCAATGCCAACTCTTAA
235	Translation of ORF number 109 in reading frame 1 on the direct strand SPWFLGPCWESKRSDSQGKRVLVVPFGLLFSCHLPHEMSLREGTEEMNQGLISPPFSQVAKQEK KKKKTGRAKRRMQYNRRFVNVPVTFGKKKGPNANS
236	ORF number 110 in reading frame 1 on the direct strand extends from base 129976 to base 130284 ccttgatgcccattgtaagagtgtgcagagcgcttttggcatccttcactgccactcaggc tttgttcaaatggaaagaccacttacaggctccttggaaggcccagatccagtcctcatatgg ggccgaggcatgtttgtgtttttccacaggatgcagaaggccctcggtggttgccagaacgat tggtgcgacatgtggaccctctacctgtgatgacattgatgascadmTTACAAAACCTTTCCAA ATGTTGTTTTATGTGTCCATCAGAAggcaatcctagggcggtgagtgtcattga
237	Translation of ORF number 110 in reading frame 1 on the direct strand PWMPMVRVLQSAFGILPLPLRLCSNGKTHLQALGKAQIQSSYGAEGMFVFFHRMQKALGGCQND WCDMWTLYLLMTLMXXLQNFNVVLCVHQKAILGRECH
238	ORF number 111 in reading frame 1 on the direct strand extends from base 130801 to base 131133 aggggaatgggacttaattggggaacagtgtgtacttccaggacattttccaagtcaagttgtc ctttcagtccttagttgtggagggcactgttcagccccaggtccagttgccgttgtagttgcag gggtggagcccagcacccttgcgggagttgaaccagcaagcttgtggttgagagcccactgg cccatgtgggctctggaaccggcagccttcaatgttaggagcacagagctccaaccgctgagc cactgggcgggcccACCCCCCTTTTTTTTTTTTAAAGAAAAAGTATTTTTTCTCTCAAAGC TTCCTTATATTAG
239	Translation of ORF number 111 in reading frame 1 on the direct strand RGMGLNWGTVCTSRTFKSSCPFLSCGGHCSAPGPVAVVSCRGWSPAPLAGVEPASLWLRHW PMWALEPAAFNVRSTELQPPEPLGRTPPFFFFKKKYFFLSKASLY
240	ORF number 112 in reading frame 1 on the direct strand extends from base 131335 to base 131946 GGGAGAATGAATGAATTAGCCTTTGAAGCTGATGTGTCTGATTGGTTCTTTTCTCTCAGGTG AAAAGCTCCGGGTCTTAGGCTACAATCACAATGGCGAATGGTGTGAAGCCCAACCAAAAATGG CCAAGGTGGGTTCACGCAACTACATCACGCCCGTCAACAGCCTGGAGAAACATTCTGGTAC CACGGGCCCCGTGTCCCGCAATGCTGCCGAGTACCTGCTGAGCAGTGGGATCAACGGCAGCTTCC TGGTGGCGGAGAGTGAGAGCAGCCCCGGGCAGAGGTCCATCTCGCTGAGATACGAAGGGAGGGT GTACCACTACAGGATCAACACAGCTTCGACGGCAAGGTgggcggggcggggcggcgggggcg

	ggcCTGAGTCTTGGGCCAGAACTCAGAGATCCCTCTGCTGGGTGGATAATGTTTTTACGACAAT ACTCGAGAAGTGGTTGGCAGACACTTTTCATGTAAACAGCAGGCGTCATTCATTAGCCTCATCGA TGATCCCCTGTGGAGGACTGATCATGTGACATTACAAGTCCACGGGCTGGGCTGGTTCTCTGGT TGTCCTGCTGGACGTTTGTGTTAACAGTTTCATAA
241	Translation of ORF number 112 in reading frame 1 on the direct strand GRMNELAFEADVSDLVFLSSEKLRVLGYNHNGEWCEAQTKNGQGWVPSNYITPVNSLEKHSWY HGPVSRNAAEYLLSSGINGSFLVRESESPGQRSISLRYEGRVYHYRINTASDGKVGAGRRGR GLSLGPRLRDPASAGWIMFLRQYSRSGWQTLSCKQQASFISLIDDLWRDTHVTLQVHGLGWFSG CPAGRLLLTVS
242	ORF number 113 in reading frame 1 on the direct strand extends from base 132532 to base 132804 GGGTGTAGCCAGATGGATTGTCGGTGTGGCTCCAGATGGTGTATATATTTTTTAGTAATATGTA ATGTATGCACACGGTTTTTAAAAAATCAATTACAGTGAAAGGTAATTTTCGTTTCTAGTTTAGT TCCCTGCCCAGAAGCAATCACTGTAACACCTTCTTCCAGAGTAACACGGTGTATATACACGG TGTATATActgtgtttccctgaaaataagacctaaccggacagtaagccctagcatgatttttc aggatgacgtccctga
243	Translation of ORF number 113 in reading frame 1 on the direct strand GCSQMDRCGRWCYFLVICNVCTRFLKKSITVKGNFVSSLVPCPEAITVTTFQSNSTVLYTR CIYCVSLKIRPNRTVSPSMIFQDDVP
244	ORF number 114 in reading frame 1 on the direct strand extends from base 134401 to base 134862 CTTGTAGAATTTGAGAGTCAGCCAATGAGGAAGCCGACCCCTCTGTCTAAAAGCTGGTGTGTGC TGGGGCTCCTTTCACTCCGGGTGGAACCTCAGGGAGTTCAATTGCTCAAGCACTGTCCACCCCCG GGCAGCTCGTCAGACAGTTCTGGGCTTCTCGccctcctccctccctccctccAGCTGTCTGAGC ACCTGGAGCCTCCTGGGCCTACAGGGTCATCGGGCAGACCCCTCTGCAGAGGCTCCTGCCTGTGT TGGGTGGGAGCACATTCCAAAAGGAGTGGAACAGTGTCTGCATGGGGAGGTACTCCAGTGATGC AGGCGACAGCCTGGCACTGAGGAGCTGCTCCAAGCGAGCTTTGAGGGGATCCTTTTAGGATTT CTAAGGGGAACATTTAAGGCTGGTAGGAGGGACAGGCTGGGGTTGAAGAAATTTAGTTCTTATT TTCAAATGAGCTGA
245	Translation of ORF number 114 in reading frame 1 on the direct strand LVEFESQPMRKPTPLSKSWCVLGLLSLRLVELREFICSSTVHPRAARQTVLGFSPSSLPPSSCLS TWSLLGLQGHRAADPLQRLLPVLGGSTFQKEWNSVCMGRYSSDAGDSLALRSCSKRSFEGILLGF LRGTFKAGRRDRLGLKKFSSYFQMS
246	ORF number 115 in reading frame 1 on the direct strand extends from base 136801 to base 137037 GGTAGTAGGATCGCTACGAAAAGACTGTCAGTTATAAAACCTCTGAGCCAGAGTTTGCTATTGG CTTGCTGACTTTTAACTGTCCATGTGTGTCTCATCTCCCAGAACagagagagagagagagagag agagagagagagaaaagagagagagaaATCTCCTTGTTAATGAATCCTGCTTACCTTCTTGAGGGT TATAGAAGGTATCAACTTGTATATGTTGTTATTTCTCTCTTTTAA
247	Translation of ORF number 115 in reading frame 1 on the direct strand GSRIATKRLSVIKPLSQSLLLA CLTFNCPVSSPQNRERERERERERERKRERISLLMNPAYLLEG YRRYQLVYVVISLF
248	ORF number 116 in reading frame 1 on the direct strand extends from base 137737 to base 138054 AAAGAGAAGAAAAATGATAGCTGTCCCCATCCACATTGCGCCCTCTGTCGTGTGCTCCTTTCCC TTCTCTCGTCTCAGTTGGTCCGGACGAGAACTCCTTGTGGAGGGGCTTCCTGCACAGGTGCTCA CCACTGTCCATCTCACAGGAGACTCATGTGCGTGTGTCTGAAAACCCTCTTCTGCCTTCCCCG CCATGGAAAAACCTGGATGGCCTTGGGCAGCCCTCCAGCCCTGCTCTGTTCTGGAGAGCACT GGCCAAGGAACACGGGGTGTATTACTGGGTACGCGGTGTACTGCAGGTCTTGATCTATGA
249	Translation of ORF number 116 in reading frame 1 on the direct strand

