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 prat 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. 11A discloses SEQ ID NOS 8, 356, 360, 9 and 361, respectively, in order of appearance. **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. 11B discloses SEQ ID NOS 15, 357-359, 362, 14, 358 and 363, respectively, in order of appearance. **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. 11C discloses SEQ ID NOS 364 and 365, respectively, in order of appearance. **FIG. 11D** shows a phylogenic 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

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 (SEQ ID NO: 350) and Z25-133:GAPDH_R1_GHB: GAAGGGGTCATTGATGGCGA (SEQ ID NO: 351)). 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 Cytofix/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

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:chr13:14,355,027-14,363,924	Cui J, et al., J Virol. 2012 Apr;86(8):4288-93.	9649 bp	Transcription profile in RNA-seq in genomic region that overlaps with the previously identified endogenous retrovirus	Previously identified endogenous 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 (SEQ ID NO: 352)
Sequence 2	m64019_210618_193151_159712964_ccs (SEQ ID NO: 353)
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 CTACTGCAGTATTTCTCAGCTAGAGTTGTGCTGGCGACTCACAGTCACTT     50

NC_009657.1    21554 -TGTCTTTTGACTCACCACTTAATGTGCCTGGGTT--TTCCTGTAACGGC  21600
                      .|....||.||..|||  ||.|...||||||..|  |.||||.|.....
m64019_210618     51 GAGGAACTTTACAAACC--TTTACAGGCCTGGACTCCTCCCTGAAGGTTT     98

NC_009657.1    21601 GCCAATGGTTCTAGCTCAGCGGAAGCCTT-TCGTTTTAACGTCAATGATA  21649
                     ...|.||..||.|.||.|...|  ||||. .|.||.|....|.|...|||
m64019_210618     99 TTGACTGAGTCAATCTAATAAG--GCCTGGACATTGTGTATTTAGAAATA    146

NC_009657.1    21650 CTAAGTTGT-TTGTTGGTGCTGGCGCTGTTACATT-GAACACCGTCGATG  21697
                     .|.....|. ||..||.|.|...|..||||.||.. |.||....|..||.
m64019_210618    147 GTTCCCCGAGTTTCTGATACAACCCTTGTTTCAAAAGTACTGAATGTATA    196

NC_009657.1    21698 GTGTTAATGTTTCTATTGTGTGCTCCAATAATGCAACACAGCCCACTAGG  21747
                     ....||||||.||..||....|.|...| ||.||...|.|..||.||||.
m64019_210618    197 AGTGTAATGTATCACTTCACAGATTTCA-AAAGCGTAAGAAACCTCTAGA    245

NC_009657.1    21748 TCAA--ACAACTTGCAGGAAGACCTGCCTTACTATTGCTTCACTAACACT  21795
                     ..|.  |||..|.|...|..|..||..|||..|.|..||.|..|...||.
m64019_210618    246 AAAGGTACAGTTAGTGAGGGGTACTTACTTCATCTCTCTCCTTTGCAACA    295

NC_009657.1    21796 AGTAGCGGCACTAATCACACTGTTAAGTTTCTTTCAGTTTTCCCGCCAAT  21845
                     .|....|||     |.||. ||....||.|.|..||||...|.||..||.
m64019_210618    296 CGCTTAGGC-----TGACT-TGAACTGTCTTTACCAGTGGGCTCGAGAAG    339

NC_009657.1    21846 CATTCGTGAGTTTGTGATCACCAAATATGGCAATGTCTATGTTAATGGCT  21895
                     .....|.||..|..|...|...|..|...|...|....|.||..|.  ||
m64019_210618    340 TGAAAGGGATATCATTGGCGGTATCTGCTGGCCTACAGAAGTACAC--CT    387

NC_009657.1    21896 ATATCTATTTGAGAACTAGACCATTGACAGCCGTGCACTTGAACGCATCC  21945
                     .|.||..|||        ..||||   |||.|.|.||||.|..|.||...
m64019_210618    388 GTTTCCTTTT--------TGCCAT---CAGTCATTCACTGGCGCTCAGAG    426

NC_009657.1    21946 TCTCATTCGCAGGACGTAGCAGGGTTTTGGACTATTGCCGCCACAAACTT  21995
                     .|||.||.|......|..|..||||..|..||||.|.. |....|..|..
m64019_210618    427 ACTCTTTAGACATTTGCTGATGGGTAGTCTACTAATAA-GTAGAATCCGA    475
```

Alignment of identified sequence with the Scotophilus bat coronavirus 512 genomic sequence.

