



US 20060240530A1

(19) **United States**

(12) **Patent Application Publication**

**Baric et al.**

(10) **Pub. No.: US 2006/0240530 A1**

(43) **Pub. Date: Oct. 26, 2006**

(54) **METHODS AND COMPOSITIONS FOR INFECTIOUS CDNA OF SARS CORONAVIRUS**

**Related U.S. Application Data**

(63) Continuation-in-part of application No. PCT/US04/23548, filed on Jul. 21, 2004.

(60) Provisional application No. 60/488,942, filed on Jul. 21, 2003.

(75) Inventors: **Ralph S. Baric**, Haw River, NC (US);  
**Rhonda Roberts**, Durham, NC (US);  
**Boyd Yount**, Hillsborough, NC (US);  
**Kristopher M. Curtis**, Frederick, MD (US)

**Publication Classification**

(51) **Int. Cl.**  
*C12Q 1/68* (2006.01)  
*C12P 19/34* (2006.01)  
(52) **U.S. Cl.** ..... 435/91.1; 435/6

Correspondence Address:  
**MYERS BIGEL SIBLEY & SAJOVEC**  
PO BOX 37428  
RALEIGH, NC 27627 (US)

**ABSTRACT**

The present invention provides a cDNA of a severe acute respiratory syndrome (SARS) coronavirus, recombinant SARS coronavirus vectors, and SARS coronavirus replicon particles. Also provided are methods of making the compositions of this invention and methods of using the compositions as immunogens and/or vaccines and/or to express heterologous nucleic acids.

(73) Assignee: **University of North Carolina at Chapel Hill**

(21) Appl. No.: **11/334,877**

(22) Filed: **Jan. 19, 2006**

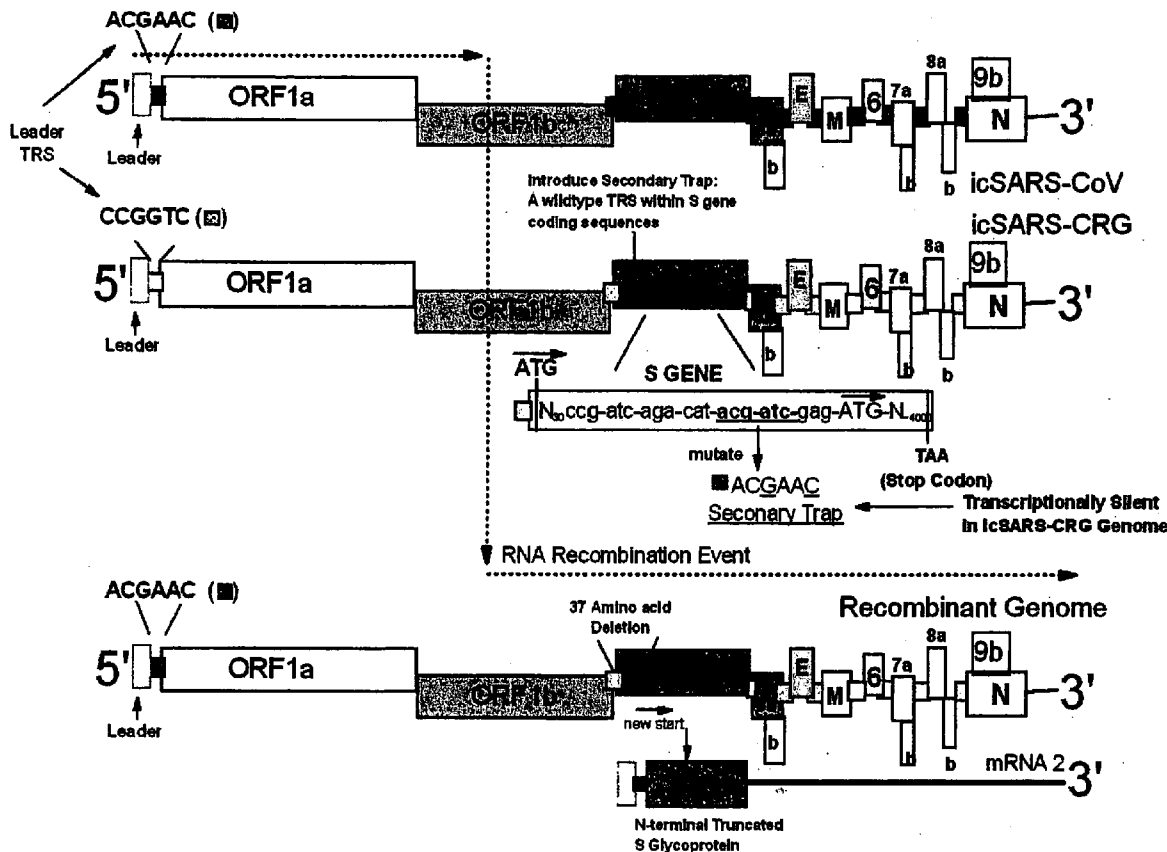
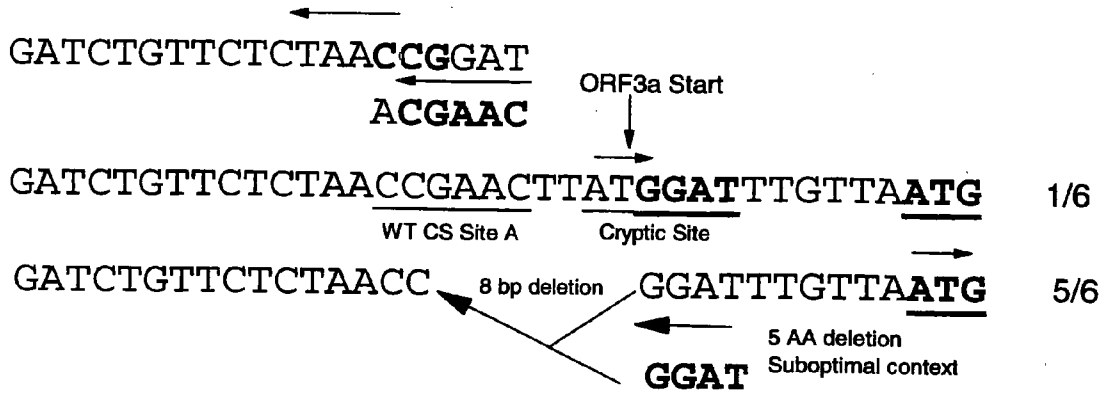