	KEKKNDSCHPHPCALCRVLLSLSSQLVTRTRPCGGASCTGAHHCPSHRRLMCVCLKTLFLPSR PWKNLDGLGQPSSPCSVPEGEHWPРНHGVVYVWTGCTAGLDL
250	ORF number 117 in reading frame 1 on the direct strand extends from base 138724 to base 139011 GGCTTCGCTGTGCATCGCGTTTCGTTAGCAGCAAAGCTGGTTCGTTGGCGTTGTTTGCGTTGGT GTCTGCTCTGTGGCCTGAAGGCTGTCCCTGTTTTCTCAGCTCTACGTCTCCTCAGAGAGCCGC TTCAACACTTTGGCCGAGTTGTTCATCATCACTCCACTGTGGCAGACGGGCTCATCACCACCTC TCCACTATCCAGCCCCCAAGCGCAACAAGCCCACCGTCTACGGCGTGTCTCCCAACTATGACAA GTGGGAGATGGAGCGCACGGACATCACCATGA
251	Translation of ORF number 117 in reading frame 1 on the direct strand GFAVHRVSLAAKLVRWRCLRWCLLCGLKAVPVFLSSSTSPQRAASTLWPSWFIITPLWQTGSSPL STIQPPSATSPSTACLPTMTSGRWSARTSP
252	ORF number 118 in reading frame 1 on the direct strand extends from base 139498 to base 139740 CCAAAAAGCGCTCAGCTCTTCTGTGGATTTTTGTTGGCAGATTTGAAATGCAAGTGCTGCTTAG TTCCTAGCAGGTTCTGTCTTTGTATTGTGTGCCAGACTTCTGGAATGAAGCAACATTAAAG GCTTCTTACTAACTCAGATCAGCCCTTCCCCCTTCTTTCTTGTATCTGTGACTTGCACCCTC GCCACTAATGCAGTGTTTTGTGGTTTCCAGGCGCTTTGTTTTCTTTTGA
253	Translation of ORF number 118 in reading frame 1 on the direct strand PKSAQLFCGFLADLKCKCCLVPSRFLFFVLVCVQTSGMKQTLRLLTNSDQFPFPSFLLSVTCTL ATNAQCLWFPGALFFF
254	ORF number 119 in reading frame 1 on the direct strand extends from base 142240 to base 142551 AAATCACTTCTTCCCCTCTCCCCTTCTCCGCCATTTGCCCCCTCAGAGTCTATAGCTGTGATC TACCTTGTCTCTTCAAGACTCCTTGGAACCCGTCAGCTCCAGCTCCAGCTTTTCGTTTGCTCA GCGGTTCTCACCAGACCTCTTCACTCTCCATGCCAGTCCCTACTGGGCACCTGAGTCTCGG TCCCCTCCTGCCTCCCTGTCTGCCTGTTTTGCCTTGCTGGCCCCGCAAAGGGCAGTGCCAGCT CCTCCTTAGCCAGCAGGGGGAGCAAGGCCGGACTTTTAACCGCGACTCCATATTGA
255	Translation of ORF number 119 in reading frame 1 on the direct strand KSLPLSPSPPFAPLRVYSCDLPCSSRLLGKPVQLQLSFAQRFSPSTSSPLHASPHWAPESR SPPASLSCLFCLAGPAKGSASSLSARGSKAGLLTATPY
256	ORF number 120 in reading frame 1 on the direct strand extends from base 143080 to base 143724 AAGCACATGGCAGCATGCTGTGGACACTGGTCTGTAGCCTACTGTCCACTGACTGTATCCGCAC AGCTGTTCCTTGTGCGGTACACATAAGGTCGCCTGTTTTTATGTGGTGGATGTCAGCATGTAGC AGCCCTCTGTGGGCATTTGCGTTCTTCCCAGTGCGTGGCTGTTACAGAAGTGCTGCAGGGATTC TCCTTGTTTGCACAGGGGACAGTGTCTTGAGGGGCCAGCACTCAGAGGGGAACGACTGCGTC AGGGGCGGTGTGTTTTGTGCTCTTCTCACACTCCCAAAGCCTCCCAAGGAGCTCGTACCTGT CTGCGCTCTGCGCGCGTGTTTGGGGAGTGCCTGCTTCCCGTCCCTGCACACTGACACAGTGTGCT TTGCTTTGGGGTTTATTTTTGTCATTTTCCCCCAGGAAATTTATTGGCAAGCTCAGAAACGAGC AGAGAAGGAAAGGTTCCGTGACAGCACTGACACTAGACCGGCCACGCAGTGGCCATGTGACTA CGCGGGGGGTGTGCACCAGGGAGAGGCCACCATTGCCGTGTGGCACTTGCTGTTACACTGGGTT CTCTTCTGGCTGTGCAGCGAGACCCAGCTGCCGTGTTTGGGGACCAGACTTCTGGGGGCTCCTC TGTGA
257	Translation of ORF number 120 in reading frame 1 on the direct strand KHMAACCGHWSVAYCPLTVSAQLFLVGTHKVALFLCGGCQHVAALCGHLRSSQCVAVTEVLQGF SLFAHRGQCPGPGPALRGERLRQGPCVFVVFLLTPKPPKELVPVICALPRVLGECLLPVPALTQCA LLWGLFLSFSRPFKFIGKLRNEQRRKGSVTALTLDRLPTQWPDCYAGGVHQGEATIAVWHLLLHWV LFWLCSETQLPCLGTRLLGAPL
258	ORF number 121 in reading frame 1 on the direct strand extends from base 145531 to base 145887

	CTTGTCTCTGGAAGTCTTCCCTCAGATCCGCGGCCAGCGGCGAATGCGGCAATCCTGGGCAGT TGTGCCGTAAGCACACCTTAGAGCCTGGTCGCCCCGAGGGGAGGTCCCACATTTCATAAACT CGATAAAGCTTTCTTCTTGGGGGAGGCTAGTTTCAAGACGTTCACTCCCCATCTCCCATACAG TCTTTCTCTTCAGACAATTCAAACCTCCCTGTGGAACTTGAAGGGTGGGCTCTTGCCCTCCCTGG TGGGCCCTTTGTAGCCAAGTTCTCACAGCAAACAGATCGTGTCAATTTACCGCCACCCGCTTCCTG TTTTGAGGGTCAGTTCAGAGGACAGTGGGTCTTTAA
259	Translation of ORF number 121 in reading frame 1 on the direct strand LVLWKSSLRSAASGECGNPGQLCRKHTLEPGRPEGQVPHFNKLDKAFFLGEASFQDVHSPSPIQ SFSSDNSNSLWKLEGWALASLVGLCSQVLTANRSCHLPPASPFCFEGQFRGQWVL
260	ORF number 122 in reading frame 1 on the direct strand extends from base 146674 to base 146928 TTTCACTACCTTTTTTCTACAGGAGGACACCATGGAGGTGGAAGAGTTTTTGAAGGAAGCTG CGGTAATGAAAGAGATCAAGCACCTAACCTAGTACAGTTACTTGGTGAGTGCGAGGAGCTCGG AAGGGGGGCGCTTTGCATTAAACCCGCTGGGGTGATCCAGGTGCTGTCAAAGAGGAGATGGCTG CCTCGCTACATGAATTCTTCTCATTGGACATCTGTTCTCTACTAACATTACGCCCTCGGTAA
261	Translation of ORF number 122 in reading frame 1 on the direct strand FHYLFFLQEDTMEVEEFLKEAAVMKEIKHPNLVQLLGECEELGRGGLCIKPAGVIQVLSKRRWL PRYMNSSHLIDICSLTFSPR
262	ORF number 123 in reading frame 1 on the direct strand extends from base 147094 to base 147399 TTTAGGCCATTTGATGTGTGCCTGGCCTTTGCTTCTGAACTCGGTGGCAGCCTCTTCTGTTTA AGTTCATTGGCTTGAGAGGAAGAAAAGAGCAGGCCATGTACCACCCCTGTCTCCCCCCCCAGA AACATCATCTCAAGTCACAGGTGCTTGAACCGTCTTAGCACTGAGTCCAGGCTTGGGGGCGAG AGTCAGATCCATTTCAAGAGCCTTTTCTTGAGGTCCAGTCTTTCTGATGCCCTGTGCTGTGTC TCGTTGGCAGGGTCTGCACCCGGGAGCCCCGCTTCTATATAATCACTGA
263	Translation of ORF number 123 in reading frame 1 on the direct strand FRPFDVCLAFASELGGSLFLFKFIGLRGRKEQAMYHPLSPPPETSSQVTGAWNRLSTESRAWGQ SQIHFRSLFLEVQSFLMPVLCLVGRGLHPGAPVLYNH
264	ORF number 124 in reading frame 1 on the direct strand extends from base 147445 to base 147708 CCGGCAGGAGGTGAACGCTGTGGTGCTGTGTACATGGCCACGCAGATCTCGTCAGCCATGGAG TACCTGGAGAAGAAAACTTCATCCACAGGTAGGAGCCTGCCGAGGCCGCTCCCCACAGGGCC CCGGCACCTTCTGTAAAAGGCCCCACCTTGAGGGGTGACCGCTCGGCCTCTCCCTCAGTGCT GGCAACATGTTAGGTCTGAGACAAGAGCGCAGCGGTGGGTCCGACGTGGCCAGCTCTGGGTGT GTGTCTAG
265	Translation of ORF number 124 in reading frame 1 on the direct strand PAGGERCGAAVHGHADLVSHGVPGEELKHPQVGACRGRRLPTGPRHPSVKGPTRLGRDRSAPFSA GNMLGLRQERSGGFRRGQLWVCV
266	ORF number 125 in reading frame 1 on the direct strand extends from base 147796 to base 148275 GGGGCATACTCAGTGTTTCATACAAGGAGTCGAGTGCTCCTTGTTCCGCCGAGCCCAGCCGGCG GGCGCCGTAGTGACCTCTTCCCCGAGCGGGTGGCCCTGCCCTGACACACGGCAAGAGCGGCCA GTGCATGGGTTTCGGTTTTGTGCTGCGTGTTTTTTTTCTCCCTTCTCTTTATTATCATTTCATT CTCCACTTAACCTTGCTGTACCGGCCCTCGCAATGTTTCCACAATTGGCAGAATTGTGTAGATG CGGCTCTAAGTGAAGTGCTTTGCTGTTTCAAAGCCCGAGTGTTGTGACCTTCAGGTGCGCCA CAATTATCCTGGTCTTCACATTCTTTGCTGGTGAAATGGCTTCCTAGCAGAGTGACAGCCTAT CCAGGGCAGAGCCTGTGGGCTTTGCCAGAGTCGTTTCATACAAGACATTCTCTGCCACCACTG TGACCTTTCCTGTCCAATTATCTCGACTATGA
267	Translation of ORF number 125 in reading frame 1 on the direct strand GAYSVFHTRSRVLLVPPSPAGGRRSDLFPGAGGPALTHGKSGQCMGFGFVLRVFFLPSLYYHFI