```
m64019_210618   5863 GGTTGAAAGAGTGAGGGTATCTAACAGGGAGGATGCTGCTGGACAGATTT   5912

NC_009657.1    27531 ACGAGGTTTTTGACACCCAGAATTGAATTCAACTAAAACAATGTACAGAA  27580
                     .....||    |.|||.|.......||..||.|  |.||...||.||.||
m64019_210618   5913 CTCCAGT----GTCACACTACCGCCAAGACAGC--ACACGGAGTTCACAA   5956

NC_009657.1    27581 TTGTAGCTATTGTTTTGGCTGAGCTTTTTCGAGCACTGGCCATTTTTGGC  27630
                     ..| |...|.||.|.|||.||.|..|......|||.|...|.|||||
m64019_210618   5957 AAG-ACAAAATGGTATGGTTGTGAGTAGGAATGCATTTTTCTTTTTT---   6002

NC_009657.1    27631 TCATTCTTCCAAATTTTTTTGCTATATTTTGATTGCATTTCCAAGGTGAG  27680
                      ..|||||||...|||||||..|.|.|||.|..|||||||....|.|.|.
m64019_210618   6003 -TTTTCTTCCTTTTTTTTTTTTTTTTTTGGTATGCATTTTTCTGATTAA   6051

NC_009657.1    27681 TTTAAGCTGTCCTACAGGACGTTGGTGTTTGCTTACATGTGCTGATTTCC  27730
                     |.|.....|...|.|.....||....|..||...|||..||  ||.|||.
m64019_210618   6052 TCTTCTTGGAGTTGCTCATTGTCACAGCATGAAAACACCTG--GAATTCT   6099

NC_009657.1    27731 TTATTCTTGTGC-TCATATTCTTTCTTTTCTTGGTGCCTTTTTCTTACTG  27779
                     |..|||.||... |..|..|...|||||..|..|. ..|||.|.||.||.
m64019_210618   6100 TGTTTCCTGACAGTGCTTATAGATCTTTGATCAGC-TATTTATGTTGCTC   6148

NC_009657.1    27780 TTTAGTGGTGTACATCGTTAA-AGATGATTGGGCCCCCTGGATGTGGTAT  27828
                     ..|.......|.|||.|..|| ||..|||...||...||.|.|.|..|.|
m64019_210618   6149 AATGTCCACTTCCATAGAAAACAGTAGATGCAGCAGTCTAGTTCTCATTT   6198

NC_009657.1    27829 GTTAACCTCTACAGGCCCCTACATGATGCCTTAATCAGATTTCTTATG-A  27877
                     |.|...|||..||.........|...||..|..|||.|||....|.|. |
m64019_210618   6199 GCTCCACTCATCAAATTAACCAAGTCTGTATCTATCTGATGATGTGTATA   6248

NC_009657.1    27878 CACCAGACTTTGCTGTCTTGGTTTTATCTTTCTTGTTCATGATCTTAACA  27927
                     .|...|..|.||.||..||.|.||.|.|||..|.........|.|...||.
m64019_210618   6249 TATGTGTGTGTGGTGCATTAGATTCAGCTTGTTGACAATAAAACAATACT   6298

NC_009657.1    27928 TG-GCTGCTGGGCATTGGAATCTTCCAATACTAGCGGT-CTTGGTCTTGC  27975
                     |. ..||...||.||....|.||.|...||...|.|.| |.|..|....|
m64019_210618   6299 TTTATTGACTGGGATACTGACCTACTTGTATATGTGCTGCCTCTTTAAAC   6348

NC_009657.1    27976 ACACAACGGTAAGCCTGTAATAATGACAGTGCAAGCAGGTTATTATTATA  28025
                     ...|.||..|....|..||..|..|.......||..|.....|.......
m64019_210618   6349 CTTCCACTTTGTTTCAATAGAATAGTATAAAAAACAAAAAGCTCTAGGAT   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 (SEQ ID NO: 354)
Sequence 2	hub_1489433_GCA_004115265.2_dna (SEQ ID NO: 355)
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 TTGTTCACTATGTATTACATATTGAATTTTCACAATAATGTTAAGAGGCA    183