Figure 1

**A: mRNA encoding ORF3a Leader-Body Junctions**

**ORF3a Transcript Organization**



**B: leader body junctions in mRNA encoding M glycoprotein**

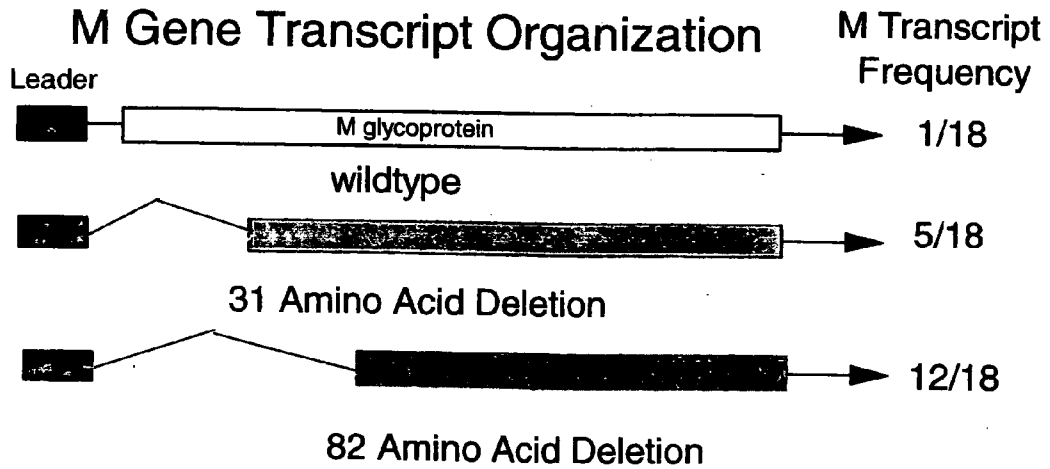
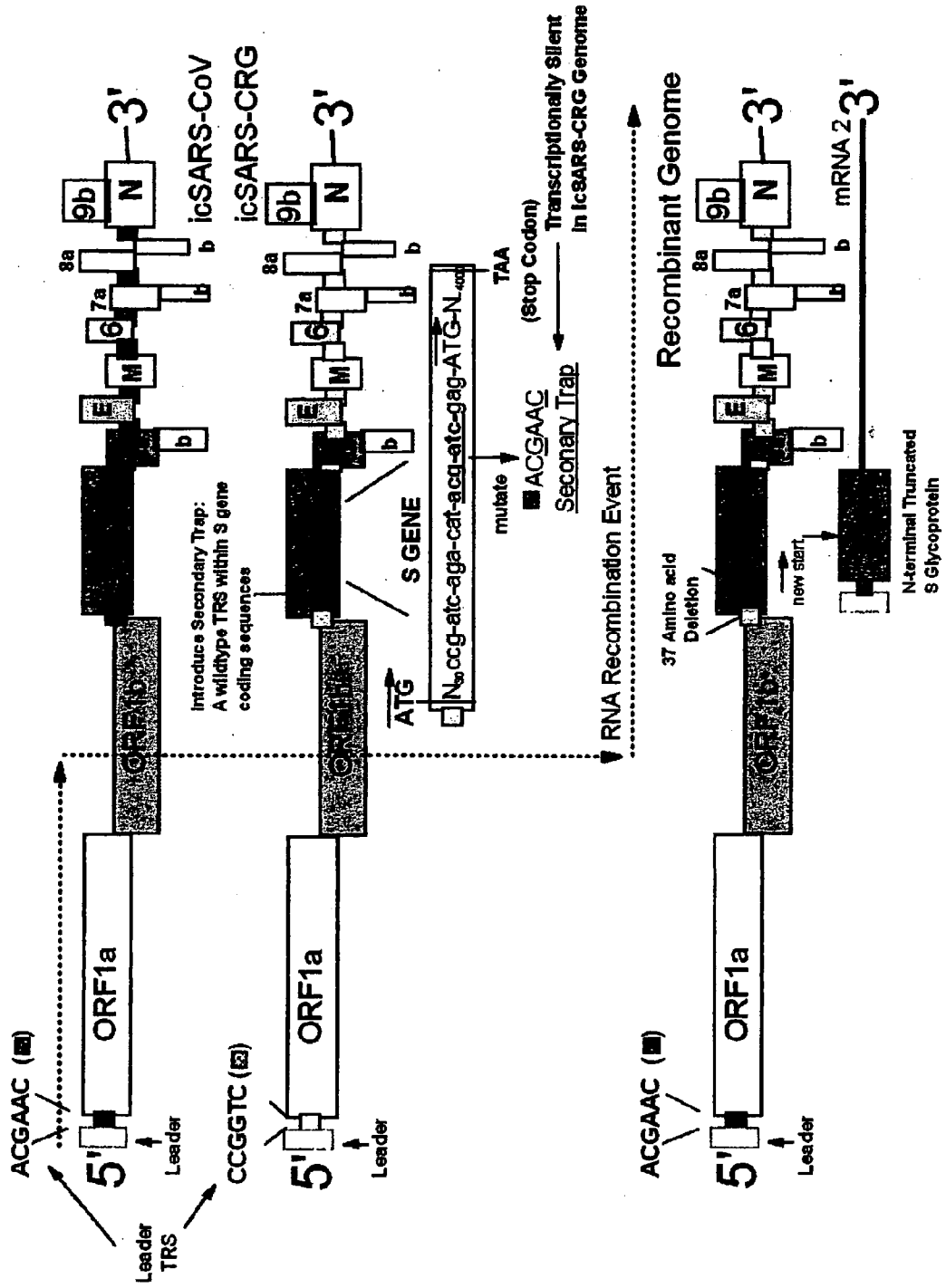


Figure 2



## METHODS AND COMPOSITIONS FOR INFECTIOUS CDNA OF SARS CORONAVIRUS

### STATEMENT OF PRIORITY

[0001] This application is a continuation-in-part application claiming priority to PCT Application Serial No. PCT/US2004/023548, filed Jul. 21, 2004, which was published in English on Apr. 21, 2005 as PCT Publication No. WO 2005/035712 and which claims the benefit of U.S. provisional application No. 60/488,942, filed Jul. 21, 2003, the entire contents of each of which are incorporated by reference herein.