	LHLTCCHRRPQCFFHNWQNCVDAALSEVSLLFQSPECCDLQVRHNPGLHILCWWKWLPSRV TAY PGQSLWALPESFIQDILSATTVTTFPVQLSRL
268	ORF number 126 in reading frame 1 on the direct strand extends from base 153391 to base 153885 AAAAAAAAAAGGAAACCAACATACCAACATGACAGCATTACTGATGGCTGCTGCTTTTTgtgtt gtttttgtgtgtgtgtgtgtatgtgTTCCTTAGAAGTGGAAAAGGAACTGGGGAAAAAAGGCAT GCGAGGGGTTGCAAGCACTCTGCTGCAGGCCCCAGAGCTGCCACCAAGACAAGAACCTCCAGG AGAGCTGTGGAACACAAAGACCCACCGACGTGCCGAGACACCCCACTCCAAGGGCCCGGGAG AGCCTGGTATGTCTGCACCCCAACCCCACTGCAGGCTCAGGGTCAGTGCCCTTAGGGCCAGGGT GGCAGACGGGGAGCAGTGCGCGCAGCCTGCACAGAAAGGCAGGCAAACTCCCATTAGTTGTCCA CGGTGGAGAAAGTTCTCTCTCCCTGCAGCATCCACCCCTCCCTCTGGGAATCGTTAGGGGCC ATTGCTTCAGCAGGTAGTTCAGTCTGATGGGCAGAGGTGCTTCTGA
269	Translation of ORF number 126 in reading frame 1 on the direct strand KKKKRPTYQHDSITDGCCFLCCFCVCMWFLEVEKELGKKGMRGVASTLLQAPELPTKTRTSR RAVEHKDPTDVPETPHSKGPGEPGMSAPHPHCRLRVSRARVADGEQCAQPAQKGRQTPISCP AVEKVLLSLQHPTLPLGIVRGHWLQQVQSDGQRCF
270	ORF number 127 in reading frame 1 on the direct strand extends from base 155347 to base 155637 AAACTGAAAAGGTCACCCCTTCTTGTTCCTCAAGCATAATGGCCAGTGCTCACTGCACTCTGT GGGATGTGTCCCGTTCCTCCAGGTCACACCTGTAGAAACCACAGTTGGCTGGTCTGAGAGG CACAGTTATGACCCTTGTCTCGGCCGTGCATAGTTTTTACTCACAAGATAGTGAGGGGACTC TGCAGATATAAAGGAAACAGTGACAGGGTGGGGGAGACGGGGACGTCCCGGCTTTTTGTCTG CTGTCTCAAGGAGAGACCTAAGCTCTTCTCTaa Translation of
271	ORF number 127 in reading frame 1 on the direct strand KLEKVTSPCFPSIMAQCHCTLWDVSRSLQVTPCRNHQLAGLRGTGYDPLLGRVIVFTHKIVRGL CRYKGNQCRGGGDGDPVAFCSAVFKERDLSSS
272	ORF number 128 in reading frame 1 on the direct strand extends from base 156277 to base 156714 GTGCGCGGGGGGGCGCGCCAGTGGGGGGGGCGCTGGAGTTGGGGCGGCAGGGCCGAGC GGCCCGGGGCGGGGAGTCGCTGTCTCGCCGAGCGCGCGGGCGCACGGGGCGCAGGTGAGCCG CGGGCGGGGCGCTGCGGCTGGGGGCTGGGGCGGCAGGGCGGCTTCGTGTGCCACTCGGCCTCG GCAGGCCAGCTCTTCGAGCTCCGTGTCCCTGGCTCTGTCTCTCTGGGACCCCAAGTGCCCT CAGGAAGGCTGTGGGGTTCCTCTGCGCCAGGCCCCACCGTGGCCATGCGCTAGGAGGTGTCTC CCACCCGCGGAGTCCCAAGGACCCCTCCCAAGAGCTCGGGCACCTGCGGCCATCACACCCAA CAGGCGAGTCGGGTGTAGGAAGTCCACTGCTCACAAGGGCACCCCTCATTA
273	Translation of ORF number 128 in reading frame 1 on the direct strand VRRGAAAASGGGRWSWGGRAPERPGAGSRCPRRARGRTGAQVSRGRGAAAGGWGRQGGFVCHSAS AGQLFELRVPGSVLLGTPQVPSGRLWGSAPRPTRGHALGGVSHPPESQGPLPRARAPCGHHTQ QASRGVGSPLLTRAPPH
274	ORF number 129 in reading frame 1 on the direct strand extends from base 156715 to base 156966 ACATCAGAAATTGGAGACACCCCGGATGGATGGGGGCTTGGCCCCAAACCCTTTTTCTGTCCC ACCTGTTTCCGTGCCCTACACCTCCTGTGGGTCTTTTCTGTCTGTGAGTCTGTGCTTACCAG GGGGAACCTTGGGCCACAGGGCCTCCTCACTACCTGCCTTGTTCCTCAGAACTTCTCATGG CTGCAGGCCCCATGGGTTCCTTAGTTAACTTatgtgggtcttctccttgagcgtaa
275	Translation of ORF number 129 in reading frame 1 on the direct strand TSEIGDTPDGWGPWPQTLFLSHLFPCTSCGSFLVCESVLTRGNPGPTGPPHSPALFSQNFWS LQAPWVSLSLTYVGLLLGA
276	ORF number 130 in reading frame 1 on the direct strand extends from base 157057 to base 157377 atacttgtcgaatgCACCGACATGCCAGTGGGGCCTGGAACCTGTCGTCGGTTGGCACTGGCC TGCTTGGGCACGCTGCTGTGTGCTCCACCGTGGCAGGACCTGTTCCCTTAGGGAGGGGGACTGG