21560-25369       58 ACTAGAACTCAGTTACCTCCTGCATACACCA---ACTCATCCACCCGTGG    104
```

	CGGCGTCCACTGTGGACGTCGCTGGGCCACCTTCTCGTCGGGGACGTCTCCCACGACTTCAAGTAACGAA GCCTCCAGGGTCGTAACCTGCACTTTCCTTCCCGTGAGAATTCTCAACTGGGACCGGGCAGAGACCTGTT TCAGTCTCTCCGGTCGAGGACCTGTTTCAGTCCTCCCCTGTTTGGGACCGGGCAGAGACCTGTTTCAGTC TCTCCGGTCGAGGACCTGTTTCAGTCCTCCCCTGTTTGGGACCGGGCAGAGACCTGTTTCAGTCTCTCCG GTCGAGGACCTGTTTCAGTCCTCCCCTATTGGAGGTGGCCAAACCTCCTTCCGCGGTTCCCTATGTAAAC CTCGGTATCGGGAGTTGTCTGTTCCCCTGAGGGGGGGCGTCCCGGGCGAGCCCCCAAATGTTAAGATCCG GAATTGTCCCCCTCAGACAGACAGAGAACCCGCACAATAATGCAAGTCACACAGCGAGGTTTATTACCAG CCGGCTGGGGTCCCCGTCTCTGCCCGACGCAGCGGGTTTTTAACAAGGACCCCAAACACTTAAAGCCAAG GGGTTATATAGTATTTTCAAGGCCTTCCATAGCTTCTGAGAGTACAAGATAAACTTACAGGACTTACATA GTTTCAAAGAGTATAAGATTTACTACAGGACAAGGAGACCAGGAGTATAAGATAAGCTACAGGACTTCTA GTCGCCATGCTGCAACTGCCCACATCCTGGAATTTTATAATTATGTTGTTTCAGGCTAGGGGCTGTTAAC CCTGACCTGAAGATAGCAGTTCTCATGCTAACAGTCTCTTACATGCTAACAGTCTCTTACAGTAAGAAGT TCCAAAGCCTGTGGTGGCAGTAAGTGAATTTCTTCCTTTTCAATAGACTATGAAGGAGGGACATTGCATT TGAACTCAGTCCATGAGTCATGATGCTCTTTATGTCCATTAAAAGGATTAACTTTCTCTCTATTCACTAT TTCTTTCACACTATTGTATAGGGTAACGTGTTTGGGGAGAAAAATCAATAAAAATGCTTAAAATAAAAGT TTCCATGCTCATAAGGTTTTTATCTTCCATTATAGGAAAATGAATCTATATGGAAGGGTACATTTTCTGA TGATGTTTTGTAAGAAGCATTATTCTATCAATCTATTAAAATATATTGATGCACTTTCC
7	Part of RFe-MD-2 sequence with Columbid/Falconid DNA homology TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGG AGGCAGv
8 and 356	Protein sequence ofRFe-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 TATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAGGCA
10	Part of RFe-V-MD3 sequence with Human herpesvirus 4 isolate HKNPC6 homology TTCATCCATGTCACAAATGACAAGATTTTGTTTTTTATAGCTGAGTAATATTCCATTGTATACATATACC ACATCTTCTTTGTGTATTCGTCTGTCAGTGAACTTTGGTTACTTCCATATCTTGGCTGTTGTAAATAATG CTGCAGTGAACATAGGGGTGTGTATATCTTTTCGAATTAGTATTTTGGATTTTTTTCAGATAAATACCCA GAAGTGGAATTGCTGGGTCATATGGTAATTCTATTTTTAATTTTTTGAGGGACCTCCATACTGTTTTCCG TAGTGGCTGCACCAATTTACAAGGTGCTTTTCTCTACATCCTTGCCAACACTTGTTGTTTATTGATTTAT