### STATEMENT OF FEDERAL SUPPORT

[0002] This invention was supported by government funding under grant numbers A123946 and GM 63228 from the National Institute of Health, Allergy and Infectious Diseases. The United States Government has certain rights to this invention.

### FIELD OF THE INVENTION

[0003] The present invention relates to compositions of infectious cDNA of the severe acute respiratory syndrome (SARS) coronavirus, recombinant SARS coronavirus vectors, SARS coronavirus replicon particles, methods of making the compositions of this invention and methods of using the compositions as immunogens and/or vaccines and/or to express heterologous nucleic acids.

### BACKGROUND OF THE INVENTION

[0004] Severe acute respiratory syndrome is a life-threatening respiratory disease that probably originated in Guangdong Province, China in the fall of 2002 [1, 2]. The agent responsible for the disease spread rapidly [3, 4]. A novel coronavirus (SARS-CoV), isolated from febrile and dying patients, is the etiologic agent responsible for the disease [5-8]. SARS-CoV infection is associated with overall case fatality rates thought to approach ~14-15%, with selected populations being at increased risk (>50% in the elderly). SARS-CoV has infected over 8,000 individuals worldwide and caused over 800 deaths, before aggressive infection control measures successfully contained the scope of the outbreak. Despite intensive efforts, no effective antiviral treatments against SARS have been described.

[0005] Coronaviruses, members of the order Nidovirus, contain the largest single-stranded, positive-polarity RNA genome in nature and are divided into three main serogroups; group I: transmissible gastroenteritis virus (TGEV) and human coronavirus 229E (HCoV-229E), group II: mouse hepatitis virus (MHV) and bovine coronavirus (BoCV), and group III: infectious bronchitis virus (IBV). Sequence analyses suggest that SARS-CoV represents the prototype strain of group IV [6, 8-10]. The SARS-CoV genomic RNA is ~29,700 base pairs in length and has several large open reading frames (ORFs) encoded in subgenomic and full-length mRNAs [8-10]. The subgenomic mRNAs are arranged in the form of a nested set from the 3' proximal end, and leader RNA sequences, encoded at the 5' end of the genome, are joined to body sequences at a highly conserved consensus sequence (CS) located just upstream of each of the ORFs. The exact SARS CS sequence has been reported as either CUAAAC or AAACGAAC by different laborato-

ries [8, 9]. The SARS-CoV genome length RNA is likely packaged by a 50-kDa-nucleocapsid protein (N) [8]. As with other coronaviruses, the virion contains several viral structural proteins including the ~140 kDa spike glycoprotein (S), a 23 kDa membrane glycoprotein (M) and a ~10 kDa protein (E).

[0006] The coronavirus gene 1, or replicase gene, comprises two-thirds of the genome. MHV contains two overlapping open reading frames, ORF1a and ORF1b, which are connected by a ribosomal frameshift structure. In MHV, three proteinases, papain-like proteinases 1 and 2 (PLP-1, PLP-2) [11-13] and 3C-like proteinase (3CLpro) [14], are expressed as part of the replicase gene polyprotein and mediate cleavage of the polyproteins into at least 15 mature proteins. Continuous protein processing is crucial for ongoing virus transcription so MHV replication is sensitive to protease inhibitors that prevent replicase processing [11]. Additional functions have been predicted for proteins processed from the replicase polyprotein, including an RNA-dependent RNA polymerase (pol), an RNA helicase (hel) and a capping enzymatic activity [6, 15, 16]. The SARS virus replicase gene is similarly organized except that the SARS virus replicase has been predicted to encode only the PLP-2 equivalent and the 3CLpro proteases [6, 7].

[0007] The present invention provides a full length cDNA of the SARS coronavirus, from which transcripts are produced that replicate and/or are infectious in vitro or in vivo, multiplication-defective replicon vector particles produced from the cDNA and methods of making and using these compositions as immunogens, vaccines and/or nucleic acid delivery vectors.