	TGACCTCAGCCTGGGCGCCTCCAGTTCGGGCTTTCTGCCTACTCAGCAACTTCTAATTTGGGTG CGTGGTTGGGAGATGCTCTCAGCTGTCACTCCTGCCCTTGGGGGGCCAGCTTCCTGCCTCTCAC AGCCATTAAGTGCAGCTGGACGCAGGACCCCTGTCCCCTCCTGGGCTGCAGGAGCCACAGGTG A
277	Translation of ORF number 130 in reading frame 1 on the direct strand ILVECTDMPSGAWNLSVGTGLPGHAAVCSTVAGPVPLGRGTGDLSLGASSSGFLPTQQLLIWV RGWEMLSAVSPALGGPASCLSQPLSAAGRRTVPVLLGCRSHR
278	ORF number 131 in reading frame 1 on the direct strand extends from base 157717 to base 158037 CACAAAGCTTTTCTGCCTGTTGCACCGAGGGGACCCTCGTCCCTCGGACCTGAGGGCACAAGAGG TGCAGGGAGGGGCTCGTGGTGCACATACTGCGTCCCAGGAGGGGTGGGGTCCCTAAGCAGTGT CCTCGCGCAGGACTCCTACCGGAAGCAAGTGGTCATCGATGGGGAGACGTGTCTGCTGGACATC CTGGACACGGCGGGCCAGGAGGAGTACAGCGCCATGCGGGACCACTACATGCGCACCGGCGAGG GTTTCTCTGCGTGTGGCCATCAACAACCAAGTCCCTTTGAAGACATCCACCACTACCGGTG A
279	Translation of ORF number 131 in reading frame 1 on the direct strand HKLFCLLHRGGPSSDLRAQEVQGGARGAHTASQEGWGSLSVLAQDSYRKQVVIDGETCLLDI LDTAQGEYSAMRDQYMRTEGEGFLCVFAINNTKSFEDIHQYR
280	ORF number 132 in reading frame 1 on the direct strand extends from base 158281 to base 158505 GCTGGCTCCCTGCCCACCTGTAGCCAGGGCCCCGCCCCGCCAGGGAGCCGTGCTCACC GC CCCTCTCCCTCGACACAGGGCAGCCGCTCTGGCTCCAGCTCCGGGACCCCGGACCCAGCGGCC CCTCGCGCTGTscadmCGGAGCCCATGCGCGGAGGAGCTgcgcgccccggccccgcccccgc ccgacccggccccggGGGCTGTCTGCTCCAGTGA
281	Translation of ORF number 132 in reading frame 1 on the direct strand AGSLPTCSQGPAPAREPCSPPLSLDTGQPLWLQLRDPGTQRPLALXXRSPCAGGAARPGPRPR PTRPGGLSLQ
282	ORF number 133 in reading frame 1 on the direct strand extends from base 158506 to base 159063 GCGGTAGTGCGGCGGGGGCGGCCGCGGCCAGTGGGGGGGGCGCTGGAGTTGGGGCGGCAGG GCCGAGCGGCCCGGGGCGGGAGTGCCTGCTCGCCGAGCGCGGGCGCACGGGGGCGCAGG TGAGCCGCGGGCGGGGCGCTGCGGCTGGGGGCTGGGGGCGGCAGGGCGGCTTCGTGTGCCACTC GGCCTCGGCAGGCCAGCTCTTCGAGCTCCGTGTCCCTGGCTCTGTCTCCTTGGGACCCCAAA GTGCCCTCAGGAAGGCTGTGGGGTTCCCTGCGCCGAGGCCACCCGTGGCCATGCGCTAGGAG GTGTCTCCCAACCGCGGAGTCCCAAGGACCCCTCCCAAGAGCTCGGGCACCCCTGCGGCCATCA CACCAACAGGCGAGTCGGGGTGTAGGAAGTCCACTGCTCACAAGGGCACCCCTCATTAACA TCAGAAATTGGAGACACCCCGGATGGATGGGGCCTTGCCCCAAACCCCTTTTCTGTCCCACC TGTTTCCGTGCCCCACACCTCCTGTGGGTCTTTTCTTGTCTGTGA
283	Translation of ORF number 133 in reading frame 1 on the direct strand AVSAAGGGRGQWGGALELGRQGRAARGESLSSPSARAHGGAGEPRAGRCGWGLGAAGRRLRVPL GLGRPALRAPCPWLCPWDPTSLRKA VGFPCA EAHWP CARRCLPPAGVPRTPPKSSGTLRPS HPTGESGCRKSTAHKGTPSLNIRNWRHPGWMGALAPNPFVPPVSVPLHLLWVFSCL
284	ORF number 134 in reading frame 1 on the direct strand extends from base 159424 to base 159651 CCTCAGCCTGGGCGCCTCCAGTTCGGGCTTTCTGCCTACTCAGCAACTTCTAATTTGGGTGCGT GGTTGGGAGATGCTCTCAGCTGTCACTCCTGCCCTTGGGGGGCCAGCTTCCTGCCTCTCACAGC CATTAAGTGCAGCTGGACGCAGGACCCCTGTCCCCTCCTGGGCTGCAGGAGCCACAGGTGAGC GGTCGGCCGTTGTTTCGGCTGCTACCCTGATGCCTGA
285	Translation of ORF number 134 in reading frame 1 on the direct strand

	PQPGRLLQFGLSAYSATSNLGAWLGDALSCQSCPWGASFLPLTAIKCSWTQDPCPTPGLQEPQVS GRPLFGCYPDA
286	ORF number 135 in reading frame 1 on the direct strand extends from base 159919 to base 160251 GCGGGGCTGACTCCCCGCCAGCCCTAATCCTGACACAAGCTTTTCTGCCTGTTGCACCGAGGG GGACCCCTCGTCTCGGACCTGAGGGCACAAGAGGTGCAGGGAGGGGCTCGTGGTGCACATACTG CGTCCCAGGAGGGGTGGGGTCCCTAAGCAGTGTCTCGCGCAGGACTCCTACCGGAAGCAAGT GGTCATCGATGGGGAGACGTGTCTGCTGGACATCCTGGACACGGCGGGCCAGGAGGAGTACAGC GCCATGCGGGACCAGTACATGCGCACCGGCGAGGGTTTCTCTGCGTGTTCATCAACAACA CCAAGTCCTTTGA
287	Translation of ORF number 135 in reading frame 1 on the direct strand AGLTPRPALILTQAFLPVAPRGTLVLGPEGTRGAGRGSWCTYCVPGGVGVPKQCPRAGLLPEAS GHRWGDVVSAGHPGHGGPGGVQRHAGPVHAHRRGFPLRVCHQQHQVL
288	ORF number 136 in reading frame 1 on the direct strand extends from base 160252 to base 160539 AGACATCCACAGTACCGGTGAGCTGCCAGCACCCGCGCAGGCCGTCCCTTCTGGCGCCCTGGA CGCAGCCTGCCGGTGGCTCACACCATCCTTCTGCAGGGAGCAGATCAAGCGGGTGAAGGACTC GGACGACGTGCCATGGTGTCTGGTGGGAAACAAGTGTGACCTGGCTGCACGCATGTGGAGTCT CGGCAGGCACAGGACCTGGCCCGCAGCTACGGCATCCCCTACATCGAGACCTCGGCCAAGACGC GCCAGGTGAGCTGGCTCCCTGCCACCTGTAG
289	Translation of ORF number 136 in reading frame 1 on the direct strand RHPPVPVSCQHPRRPSLLAPWTQPAGGSHHPPCREQIKRVKSDDDVPMVLVGNKCDLAARTVES RQAQDLARSYGIPYIETSAKTRQVSWLPAHL
290	ORF number 137 in reading frame 1 on the direct strand extends from base 160720 to base 161094 gtcaattttacaaaaataaaaaagggggagttgtatcccctgacgccccataattacctgtctc attctctctttattcaaaatttttgaccttgatgcccatggtaagagtgtctgcagagcgctt ttggcatccttctactgccccctcaggctttggtcaaatgaaagacccatttacaggctcttg caaggcccagatctagtccctcatatggggccgagggcatgtttgtgtttttccacaggatgcag aaggccctcggtggtgcccagaacgattggtgcgacatgtggaccctctacctgctgatgacat tgattactctcagcaatgtagaagaagaccagacatatattggcctatgttcctga
291	Translation of ORF number 137 in reading frame 1 on the direct strand VNLQKIKKGELYPLTPHNYLSHSLFIQNFLTLDANGKSAERFWHPSTAPQALVKWKDPFTGSW QGPDLVLIWGRGHVCVFPQDAEGPRWLPERLVRHVDPLPADDIDYSQQCRRRPDILGLCS
292	ORF number 138 in reading frame 1 on the direct strand extends from base 163255 to base 163488 GGCGTGAGTGTCATTGACATAGTCTGGAATCTCAGGaccttccatacagcagggtggagaata ggtggatcaggtacgtaggcccaatacgtctggtcttcttctgcattgctgagggctcatcaatg tcatacagcaggtagagggtccacatgtcgcaccaatcggtctggcagccaccgagggccttctg tatcctgtggaaaaacacaaacatgccctcgcccccatatga
293	Translation of ORF number 138 in reading frame 1 on the direct strand GVSVIDIVWNLRTFPYSRVENRWIRYVGPIRLVFFCIAEGHQCHQQVEGPHVAPIVLAATEGLL YPVEKHKHALGPI
294	ORF number 139 in reading frame 1 on the direct strand extends from base 163810 to base 164130 ccggagccattatctgttttaagtttttaggagtggcagaagggtgtggtaacccscadmtgg tcaaatggaaagaccacttacgggctcttggaaggccagatccagtcctcatatggggccg aggcatgtttgtgttttccacaggatacagaaggccctcggtggctgccagaacgattggtg cgacatgtggaccctctacttgctgatgacattgatgacctcagcaatacagaagaagaccag acgtattscadmcaagcaGATACATTAACAGATTTTTTAGACCAGTCTCTAGTCCCATCTTGTA A