TGATGATGGCCATTCTGACACGTGTGACATGATAGCTCATTGTGGTTTTAATTTGCATGTCCCTGATGAT TAGTGACATTGAGTATTTTTTCATATGTCTATTGGCCATCTCTGTGTCCTCTGGAGAAATGTCTGTTCAG GTCCTCTGCCCATTTTTTAAATCAGATTGTTTCGTTTTGTGTGTTAAGTTGTATGAGTTCCTTATATATT TTGGATATTAAACCCTTATTGGCATCTTCTCCCATTCAGCAGGTTATCATTTTGTTTTGCTAATGGCATC CTTCACTGTGCAAAAACTGTTTAGTTTGATGTAGTCCCATTTGTTTATTTTTTTTCTTTTGTTTCCCCTG CCAGAGGAAACATATTCAAAGAAATACTACTAAAAGAGATGTAAAAGCGTTTACTGCCTATATTTTCTTC TAGGAGTTTTACGGTTTTGGGTCTTAAATTTAACTCCTTAATCCATTTTTAGTTTATTCTTATATGTATA CAGTGATCCAGTTTCATTCTTTTGCATGTATCTGTCTATAGTTTTTCCAACACCATTTACTGAAGAGACT GTCTTTACCCAATTATATATTTTTGCCTCCTGTCATAGATTAATTGACCATGTGGGCATGGGTTTATTTC TGGGTTCTGTTCCATTGATTTATGTGTCTGTTTTTATGTCAGTACCATGATGTTTTGATTACTATGGTCT AGTAGTATAGTTTGATATCAAGTAGCATGATACCTCCAGCTTTGTTCTTCTTTATCAAGATCGCTTTAGC TATCTGGGGTCTGTTGTGGGGTCTACAAATTTTAGGGTTACTTGTTCTGGTTCTGTGAAAATGCCATTGG TATTTTGATAGGAATTGCATTGAATCTGTAGATTGATTTGGGTAGTATGAACATTTTAATGATGTTAATT CTTTCTATTCACAAACATAGTATATGCTTCCATTTATTAGTATCTTAACTTTCATTCTTCAGTGTCTTAC AGTTTTCCAAGCACAGGTCTTTTACTTCCTTAAATTCATTCCTAGGTATTTTATTCTATTTAATGCAATT TTAAATGGGATTGTTTTCTTAATCTCTCTTTCTGATAGTTTGTTATTGGTGTATAAAAATGCAACCAATT TCTGAATATTAATTTTGTGTCCTGATACTTTACTGAATTCATTTATTAGTTCTAATTGTTTTTTTGGTGG AATCTTAAGGTTCTCTCTATATAGTATCATGTCATCTGTGAATAATGACAATTTTACTTCTTCCTTTTCA ATTTGGATGGCTTTTATTTCTAGTCTGACTGCTGTGGCTGGGACTTCTAGTACTATGTTGAATAAAAGTG AAAGTGGCTTGTTCCTGATCTTAAAGGAAAAGCTTTCAGCTCTTCACTACTGAGTATGATGTTAGCTGTG GGTTTGTCCTATATGGCCTTTATTATGTTGAGGTATTTTCCCTCTATTCCCAATTTGCTGAGAGTTTTTA TCATAAATAGATGTTGGATTTTGTCAAATGCTTTTTCTGCATCTATTGATATGATCATATGATTTTTATC TTTCATTTTGTTTATATAGTTTATCACATTAATTGATTTGCAAATATTGAACCAACCTTGCATGCCAGGA ATAAATCCCACTTAATCATGGTGTATGAACTTTTTAATGTACTGCTGAATTTGGTTTGCTAATACTTTGT TGAGGATTTTTGCATCTATGTTGTTCATCAGGGATGTTGGGCATTTTTTTTTTTTTTTTGTATTGTCTCT GGTTTTGGTATCAGGCTAATGCTGGCCTTGTAAATGAGTT
11	Part of RFe-V-MD3 sequence with Human herpesvirus 4 isolate HKD40homologyTAAATACCCAGAAGTGGAATTGCTGGGTCATATGGTAATTCTATTTTTAATTTTTTG