### BRIEF DESCRIPTION OF THE DRAWINGS

[0008] FIGS. 1A-B. Leader body TRS junctions in wild type and chimeric recombinant viruses. Leader-containing cDNAs were isolated, subcloned and sequenced as described. The expected leader body junctions were noted for icSARS-CoV and icSARS-CoV CRG, using the body TRS CS junctions ACGAAG and CCGGAT. Leader containing transcripts were analyzed for M, and ORF3a-encoding mRNAs. Panel A: Leader-body junctions in ORF3a-encoding mRNAs. Panel B: Leader-containing junctions in ORF M-encoding mRNAs.

[0009] FIG. 2. Mechanism of secondary genetic trap. In the recombinant virus, there is miscommunication between the leader TRS CS and the body TRS CS. In this example, subgenomic transcripts of the essential S gene are directed to initiate within the S gene, resulting in N terminal truncations in the S glycoprotein product. Similar mutations can be introduced in the essential M and N structural protein genes.

### SUMMARY OF THE INVENTION

[0010] The present invention provides a cDNA of the SARS coronavirus, from which transcripts are produced that replicate and/or are infectious in vitro or in vivo. Two examples of a nucleic acid sequence encoding a cDNA of this invention are provided in the attached Sequence Listing as SEQ ID NO:1 and SEQ ID NO:3. As this invention encompasses all such cDNAs of all SARS coronavirus isolates, in one embodiment, the present invention provides an isolated nucleic acid comprising, consisting of and/or

consisting essentially of, a nucleotide sequence selected from the group consisting of: a) SEQ ID NO:1 or SEQ ID NO:3 or a fragment of at least 25 contiguous nucleotides of SEQ ID NO:1 or SEQ ID NO:3; b) a nucleotide sequence that is functionally equivalent to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3 but comprises different codons encoding the same amino acid sequences; c) a nucleotide sequence having at least 70% homology to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3; d) a nucleotide sequence having at least 95% homology to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3; e) a nucleotide sequence that hybridizes to the complement of SEQ ID NO:1 or SEQ ID NO:3 under stringent conditions; and f) a nucleotide sequence having complete complementarity to the nucleotide sequence of (a)-(e) above.

[0011] Further provided herein is a cDNA of a SARS coronavirus, wherein all or part of a nucleotide sequence of the cDNA is deleted and wherein the nucleotide sequence is selected from the group consisting of a nucleotide sequence encoding ORF1, ORF2, ORF3a, ORF3b, ORF4, ORF5, ORF6, ORF7a/b, ORF8a/b, ORF9a, ORF9b and any other ORF of a SARS coronavirus now known or later identified, and any combination thereof.

[0012] The present invention further provides a cDNA of a SARS coronavirus, comprising a mutation in a nucleotide sequence of the cDNA, selected from the group consisting of a nucleotide sequence encoding ORF1, ORF2, ORF3a, ORF3b, ORF4, ORF5, ORF6, ORF7a/b, ORF8a/b, ORF9a, ORF9b, ORF 10, ORF 13, ORF 14, any other ORF of a SARS coronavirus now known or later identified, and any combination thereof, wherein the mutation results in a nonfunctional gene product.

[0013] In an additional embodiment, the present invention provides a cDNA of a SARS coronavirus, wherein the order of nucleotide sequences of the cDNA encoding replicase, accessory ORFs and/or structural proteins S, E, M and N is rearranged in comparison to the order in wild type SARS coronavirus.

[0014] Additionally provided herein is a cDNA of a SARS coronavirus, wherein one or more of the nucleotide sequences encoding replicase, accessory ORFs and/or structural proteins S, E, M and N is present two or more times.

[0015] The present invention further provides a cDNA of a SARS coronavirus, comprising an attenuating mutation in a consensus sequence of the nucleotide sequence of the cDNA, selected from the group consisting of a leader consensus sequence, an S(ORF2) consensus sequence, an ORF3a consensus sequence, an E consensus sequence, an M consensus sequence, an ORF6 consensus sequence, an ORF7 consensus sequence, an ORF8 consensus sequence, an N consensus sequence, and any combination thereof.

[0016] In one embodiment provided herein, the present invention provides a cDNA of a SARS coronavirus comprising a 29 bp insertion in ORF8a/b.