295	Translation of ORF number 139 in reading frame 1 on the direct strand PEPLSVLSFLGVAEGCGNPXXVWKDPLTGSWQGPDPVLIWGRGHVCFVFPQDTEGPRWLPERLV RHVDPLLADDIDDPQQYRRRPDVXXXSRVINRFFRPVSSPIL
296	ORF number 140 in reading frame 1 on the direct strand extends from base 164356 to base 164601 agggtccacatgtcgcaccaatcattctgtgcagccaccgagggccttctgcatcctgtggaaaa acacaaacatgccctcgccccatatgaggactggatctggccttgccaagagcctgtaagtg ggtctttccatttgaccaaagcctgagtggcagcagaaggatgccaaaagcgctccgcagcact cttaccatgggcatccascadmCTCTAGTCCCGTCTTGTAATCAGTCACCTGA
297	Translation of ORF number 140 in reading frame 1 on the direct strand RVHMSHQSFQPPRAFCLWKNTNMPSPAPYEDWIWALPRACKWVFPFDQSLSGSRMPKALRST LTMGIXXXLVPSCSVT
298	ORF number 141 in reading frame 1 on the direct strand extends from base 164788 to base 165093 gggtcatcaatgtcatcagcaggtagagggccacatgtcacaccaatcgcttctgcagccacc gagggccttctgtatcctgtggaaaaacacaaacatgccctcgccccatatgaggactggata tggcscadmatttggccagcttaattcaagaaagccgtttggaagctcgaaaatattatgg gaaagagccagatttgattgtgttctctttacaaaaacacagattcaaggcttgatgcagttt acagacagttttcccatcgcttggctcattttgaggaaacttttagataa
299	Translation of ORF number 141 in reading frame 1 on the direct strand GSSMSSAGRGSTCHTNRSGSHRGPSVSCGKTQTCPRPHMRTGYGXXICGQLNSRKPFSSKILW ERARFDCCSFYKNTDSRLDAVYRQFSHRLGSFCRNFR
300	ORF number 142 in reading frame 1 on the direct strand extends from base 165112 to base 166104 attgcttcagtttttcaacatcatgatccaatttttcttcaattgtgtcacatgctcctcttc ctgcggtaccaaagtctttactgatggatctaacaatgggtgcgctggttatgactcaataa acaaattaaaaagatccagacacctccagcttcagctcaaatagttgagcttcgagcagttcat atgggtgttgcttgattttgcttcccagctcttttaatttattctctgacagccattatgtggttc gtgcagtcaaaaatttagaaacagtaccgtttattaataaccagtaatcctgttattcaggattt atttcttcagatacaacaagccattcagctgcgctgtaaaaaattttatattggccatattaga gctcactctagtcttccaggcccttttagcagcaggcaatcaaattgcagattctgccacgcagc ttattgccttaactcaaataagaaaaagcacaaaaggctcatagcctccaccatcaaaaacagcca gagcctaagattacagtataagatccccagagaagcagcagccagattgtaaagcaatgtcct gactgttcacatttacagcctgtgcctcattatggagttaaccctcggggcttgcgctcccaatg atctgtggcagacggatgtgactcatataacctgaatttgggaaattaaaatagctccatgtctc tatagacacgttctctggctttgttaattacttctggtcaatcaggagaagctacgtctcatgtt atcagacactgtcttgctgcttttgccatgattggcactcctaaaaaacttaaacagataatg gctccggctacaccagcaagaaatttgctttattttgcccagcaattttcaattaatcatgttac tggcattccttacaatccccaaggacaagggattgttgaacgcactcatggcacattaaaagtc attttcaaaaaataaaaaagggggagttatag
301	Translation of ORF number 142 in reading frame 1 on the direct strand IASVFQHHDPFIPSIIVSHAPLPVAPNVFTDGSNNGVAVYALNKQIKKIQTTPPASAQIVELRAVH MVLDFASQSFNLFSDSHYVVRVAVKNLETVPFINTSNPVIQDLFLQIQQAIQLRCKKFYIGHIR AHSSLPGPLAAGNQIADSATQLIALTQIEKAQKAHSLHHQNSQSLRLQYKIPREAAARQIVKQCP DCSHLQVPVPHYGVNPRGLRPNDLWQTDVTHIPEFGKLKYVHVSIDTFSGFVITSGQSGEATSHV IRHCLAAFAMIGTPKKLKTDNNGSGYTSKKFALFCQQFSINHVTGIPYNPQQGQIVERTHGTGLKV ILQKIKKGEL
302	ORF number 143 in reading frame 1 on the direct strand extends from base 166105 to base 166485 cccctgacgccccataattacctgtctcattctcttttattcaacattttttgaccttgatg cccatggtaagagtgtgcagagcgcttttggcatccttctactgccactcaggctttggtcaa

	atggaagactcacttacaggctcttggcaaggcccagatccagtcctcatatggggccgaggg catgtttgtgtttttccacaggatgcagaaggccctcggtggctgccagaacgattggtgcgac atgtggaccctctatttctgatgascadmGCCATGCACTGTGTCCGCGTCCCGCTCGCTACCA TTGGGAACACAGCAGCAGCCGCTGCAGCTCTCGCCCCTGAAGGGGCTCAGCCTAGCGGATAA
303	Translation of ORF number 143 in reading frame 1 on the direct strand PLTPHNYLSHSLFIQHFLTLDAHGKSAERFWHPSTATQALVKWKDSLTSWSWQGPDPVLIWGRG HVCVFPQDAEGPRWLPRLVRHVDPLFADXXXHALCPRPARYHWEPAAAAAALAPEGAQPSG
304	ORF number 144 in reading frame 1 on the direct strand extends from base 168031 to base 168300 TGCAACCAATGTCCAGTGACCCAGATTGCGCTGAACTTTGATGTGTTTACCACTAGGTGGAGCG GTTTAGCCAAGAAGTTCAAGATTACAGAAGCCCGCTGTTTCTATGGCTTCCAAATTGCCATGGAA AACATACATTCTGAGATGTATAGTCTCCTCATTGACACTTACATCAAAGATTCCAAGGAAAGGT GAGTATTTGAGTGGTATGCCAACATGTTTGGGACTCACTAATTGTTTATTTCAAGTTTTTGGAT TCAGACCGGGATAG
305	Translation of ORF number 144 in reading frame 1 on the direct strand CNQCPVTQIALNFDVFTTRWSGLAKKFRLLQKPAVSMASKLPWKTYILRCIVSSLTLSKIPRKG EYLSGMPTCLGLTNCLFQVGFGRPG
306	ORF number 145 in reading frame 1 on the direct strand extends from base 172837 to base 173121 GCACGCTCGGGCCGGGTGGGGGTGGCGGGTACCTGGGGGACTCGGGCATGCCTCTCACCGCATG TCTCCCCGAGCCACCCGCTCTCAACGGCACCCGCGTGTGGCCAGCAAGGCGGCCGAGGAT CTTCCAGGAGGCGGCGGAGTCCGTGGAGCCGGTGAGCGGATGCCCCAGGGCGGAGACAGCGCAG TGGGCGTGGCCAGCGCGCAGCGCCTGGGGGCGACAGCCGACTTCGCCGGCTCTCTGGCGCCATG GCTTCTTTGTCTTTCTACTTACTCATAA
307	Translation of ORF number 145 in reading frame 1 on the direct strand ARSGRVGVAGTWGTRACLSPHVSPQPPALNGTRVLASKAARRIFQEAESVEPVSGCPRAETAQ WAWPARSAWGRQPTSPALWRHGFLCLSTYS
308	ORF number 146 in reading frame 1 on the direct strand extends from base 173212 to base 173502 CAGCTGACACGTAAGACACTggaccacatgaaattgccgacaattgaatgtaactggatgggaa aaatggcaatttcatatggttcgaTGGATACTTACATTTTCATTACTTTCTCCCCAACAGAA AACTAAGGTGTCTGCCCTCAGCGGGCAGGATGAACCACTGCTGAGAGAAAACCCCCGCGCTTT GTCTGCTTTCCCATGAATACCATGATATCTGGCAGATGTATAAGAAAGCGGAGGCTTCCTTTT GGACAGCTGAGGAGGTAATCAGATTCAGGAGCTAG
309	Translation of ORF number 146 in reading frame 1 on the direct strand QLTRKTLDHMKLPITIECNWMGKMAISYGSMDSHFHYFLPQQKTKVSALSGQDEPLLRENPRRF VVFPIEYHDIWQMYKKAESFWTAEEVIRFS
310	ORF number 147 in reading frame 1 on the direct strand extends from base 178783 to base 179067 GCACGCTCGGGCCGGGTGGGGGTGGCGGGTACCTGGGGGACTCGGGCATGCCTCTCACCGCATG TCTCCCCGAGCCACCCGCTCTCAACGGCACCCGCGTGTGGCCAGCAAGGCGGCCGAGGAT CTTCCAGGAGGCGGCGGAGTCCGTGGAGCCGGTGAGCGGATGCCCCAGGGCGGAGACAGCGCAG TGGGCGTGGCCAGCGCGCAGCGCCTGGGGGCGACAGCCGACTTCGCCGGCTCTCTGGCGCCATG GCTTCTTTGTCTTTCTACTTACTCATAA
311	Translation of ORF number 147 in reading frame 1 on the direct strand ARSGRVGVAGTWGTRACLSPHVSPQPPALNGTRVLASKAARRIFQEAESVEPVSGCPRAETAQ WAWPARSAWGRQPTSPALWRHGFLCLSTYS
312	ORF number 148 in reading frame 1 on the direct strand extends from base 179158 to base 179448 CAGCTGACACGTAAGACACTggaccacatgaaattgccgacaattgaatgtaactggatgggaa