	AGGGACCTCCATACTGTTTTCCGTAGTGGCTGCACCAATTTACAAGGTGCTTTTCTCTACATCCTTGCCA ACACTTGTTGTTTATTGATTTATTGATGATGGCCATTCTGACACGTGTGACATGATAGCTCATTGTGGTT TTAATTTGCATGTCCCTGATGATTAGTGACATTGAGTATTTTTTCATATGTCTATTGGCCATCTCTGTGT CCTCTGGAGAAATGTCTGTTCAGGTCCTCTGCCCATTTTTTAAATCAGATTGTTTCGTTTTGTGTGTTAA GTTGTATGAGTTCCTTATATATTTTGGATATTAAACCCTTATTGGCATCTTCTCCCATTCAGCAGGTTAT CATTTTGTTTTGCTAATGGCATCCTTCACTGTGCAAAAACTGTTTAGTTTGATGTAGTCCCATTTGTTTA TTTTTTTTCTTTTGTTTCCCCTGCCAGAGGAAACATATTCAAAGAAATACTACTAAAAGAGATGTAAAAG CGTTTACTGCCTATATTTTCTTCTAGGAGTTTTACGGTTTTGGGTCTTAAATTTAACTCCTTAATCCATT TTTAGTTTATTCTTATATGTATACAGTGATCCAGTTTCATTCTTTTGCATGTATCTGTCTATAGTTTTTC CAACACCATTTACTGAAGAGACTGTCTTTACCCAATTATATATTTTTGCCTCCTGTCATAGATTAATTGA CCATGTGGGCATGGGTTTATTTCTGGGTTCTGTTCCATTGATTTATGTGTCTGTTTTTATGTCAGTACCA TGATGTTTTGATTACTATGGTCTAGTAGTATAGTTTGATATCAAGTAGCATGATACCTCCAGCTTTGTTC TTCTTTATCAAGATCGCTTTAGCTATCTGGGGTCTGTTGTGGGGTCTACAAATTTTAGGGTTACTTGTTC TGGTTCTGTGAAAATGCCATTGGTATTTTGATAGGAATTGCATTGAATCTGTAGATTGATTTGGGTAGTA TGAACATTTTAATGATGTTAATTCTTTCTATTCACAAACATAGTATATGCTTCCATTTATTAGTATCTTA ACTTTCATTCTTCAGTGTCTTACAGTTTTCCAAGCACAGGTCTTTTACTTCCTTAAATTCATTCCTAGGT ATTTTATTCTATTTAATGCAATTTTAAATGGGATTGTTTTCTTAATCTCTCTTTCTGATAGTTTGTTATT GGTGTATAAAAATGCAACCAATTTCTGAATATTAATTTTGTGTCCTGATACTTTACTGAATTCATTTATT AGTTCTAATTGTTTTTTTGGTGGAATCTTAAGGTTCTCTCTATATAGTATCATGTCATCTGTGAATAATG ACAATTTTACTTCTTCCTTTTCAATTTGGATGGCTTTTATTTCTAGTCTGACTGCTGTGGCTGGGACTTC TAGTACTATGTTGAATAAAAGTGAAAGTGGCTTGTTCCTGATCTTAAAGGAAAAGCTTTCAGCTCTTCAC TACTGAGTATGATGTTAGCTGTGGGTTTGTCCTATATGGCCTTTATTATGTTGAGGTATTTTCCCTCTAT TCCCAATTTGCTGAGAGTTTTTATCATAAATAGATGTTGGATTTTGTCAAATGCTTTTTCTGCATCTATT GATATGATCATATGATTTTTATCTTTCATTTTGTTTATATAGTTTATCACATTAATTGATTTGCAAATAT TGAACCAACCTTGCATGCCAGGAATAAATCCCACTTAATCATGGTGTATGAACTTTTTAATGTACTGCTG AATTTGGTTTGCTAATACTTTGTTGAGGATTTTTGCATCTATGTTGTTCATCAGGGATGTTGGGCATTTT TTTTTTTTTTTTGTATTGTCTCTGGTTTTGGTATCAGGCTAATGCTGGCCTTGTAAATGAGTTTGAGAGC CTTCCCTCCTTTTCAGTTTTTTGGAATGTTTGGTAAAATTTACCTGTGAAGTCATTTGGTTCAGGGCTTT TGTTTGTTGGGAGTTTTTTGATTACTGATTCGATTTTGTTAGCAGTTACTGGTCTGTTCAGATTTTCTGT TACTGATTCAGCCTTAATTTTCTGCTGATTCAAGCCTTGGAAGATTGTATGTGTCTAGCGATTTATCCAT CTCTTCCAGTTTGTCCAATTTGTCAGCATATAGTTGTTCTAGTGTTTCCTTATACTTCTTTGTATACCTG TGGTGTCAGTTGTCGTATCTCTTTCATTTCTGATTTTATTTTGGCCCTCTCTCTTTTCTTCTTGAGTCTG GCTAAAGGTTTATCAAT