	aaatggcaatttcatatggttcgaTGGATACTTACATTTTCATTACTTTCTCCCCAACAGAA AACTAAGGTGTCTGCCCTCAGCGGGCAGGATGAACCACTGCTGAGAGAAAACCCCGCCGCTTT GTCGTCTTTCCCATCGAATACCATGATATCTGGCAGATGTATAAGAAAGCGGAGGCTTCCTTTT GGACAGCTGAGGAGGTAATCAGATTCAGGAGCTAG
313	Translation of ORF number 148 in reading frame 1 on the direct strand QLTRKTLDHMKLPTIECNWMGKMAISYGSMDTSHFHYFLPQQKTKVSALSGQDEPLLRENPRRF VVFPIEYHDIWQMYKKAESFWTAEEVIRFRS
314	ORF number 149 in reading frame 1 on the direct strand extends from base 186598 to base 186852 ctttggatgcccatggtaaaagtgcagctgaacgtttttggcatccttcaactagccctcaggc cttgggtcaaatggaaggaccacttacgggtgtcttggaaggcccagatccagtcctcatatgg ggcgaggcatgtttgtgtttttccacaggatgcagaaggccctcggtgggtgccagaacgat tggtgcgacatgtggaccctctacctgctgatgacattgatgascadmCTCCGCTTCAGCTAG
315	Translation of ORF number 149 in reading frame 1 on the direct strand LWMPMVKVQLNVFGILQLALRPWSNGRTHLRVSGKAQIQSSYGGEGMFVFFHRMQKALGGCQND WCDMWTLYLLMTLMXXLRFS
316	ORF number 150 in reading frame 1 on the direct strand extends from base 187354 to base 187623 gacaggagctgatgaatcttttcaagattttgtgtctcgccttactgttgctgcgggacggac ctttggagcgtccgtggctacggaggctttcattaacagcttgcttatgaaaatgcaaattct gcctgccaagcgattattaggccattaaagaaaaaggcactatctctgattttatccgttcct gtgccgatgtcggccctccttttcacaggagtgccctcggtgccgctttacaaggaaaaag cattcatgaagtaa
317	Translation of ORF number 150 in reading frame 1 on the direct strand DRELMNLFKILCLALLLLRDGPLERPWLRLRLSLNSLLMKMQILPAKRLLGPLRKKALSILSV VPMSAPPFHREWPWLPYKEKAFMK
318	ORF number 151 in reading frame 1 on the direct strand extends from base 187624 to base 187863 tgcagcaacaggccaagcttcatgctagtggccgagcagagcttgttttaactgtggaaaaat gggacatcgagcttctcaatgccacataaaatggaggctaacaatccgtcggtactgtgtg gttaaaaaacctccaggccttgtcccaggtacaagaaaggcgctcattgggctaataaatgta aatccaaaactgacaaagacggcaaaccttacagggaactgggtga
319	Translation of ORF number 151 in reading frame 1 on the direct strand CSNRPSFMLVAAQELVLTVEKWDIELLNAHIKWLRTIRLLLLWLKNLQGLVPGTRKALIGLINV NPKLTKTANPYRETG
320	ORF number 152 in reading frame 1 on the direct strand extends from base 188323 to base 188637 ttacttgtctttttattcaaaatttttttgacttttgatgcctatgttaagagtgcagctgaac gtttctggcatccttctgccgtccctgaggctttggtcagaaagaaggatccacttactggatc atggcaaggcccagaccagtcctcatatggggccgaggcatgtttgtgtttttccacaggat gcagatagtcctcggtggttgccagaacgattggtgcgacatgtggaccctctacctgctgatg acattgatgaccctcagcaatacagaagaagaccagacgtattgggcctacgtacctga
321	Translation of ORF number 152 in reading frame 1 on the direct strand LLVFLFKIFLTLDAYVKSAAERFWHPSAVPEALVRKKDPLTGSWQGPDPVLIWGRGHVCVFPQD ADSPRWLPERLVRHVDPLPADDIDDPQQYRRRPDVLGLRT
322	ORF number 153 in reading frame 1 on the direct strand extends from base 188725 to base 189525 tggacacatgaaacaacaTTTGAAAGTTTGTAAATCAGGCACTCCCTGCAGTCAGGTGACTG ATTTACAAGACGGGACTAGAGACTGGTCTAAGAAATCTGTTAATGTATCTGCTTGTGTTCTTC CCCTTATACACTTTTGATTGGAATATTAATGTACATTTGTAGGAGTTCAGTTTAtggaagat

	gtgattcagagtataaaagttaaattctattttaaaatgtcattcagaatatcattggatatgtg ttacttscadmccccggcgacggggcgcgggggggcgggggcggggactgtgccagtgcgcccc ggcggggtcgcgccgtcgggccccgggggtttccaggcgccacgccgtgaccaaagcacagcga agcgagcgcacgggggtcagcgggcatgtcggccaccacccgaccggtcttgaaacacggacca aggagtctaacacgtgcgcgagtcaggggctcgcacgaaagccgcccgtggcgcaatgaaggtga aggccggcgccgctcgccggccgaggtgggatcccgaggcctctccagtcgccggagggcgcac caccggccccgtctcgcccgagcgccggggaggtggagcacgagcgcacgtgttaggacccgaa agatggtgaactatgcctgggcagggcgaaagccagaggaaactctggtggaggtccgtagcggg cctgacgtgcaaactcggtcgctccgacctgggtatagggcgaaagactaatcgaaacctatgtagt agctgggtccctccgaagtctccctcaggatag
323	Translation of ORF number 153 in reading frame 1 on the direct strand WTHETTFGKFKSGTPCSQVTDLQDGRDWSKKS VNVSACVPSPYTL LIGNIN VHFVGVQFMED VIQSIKVKSYLKCHSEYHWICVTSXXPATGRAGGGADCAQCAPGGSRRRARGVSRRHAVTKAQR SERTGSAAMSATHPTRLETRTKESNTCASQGLARKPPWRNEGEGRRRSPA EVGSRGLSSPPRAH HRPVSPAAPGRWSTSARVRTRKMVN YAWAGRSQRKLWRRSVAVLTCKSVVRPGYRGERLIEPSS SWFPPKFPSG
324	ORF number 154 in reading frame 1 on the direct strand extends from base 189922 to base 190194 ccttggtatgcccatggttaagagtgtcgcggagcgcttttggcatccttctgctgccactcaggg tttggtcaaatggaaagaccacttacaggctccttggaaggccagatccagtcctcatatgg ggccgagggcatgtttgtgttttccacaggatgcagaaggccctcggtggctgccagaacgat tggtgcgacatgtggacctctacctgtgatgacattgatgaccscadmgttaggggtcatca atgtcatcagcaagtag
325	Translation of ORF number 154 in reading frame 1 on the direct strand PWMPMVRVLRSAFGILLPLRLWSNGKTHLQALGKAQIQSSYGAEGMFVFFHRMQKALGGCQND WCDMWTLYLLMTLMTXXLRVINVISK
326	ORF number 155 in reading frame 1 on the direct strand extends from base 190195 to base 190644 aggggtccacatgtcgcaccaatcggttctggcagccaccgagggccttctgcatcctgtggaaaa acacaaacatgccctcgcccccatatgaggactggatctgggccttgccaagagcctgtaagt ggcttttccatttgaccaaagcctgagtggcagcagaaggatgccaaaagcgctccgcagcact cttaccatgggcatccaaggtcaaaaaattttgaataaagagagaatgscadmGACCGGGCCGG GCTCATCGCCCGGCCGCCGCCGCCGCTTCTCGTtaatgateccttccgcaggttcacctac ggaaacctgttacgacttttacttctctagatagtcaggttcgaccgtcttctcagcgctcc gccagggccgtggggcgaccccgggggcgatccgagggcctcactaaacctccaatcggt ag
327	Translation of ORF number 155 in reading frame 1 on the direct strand RVHMSHQSFQPPRAF CILWKNTNMPSAPYEDWIWALPRACKWVFPDQSLSGSRMPKALRST LTMGIQGGKILNKERMXXTGPGSSPGRRRRFLVNDPSAGSPTETLLRLLLPLDSQVRPSSQRS ARAVGRPRRRRSEGLTKPSNR
328	ORF number 156 in reading frame 1 on the direct strand extends from base 191302 to base 191622 tcgtcttcgaacctccgactttcgttcttgattaatgaaaacattcttgcaaatgcttttcgct ctggtccgtcttgccgggtccaagaatttcacctctagcggcgcaatacgaatgccccggcc gtccctcttaataatcatggcctcagttccgaaaaccaaaaaatagaaccgcggtcctattccats cadmttgctgaggggtcatcaatgtcatcagcaggtagaggggtccacatgtcgcaccaatcggtc tggcagccaccgagggccttctgcatcctgtggaaaaacacaaacatgccctcgcccccatatg a
329	Translation of ORF number 156 in reading frame 1 on the direct strand SSSNLRLSFLINENILGKCFRSGPSCAGPRISPLAAQYECPRPSLLIMASVPKTNKIEPRSYX XXAEGHQCHQQVEGPHVAPIVLAATEGLLHPVEKHKHALGPI

330	ORF number 157 in reading frame 1 on the direct strand extends from base 191674 to base 191952 ccaaagcctgagtggcagtggaaggatgccaaaagcgctccgcagcactcttaccascadmtgt catcagcaggtagagggtccacatgtcgccaccaatcgttctggcagccaccgagggccttctgc atcctgtggaaaaacacaaacatgccctcgccccatatgaggactgggtctgggccttgccat gatccagtaagtggatccttctttctgaccaaagcctcaggagcggcagaaggatgccagaaac gttcagctgcactcttaacatag
331	Translation of ORF number 157 in reading frame 1 on the direct strand PKPEWQWKDAKSAPQHSYXXXSSAGRGSTCRNRSGRGPSASCGKTQTCPRPHMRTGSGPCH DPVSGSFFLTKASGTAEGCQKRSALLT
332	ORF number 158 in reading frame 1 on the direct strand extends from base 192412 to base 192966 CACTGCCCTTCCTTCGAGCACAGGCTGACCTCAGTGACAGATGAACTGGCTGCGGTCACCGCAG TGGTGTTCAGCCGGCAGGAGGTGGTCACCCAGCTGCAGCGCGAGCTGCGGAATGAGGAACAGAA CATCCACCCCCCGGCAGCGGTGAGTGGGTCCACCTATTGTAGCCTTGTGCCCCGCGCCCCACCCC ACACACCTGCCCTGCAGCCAGCTGCAGGCTGAGCCCTCTCTCTGCCCCCTCCACCTCCACCT GCCTGTCTCCTTTCAGGGTTTACCTGCTGGGCAAGAGGCAGGTATTGCAGGAGGAGCTCCAGGG GCTGCAGGTGGCACTGTGCAGCCAGGCCAAGCTGGAGGCCAGCAGGATCTTTTGCAGGCCAAG CTGGAGCAGCTGGGCCCCGGGGATCCCCCGCTGTGCCGCTCTACAGGACGACCGCCACTCTA CCTCCTCCTCGGTGAGTGCCTACTGCCCTCCGTGGTCACCTTGCTGCCAGCCAGGCTGTGTC CTCATTTTCGCCCTCCCCCTCCCCAAGCTGGCCACCCGCTGA
333	Translation of ORF number 158 in reading frame 1 on the direct strand HCPSFEHRLTSVTDELAAVTAVVFSRQEVVTQLQRELRNEEQNIHPRQRSVGPITYCSLVPAHP THLPCSQLQAEPCLPLPPPTCLSPFRVYLLGKRQVLQEELQGLQVALCSQAKLEAQDQLLQAK LEQLGPGDPPVPLQLDDRSTSSSVSALLPSVVTLLPAQAVSSFSPSPSPSLATR
334	ORF number 159 in reading frame 1 on the direct strand extends from base 192967 to base 193197 CGTCTGTCCCTGGCCTCAGGAGCAGGAGCGGGAAGGGGTACGGACGCCTACCCTGGAGCTCCTG AAGAGCCACATCTCAGGAATCTTTCGCCCAAGTTTTCGGTGAGTGGCACCTGTCTGGGCCTGC GCCTCTGCCCTTCTCAAGGGGTGGGCTGGGCCAGGGGTCTCAGACATGCCCCACTGCACCCC GCCACATGGGTGTTCTGTTAGCCCCCTGGGTTGCCCTAA
335	Translation of ORF number 159 in reading frame 1 on the direct strand RLSLASGAGAGRGTDAYPGAPEEPHLRNLSPQVFGEWHLSGPAPLPFSKGWAGPGVSDMPPLHP AHMVFWLAPGLP
336	ORF number 160 in reading frame 1 on the direct strand extends from base 193198 to base 193455 AGAGGAGGCTCTCTCCACGCCGCTTTTATTGGGGTGCCAAGCACCAACGTCCCCAGATCCTGCC ACTCTCACACCCCCCTTCTCTCTGCCATCACATGTGCTGAAGGGACTCACAGCTTTAGTGACCC CATGGCTCTCCCTGCTCCAGGAGTGGTTGGGGGGCCGAGCCTGGTGAAAAGGCAAAAGTTTG GTTTGGGACCACTCAGCCGGCCCCCCCCATCCAGCTGTGCCTGGGCCAGTCTATGGCCTGCTCT AG
337	Translation of ORF number 160 in reading frame 1 on the direct strand RGGSLHAAFIGVPSTNVPRSCHSHTPFFSAITCAEGTHSFSDPMALPAPGVVGGPQPGGKGKSL VWDQSAGPPIPAVPGPVYGLL
338	ORF number 161 in reading frame 1 on the direct strand extends from base 193816 to base 194112 CGTGAGTGGTGCCAGGACCCGCGCCACCTGCCCCACCTTCCCTGTCACCAGAATGACCTTG AGAGGGTAGGAAGAAAGGGGCTGCTAGTCTTAGATGCTAGTCAGAGCTGCAAGGGGCCATGGAG ACCACTTAGTCCCTATAACAGAACAGGCGTAAGTAGCATGGGTAGCAGGTGTGTTGGGCGCCAT GAGGTGCTGCCTTCCCTGCAGTGTCTCTGCCTCTCGTCCCAGGCAGGCCCTTTCTCCCTGCTACT CTCCCGCTCCCCCTCCAGGGCTCAGGCCCCCTCAGCAGTAG

339	Translation of ORF number 161 in reading frame 1 on the direct strand REWCQDPRPPCPTLPCHQNDLERVGRKGLLVLDASQSKGPWRPLSPYNRTGVSSMGSRVCVRH EVVPSCSVSASRPRQALSPCYSPAPLPGLRPPQQ
340	ORF number 162 in reading frame 1 on the direct strand extends from base 194113 to base 194427 AGGCTGCTGACCCCCAAGTTGCCCTGCCCTGCAGAACCTGTACCGACTGGAAGGTGATGGTTTTTC CCAGCGTCCCCTTGCTCATTGACCACCTGCTGCAGTCCCAGCAGCCCCCTACCAAGAAGAGCGG TATTGTCCTGAACAGAGCTGTGCCCAAGGTGAGCCTGCACCCACCGGCCACACCAACCCACCA CAGGGTTTGGGGAGCGCGGGTTCAGGCCACAGAATCGGGGCAGGAGGGGCTTCCAGGTCTCT GGTCTACGGTCTGGGTACCACGCGACTCCTCACTCTCCAAGGGGTGAGCTCCCTCCTAG
341	Translation of ORF number 162 in reading frame 1 on the direct strand RLLTPSPALQNLRYLEGDGFPSVPLLDHLLQSQQPLTKKSGIVLNRAVPKVS LHPTGPHHP QGLGSAGSGPQNRGRGFPLWSTVWVPRDSSLSKGSAPS
342	ORF number 163 in reading frame 1 on the direct strand extends from base 196108 to base 196377 GTGCGGGCAGGGCTCGTGCTGCCCACGCCAGCCCCCAGTAACCCCGCCCAAGCACAGGCCAT GCTGTACCCCGTGCCCTTTCCCGAGGGACCATGAGTCCTGGGCAGGGAGCGGCCCTTGTTT ATGTCTATGTGTGGAGTCCCCAGCTCAGGGAGGTGACGGGTGCGGTGTGTGGTGGCTGAGTGAG CCCCTTTCTGCTTTATCCAGGGACCTTGCTGCTCGGAAGTGCCTGGTCACAGAGAAGAATGTC TTGAAGATCAGTGA
343	Translation of ORF number 163 in reading frame 1 on the direct strand VRARPRAAHASPPVTPPKHRPCCHPVPPFPRDHESWAGSGPCSLCVESPAQGGDGC GVWVLSE PLSCFIQGPCCSELPGHREECLEDQ
344	ORF number 164 in reading frame 1 on the direct strand extends from base 196516 to base 196761 GGCTGGGCGTGCTCTGGCTGATGGACGTGGGTGGCTCACTCACACTGCCTCACCTCCTTGACG GCCGTATTTCGTCCGAGAGCGATGTGTGGAGCTTTGGCATCTTGCTCTGGGAGGCCCTTCAGCCT GGGGGCTCCCCCTACCCCAACCTCAGCAATCAGCAGACTCGGGAGTTTCGTAGAAAAAGGTAAG GCAACCCCACTGCATGACAGCAGCCGACCCACGCGCTCATCCAGTGCTATAG
345	Translation of ORF number 164 in reading frame 1 on the direct strand GWACLWLMDVGGSLTLPHELLAGRYSES DVWSFGILLWEAFSLGASPPNLSNQQTREFVEKGK ATPLHDSSPTHALIPVL
346	ORF number 165 in reading frame 1 on the direct strand extends from base 197161 to base 197598 CGCTGTGTTCAAGGCTCATGGAGCAGTGCTGGGCTACGAGCCAGTCAGCGACCCAGCTTCAGC ACCATCTACCAGGAGCTGCAGACCATCCGAAAGCGGCATCGGTGAGGCTCGGCCGCTTCTCAA GCCAGTGGCTTCTGTTGGCAAGATTATACCTCCTCCCCAGCTCCAGCTCACACCGTGGGACAGC CCTTCCAGTCTGGAATCTGGCCGCCGGCATCCATGCTGCCAGGGGGGATGCAGCTCCATGTC TGCTGTGCGTCCCCATTCCTGCCAGscadm gatttaacctttatgctttgaatgacatctcca TATACTGAACCTCTACAAAATGTACATTAATATTTCCAATCAAAAGTGTATATGGGGAAGGAAC ACAAGCAGATATATTACAGATTTCTTAGACCAGTCTCTAGTCCCGTCTGGTAA
347	Translation of ORF number 165 in reading frame 1 on the direct strand RCVQAHGAVLGLRAQSATQLQHHLPGAADHPKAASVRLGPLLKPVASVGKII PPPQLQLTPWDS PSQSWTLAAGIHAARGDAAPCLLCVPIPAXXRFNLYALNDISHILNSYKMYINISNQKCIWGRN TSRYINRFLRPVSSPVW
348	ORF number 166 in reading frame 1 on the direct strand extends from base 197797 to base 198024 gggtcatcaatgtcatcagcaggtagagggccacatgttgaccaatcgttctggcagccacc gaggactatctgcacccctgtggaaaaacacaaacatgccctcgcccccatatgaggactgggtc

	tgggccttgccatgatccagtaagtggatccttccttctgaccaaagcctcagggacggcagaa ggatgccagaaacgttcagctgcactcttaacatag
349	Translation of ORF number 166 in reading frame 1 on the direct strand GSSMSSAGRGSTCCTNRSGSHRGLSASCCKTQTCPRPHMRTGSGPCHDPVSGSFLLTKASGTAE GCQKRSALLT