12	Part of RFe-V-MD3 sequence with Human respiratory syncytial virus (Kilifi isolate) homologyTTTTCTTCTAGGAGTTTTACGGTTTTGGGTCTTAAATTTAACTCCTTAATCCATTTTTAGTT TATTCTTATATGTATACAGTGATCCAGTTTCATTCTTTTGCATGTATCTGTCTATAGTTTTTCCAACACC ATTTACTGAAGAGACTGTCTTTACCCAATTATATATTTTTGCCTCCTGTCATAGATTAATTGACCATGTG GGCATGGGTTTATTTCTGGGTTCTGTTCCATTGATTTATGTGTCTGTTTTTATGTCAGTACCATGATGTT TTGATTACTATGGTCTAGTAGTATAGTTTGATATCAAGTAGCATGATACCTCCAGCTTTGTTCTTCTTTA TCAAGATCGCTTTAGCTATCTGGGGTCTGTTGTGGGGTCTACAAATTTTAGGGTTACTTGTTCTGGTTCT GTGAAAATGCCATTGGTATTTTGATAGGAATTGCATTGAATCTGTAGATTGATTTGGGTAGTATGAACAT TTTAATGATGTTAATTCTTTCTATTCACAAACATAGTATATGCTTCCATTTATTAGTATCTTAACTTTCA TTCTTCAGTGTCTTACAGTTTTCCAAGCACAGGTCTTTTACTTCCTTAAATTCATTCCTAGGTATTTTAT TCTATTTAATGCAATTTTAAATGGGATTGTTTTCTTAATCT
13	Part of RFe-V-MD3 sequence with SARS-CoV-2 homologyAGGTTCATCCATGTCACAAATGACAAGATTTTGTTTTTTATAGCTGAGTAATATTCCATTGT ATACATATACCACATCTTCTTTGTGTATTCGTCTGTCAGTGAACTTTGGTTACTTCCATATCTTGGCTGT TGTAAATAATGCTGCAGTGAACATAGGGGTGTGTATATCTTTTCGAATTAGTATTTTGGATTTTTTTCAG ATAAATACCCAGAAGTGGAATTGCTGGGTCATATGGTAATTCTATTTTTAATTTTTTGAGGGACCTCCAT ACTGTTTTCCGTAGTGGCTGCACCAATTTACAAGGTGCTTTTCTCTACATCCTTGCCAACACTTGTTGTT TATTGATTTATTGATGATGGCCATTCTGACACGTGTGACATGATAGCTCATTGTGGTTTTAATTTGCATG TCCCTGATGATTAGTGACATTGAGTATTTTTTCATATGTCTATTGGCCATCTCTGTGTCCTCTGGAGAAA TGTCTGTTCAGGTCCTCTGCCCATTTTTTAAATCAGATTGTTTCGTTTTGTGTGTTAAGTTGTATGAGTT CCTTATATATT
14 and 358	Part of RFe-V-MD3 with RNA-dependent DNA polymerase of Erythrocytic necrosis virus homology PTSLMNNIDAKILNKVLANQIQQYIKKFIHHD*VGFIPGMQGWFNICKSINVINYINKMKDKNHMIISID AEKAFDKIQHLFMIKTLSKLGIEGKYLNIIKAI
15	Part of RFe-V-MD3 with RNA-dependent DNA polymerase of Lymphocystis disease virus homology

and 357 -359	MTSQVNFTKHSKKLKRREGSQTHLQGQH*PDTKTRDNTkkkkKC–PTSLMNNIDAKILNK VLANQIQQYIKKFIHHD*VGFIPGMQGWFNICKSINVINYINKMKDKNHMIISIDAEKAF DKIQHLFMIKTLSKLGIEGKYLNIIKAI*DKPTANIILSSEELKAFPLRSG