CLAIMS

What is claimed is:

1. An induced pluripotent bat stem cell (bat IPSC), wherein the cell is in a pluripotent state.
2. The bat IPSC of claim 1, wherein the cell is in a pluripotent state characterized by the expression of one or more factors selected from the group of Klf4, Klf17, Essrb, Tfcp2l1, Tfe3, Dppa, Oct4, Sox2, Nanog, and Dusp6.
3. The bat IPSC of claim 1 or 2, wherein the cell is in a naïve pluripotent state.
4. The bat IPSC of any one of claims 1-3, wherein the cell further is characterized by the expression of one or more factors selected from the group of Otx2 or Zic2.
5. The bat IPSC of any one of claims 1-4, wherein the cell is derived from a bat fibroblast.
6. The bat IPSC of claim 5, wherein the cell is derived from a bat embryonic fibroblast or a bat fibroblast from an adult bat.
7. The bat IPSC of any one of claims 1-6, wherein the cell is derived from a *Rhinolophus* bat or a *Myotis* bat.
8. The bat IPSC of claim 7, wherein the cell is derived from a *Rhinolophus ferrumequinum* bat or a *Myotis myotis* bat.
9. The bat IPSC of any one of claims 1-8, wherein the cell is capable of differentiating into embryonic bodies.
10. The bat IPSC of claim 9, wherein the embryonic bodies are capable of differentiating into three-dimensional structures comprising three germ layer markers.
11. A method of producing induced pluripotent bat stem cells (bat IPSCs), the method comprising:
 - (i) reprogramming isolated bat cells with Oct4, Sox2, cMyc, and Klf4 factors;

- (ii) culturing the reprogrammed cells on feeder cells in a medium comprising FGF, Leukemia inhibitory factor (Lif), SCF, and Forskolin until colonies appear; and
 - (iii) splitting cells using a low concentration EDTA buffer;
- thereby producing iPSCs from bats.
12. The iPSCs produced by the method of claim 11.
13. The method of claim 11 or claim 12, wherein the isolated bat cell is a bat fibroblast.
14. The method of claim 13, wherein the isolated bat cell is a bat embryonic fibroblast or an bat adult fibroblast.
15. The method of any one of claims 11-14, wherein the isolated bat cell is derived from a *Rhinolophus* bat.
16. The method of claim 15, wherein the isolated bat cell is derived from a *Rhinolophus ferrumequinum* bat.
17. The method of any one of claims 11-16, wherein the Lif is at a concentration of 10^4 U/ml.
18. The method of any one of claims 11 -17, wherein the FGF is at a concentration of 100 ng/ml.
19. The method of any one of claims 11,-18 wherein the SCF is at a concentration of 100 ng/ml.
20. The method of any one of claims 11-19, wherein the Forskolin is at a concentration of 20 nM.
21. The method of any one of claims 11-20, wherein the feeder cell is a mouse CF1 mouse embryonic fibroblasts (MEF).
22. The method of any one of claims 11-21, the method further comprising passaging the bat iPSCs every 5 days onto feeder cells.

23. The method of any one of claims 11-22, wherein the bat iPSC is further differentiated into embryonic bodies.
24. The method of claim 23 wherein the embryonic bodies are further differentiated into three-dimensional structures comprising three germ layer markers.
25. A method of producing induced pluripotent bat stem cells (bat iPSCs), the method comprising:
- (i) reprogramming isolated bat cells with Oct4, Sox2, cMyc, and Klf4 factors;
 - (ii) culturing the reprogrammed cells in feeder free medium comprising FGF, Leukemia inhibitory factor (Lif), SCF, and Forskolin until colonies appear; and
 - (iii) splitting cells using a low concentration EDTA buffer
- thereby producing iPSCs from bats.
26. A composition for reprogramming a bat cell to produce pluripotent stem cells comprising a medium comprising FGF, Leukemia inhibitory factor (Lif), SCF, and Forskolin.
27. The composition of claim 18, wherein the Lif is at a concentration of 10^4 U/ml.
28. The composition of claim 18, wherein the FGF is at a concentration of 100 ng/ml.
29. The composition of claim 18, wherein the SCF is at a concentration of 100 ng/ml.
30. The composition of claim 18, wherein the Forskolin is at a concentration of 20 nM.
31. A method of obtaining viral sequences from bat iPSCs, the method comprising
- obtaining bat iPSCs;
 - identifying viral sequences residing in the bat iPSC genome or intracellular virus genome; and
 - assembling the viral sequences;
- thereby obtaining viral sequences from the bat iPSCs.

32. The method of claim 31, wherein the identifying comprises sequencing the bat genome or the genome of viral particles residing in the bat iPSCs, or of viral particles shed by the bat iPSCs.
33. The method of claims 31 or claim 32, wherein the identifying comprises sequencing the RNA of the bat genome or the genome of viral particles residing in the bat iPSCs, or of viral particles shed by the bat iPSCs.
34. The method of claim 31, wherein the identifying the proteins and peptides produced by the viral genome by proteomics e.g., LC-MS.
35. The method of claim 31, further comprising translating the sequence into a protein sequence and determining whether the translated sequence has a significant homology to a known protein sequence in a viral protein database.
36. The method of claim 35, wherein the sequence is selected from SEQ ID NO: 1-349.
37. The method of claim 31, wherein the virus is selected from the group of a SARS-CoV-2 virus, endogenous retrovirus (RfRV), and sindbis virus.
38. The method of claim 31, wherein the virus is a coronavirus.
39. The method of claim 35, wherein the sequence is encoding a gag protein, a pol protein, or an env Protein.
40. A method of obtaining viral sequences from virus particles shed by bat iPSCs or cells derived from bat iPSCs, the method comprising
 - obtaining bat iPSCs or cells derived from bat iPSCs;
 - culturing the bat iPSCs or cells derived from bat iPSCs under conditions that allows shedding of virus particles into the culture media;
 - collecting the culture media;
 - identifying viral sequences residing in the culture media; and
 - assembling the viral sequences,

thereby obtaining viral sequences from virus particles shed by bat iPSCs or cells derived from bat iPSCs.

41. Use of any one of the viral sequences of claims 31- 40 for the development of a vaccine.
42. A recombinant nucleic acid molecule, comprising

a promoter, and

a nucleic acid selected from SEQ ID NO: 1-349 encoding for a viral protein or fragment thereof.
43. A recombinant, replication deficient adenovirus, comprising the nucleic acid of claim 42.
44. A mRNA comprising the nucleic acid of claim 42.
45. An expression vector comprising

a promoter and

a nucleic acid selected from SEQ ID NO: 1-349 encoding for a viral protein or fragment thereof.
46. An isolated protein or peptide comprising an amino acid sequence encoded in a nucleic acid set forth in SEQ ID NO: 1-349, wherein the peptide is no more than 100 amino acids in length, and an optional pharmaceutically acceptable carrier.
47. The isolated protein or peptide of claim 46, wherein the protein or peptide is no more than 30 amino acids in length or 20 amino acids in length.
48. The isolated protein or peptide of claims 46 or 47, where the protein or peptide is synthetic.
49. A pharmaceutical composition comprising the adenovirus of claim 43, the mRNA of claim 44, or the protein or peptide of any one of claims 46-48 and a pharmaceutically acceptable carrier or excipient.

50. A pharmaceutical composition comprising a plurality of (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) proteins or peptides of any one of claims 46-48 and a pharmaceutically acceptable carrier or excipient.
51. A pharmaceutical composition comprising a nucleic acid encoding the mRNA of claim 44 or the protein or peptide of any one of claims 46-48 and a pharmaceutically acceptable carrier or excipient.
52. A pharmaceutical composition comprising one or more nucleic acids encoding a plurality of (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) mRNAs of claim 44 or proteins or peptides of any one of claims 46-48, and a pharmaceutically acceptable carrier or excipient.
53. The pharmaceutical composition of any one of claims 49-52, further comprising a liposome, wherein the protein or peptide or the nucleic acid encoding the protein or peptide is disposed within the liposome.
54. The pharmaceutical composition of any one of claims 49-52, further comprising a lipid nanoparticle, wherein the protein or peptide or the nucleic acid encoding the protein or peptide is disposed within the lipid nanoparticle.
55. The pharmaceutical composition of any one of claims 49-54, further comprising an immunogenicity enhancing adjuvant.
56. The pharmaceutical composition of any one of claims 49-55, wherein the protein or peptide or nucleic acid encoding the protein or peptide is synthetic.
57. A vaccine that stimulates a T cell mediated immune response when administered to a subject, the vaccine comprising the pharmaceutical composition of any one of claims 49-56.
58. A vaccine comprising the pharmaceutical composition of any one of claims 49-57.
59. The vaccine of claims 57 or 58, wherein the vaccine is a priming vaccine and/or a booster vaccine.
60. A recombinant cell comprising a nucleic acid or a portion of a nucleic acid set forth in SEQ ID NO: 1-349.
61. A recombinant cell comprising a protein or a portion of a protein encoded by a nucleic acid set forth in SEQ ID NO: 1-349.

62. A composition comprising an inhibitor of a protein encoded by a nucleic acid selected from SEQ ID NO: 1-349.