16	Prediction of a potential new spike protein sequence (RFe–SP2) (M)FVVVVVVVFPFRLPHHSGVSYCFSVLCRTQLPGPCWYYEPCAPPSGSRLLVGPLGHSQHFMSVLAGC RSLTLATSRSWQHMVAHGSPWQPSSRESRCSQSLRMQRTGPRGNRTSMALQPPEPPCEGCEGWLTKVFNF DEIQLFSFVACALMLYLRRHCLIQGHGDLHLCFLQVLLHLFVYLSISSFLYMVGRVSKFILLHVDIQLSH HLLKRLFSSIELSWHIYQKSIDHGFILDSSIPLIYMSVFMPVPHSLDYCSFAVSFIRKYESSHFVPFFQD CFGYSGFLAFPCKITQLVNFCRKIKVAKIFVGFAVNNSNGVLLFQNVFSTKAGFKFYLGLFLSTMFCCFP RTSVTLLCIYSLISFCYHKWCCFLISFSDCSLLVYRNTTFFFPYPCILKSCIHLLALIVLLGWGLGVDSL AFSKGHVIKIVLVLLHFQSFFLFHFLTNLARTSGRMLNSSGESRHPCLVTDLRKKASSLSPLSMVLAVGF SMLFIRLRKFPPTFASVFFSFPSNHMIPIWHTGKQMTNVEMCSRYIKLEMRPRYPDSHSTVLSTILYYFL WSPAATFRDQKKVHPFHLLIKKVSSITVGFVSGFVPLFPWNLKVPGTEVLVVSRKSQFRQIPLEPLLCAR QCAQYSSPQRDCGGGDETSYKKIIRRDLIVGNRRQLPEAEPFVNKKVFVEETGMNRALKEGQLTCVCGST LGRIFDRKLKNVLLFNFYMKSLKPITITRDMQSKNNNRVHIRSGNILFFSDLFFGLRLKLSKSAFSFCPE NLGSICFIPPIENFFRQPVSDVQVLFFVYCSMLLLQVFSVLRARLLILHTDHMWLKTTGSHHEQVRAAIW ELTIHHTFIDYPLARYMNDRGSIQADMQISYYNLIDNPREVFFCHSLDVFMLHPIETCFRGIIIVPSTDI FEHLILRSKSRANIDVELCSMQSPSPIVLLLIISEFSVSSGFERPPEPLFHNNSTLSSLIPNSTQKIKGE RKVCSTDKNFSIRISTRCDTIQTLLLSAIQWKGRDIQYKKLHVKPCVTSFNLFGAVSWVDTLERRCVLQC AVNHPLSQCSNIASGIQQFLSTKDIMVCPHPPYPDLAPCDCWLFPKEKMTTKGKLFELIQDIKAATTVQL KTLIKEDFQNCFRKWQKQRDTCVQSKGEYFEENWCVFYCNKFFITFTVFFLSHLIQKKTSRRKLLHFVPP QLQVVTGSAEYNGHMEKQFSWKFWMFTKNVFENKVIYHHHHLLCRNKTQSEVPWKKSSGWKIPALSPLIT SCDFSHFSLNCHPHDAPSPRVFHDNSVLLDEAFWWGASESPSRARVCVCVCVSLYSSEQFYRTWRRHGQS TQRECSLIMNYVKNISPGPVIMVPYDVHSTFMNIEAHLFPMKIRKLGEKIKRTHSLTPNKYWHDGFGFSL MTLILRVLALILYSILSAQILKLTGWVRIPAAVYQGVVPWAHYVTSLPFSHLKVEVLIVPASHYCENSIC NTTMPGTSVLTSIYMPQSMAEFCQPPHLLCHWTHPTIPGVSGGGXXXXXXXXXXCX
17	Prediction of a potential new spike protein sequence (RFe–SP1) (M)YCSISQLELCWRLTVTGTLQTFTGLDSSLKVFDVNLIRPGHCVFRNSSPSFYNPCFKSTECISVMYH FTDFKSVRNLKRYSGVLTSSLSFATRLGLELSLPVGSRSERDIIGGICWPTEVHLFPFCHQSFTGAQRLF RHLLMGSLLISRIRPLKKEFCISVHAQVVRSINVYHQHAFPAAVRNHEPSFLKLCDRLSQYVCLIPTALA SGGVTSKHQEPGVRTKTERNCFNNLESAVLCRNVFDQSVKNKCKWTVVISFEKFLKWVKVMTSCLYCKLR PNFWIKRRIPKKGKILHTNIHIIHNHLETANIKSQVPYRLNVSPGYVFASMYSTLTILETHVECGCQASL KNQTWSFLITYCAVPFRSDMCSLVCYCPHLGSSLTLVTFFVSISTGGYDSCLIVYEVQLHSIHSRKSNAY LIGEYHCGHLQSTPLGKLCTDASASTLQSNFGTLFLEWSSELSSCYPCPECHQNVFLSIFPLSSGKERRH WGPGEVSREGVPIRLFVCWLKTPSQTVAGVLSCKIHELLEKRSGHFRLKFFTQPFLHFIVNLVNCLSSWY KFIMNNPSDIKKGQQHRLDPFGLMELFSSWRIGLPFCTLTFCYRIILVHPCFITSDNMANVMLPLQKVLT NLDSLLFLFTGELLRDHFCTHLPHAKIFSSDFVFLYFYLGLLCFSESANNFDGSFDEFLQFFSSVLLVIG CPDLSLSVARSLPSEKTFTPLVHCHFILGIILIDLDHVPDFTSTLARFTKKFNVCSLFFKLRHSCDATQK HTTTIMNAIQQHRHMSTRTQLDLYTKVMKPYRAVCTVLPMTQGGKFSNLILRPHILTNISRFSIQSMFYV DLLVFYLNFFIVLYRGTVCFSRAHQLQVIALQSRCGGHSSAPSPVAVFQSLVAGGAAHHPMRELNQQPCC ELTVPTEPLGHPNKTYCLFQKYRKLGEKRKNHSYSQYPKTKTQHCFTFISLKCLLFISLHYTFENQIVSL AMISPCLLEVFLYCALLNIGILQLLTLNPFSHSISICPAFSHLADKSQINIFYIHYFLIIKSIFPFPYSI IHVDVFKFKHPCLPFLLYSLQISLIASKRKSKSNCVEKTKTKNSKPFGSRIRLVGCRSSKSHSCAISVLL VPSHFSFFFLISPSECWSFVTTNDYHHLPKLHASRFPGRKELCSFCFYYFPSISTKVLTVTQIRTAKVTC PTLNQIRPERCNELLCLAVSHHRRAHGYLPRAESGGKRDRTSNETQSCCQNKANGCQELTNNTPSFIEWH SLIQFQNLASCETTLLPLLPSHLFVFSCLPLSLTRTLEITMDPHRYKTISPSQSFNHLYKVHFAVSNMLT YFRDDHILRGHYFACHTHIFLNHLHNLILEMCSGIMFILGCFSLLAHSVSFTLALNTHSVPHATLVSHAK QLFIHFAIMPNSVWHNQGNLFSPLTINLKAFLIPKSQINLKVLLSESQNYFTFERNLAKVHFTFISNKPI PLNFSIYNDVPLFLVNETKHNTFLKRVKKKKSLKLLGFICYNEVSSASERSSRKNNKTVDITEQLIHELT TVCLNFNTLCHIDPNVPSSASLRALFSHGSESILKSSTHPSADAMGSLPASMALWLKDISVVQSPPLYPP FHWKLSSLPHKCEPEEIICQEGFLMVNAQILNRVQPFLTQASRTLCISGKSHRLKFPPPELCSCCCLSQP LSGTSWNSWNFQRTETSPIQSELWRYILICIYTDQSELIHNNQSEHMMLTDQNCVIWISHLHKNGPNGNQ GTNFLCKRPLPLCLGVHFRFFLFTNCSAESLSSFHTSYWKLLLIGRLWSQFTRGPLKLSEMIFKQEDSSR FEGRGYCPRSSLRSEVCQKREKAVMALVIFQWKQYISLVERVRVSNREDAAGQISPVSHYRQDSTRSSQK TKWYGCEECIFLFFFFLFFFFFLVCIFLINLLGVAHCHSMKTPGILVSQCLIFDQLFMLLNVHFHRKQMQ QSSSHLLHSSNPSLYLSDDVYICVCGALDSACQNNTFIDWDTDLLVYVLPLTFHFVSIEYKKQKALGFA
18	ORF number 1 in reading frame 1 on the direct strand extends from base 610 to base 837 TCTCACCTAGCAGGAAGGscadmtctcaggaccatcccatacagcagggtggaggattggtgga tcaggtacataggcccaatacgtctggtcttcttctgcattgctgagggtcatcaatgtcatca