[0017] Also provided herein is a cDNA of a SARS coronavirus, wherein all or part of a nucleotide sequence of the cDNA is deleted and wherein the nucleotide sequence is selected from the group consisting of a nucleotide sequence encoding ORF1, ORF2, ORF3a, ORF3b, ORF4, ORF5, ORF6, ORF7a/b, ORF8a/b, ORF9a, ORF9b, or any other ORF of a SARS coronavirus now known or later identified,

and any combination thereof and furthermore, wherein the order of nucleotide sequences of the cDNA encoding replication, accessory ORFs and/or structural proteins S, E, M and N is rearranged in comparison to the order in wild type SARS coronavirus. In this embodiment, the cDNA can further comprise an attenuating mutation in a consensus sequence of a nucleotide sequence of the cDNA, selected from the group consisting of a leader consensus sequence, an S consensus sequence, an ORF3a consensus sequence, an E consensus sequence, an M consensus sequence, an ORF6 consensus sequence, an ORF7 consensus sequence, an ORF8 consensus sequence, an N consensus sequence, and any combination thereof.

[0018] The present invention also provides a SARS coronavirus replicon RNA comprising a coronavirus packaging signal and a heterologous RNA sequence, wherein the replicon RNA lacks a sequence encoding at least one coronavirus structural protein.

[0019] Furthermore, the present invention provides an infectious, multiplication-defective, coronavirus particle, comprising a SARS coronavirus replicon RNA, wherein the replicon RNA comprises a coronavirus packaging signal and a heterologous RNA sequence, and wherein the replicon RNA lacks a sequence encoding at least one coronavirus structural protein.

[0020] Additionally provided is a population of infectious, multiplication defective, coronavirus particles, wherein each particle comprises a SARS coronavirus replicon RNA, and wherein the replicon RNA comprises a SARS coronavirus packaging signal and a heterologous RNA sequence, and wherein the replicon RNA lacks a sequence encoding at least one coronavirus structural protein, wherein the population contains no detectable replication-competent coronavirus particles as determined by passage on coronavirus permissive cells in culture.

[0021] The present invention further provides a method of introducing a heterologous RNA into a subject, comprising administering to the subject an effective amount of the particles or populations and/or compositions of this invention.

[0022] Also provided herein is a method of inducing an immune response and/or treating and/or preventing a SARS coronavirus infection in a subject, comprising administering to the subject an effective amount of the viruses, vectors, particles or populations and/or compositions of this invention.

[0023] In further embodiments, the present invention provides a helper cell for producing an infectious, multiplication-defective, coronavirus particle, comprising: (a) a SARS coronavirus replicon RNA comprising a coronavirus packaging signal and a heterologous RNA sequence, wherein said replicon RNA lacks a sequence encoding at least one coronavirus structural protein; and/or (b) at least one separate helper RNA encoding the at least one coronavirus structural protein absent from the replicon RNA, said helper RNA lacking a coronavirus packaging signal; wherein the combined expression of the replicon RNA and the helper RNA produces an infectious, multiplication-defective coronavirus particle. Thus, the present invention includes the embodiment of a helper cell comprising a helper RNA encoding at least one coronavirus structural protein and the

embodiment of a helper cell comprising a SARS coronavirus replicon RNA comprising a coronavirus packaging signal and a heterologous RNA sequence, wherein said replicon RNA lacks a sequence encoding at least one coronavirus structural protein.

[0024] The present invention additionally provides a method of making infectious, multiplication-defective, coronavirus particles, comprising: a) providing the helper cell of this invention; and b) producing coronavirus-particles, in the helper cell.

[0025] The foregoing and other objects and aspects of the present invention are explained in detail in the specification set forth below.

#### DETAILED DESCRIPTION OF THE INVENTION

[0026] As used herein, “a” or “an” or “the” can mean one or more than one. For example, “a” cell can mean one cell or a plurality of cells.

[0027] Also as used herein, “and/or” refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative (“or”).

[0028] Furthermore, the term “about,” as used herein when referring to a measurable value such as an amount of a compound or agent of this invention, dose, time, temperature, and the like, is meant to encompass variations of  $\pm 20\%$ ,  $\pm 10\%$ ,  $\pm 5\%$ ,  $\pm 1\%$ ,  $\pm 0.5\%$ , or even  $\pm 0.1\%$  of the specified amount.

[0029] As used herein, the transitional phrase “consisting essentially of” means that the scope of a claim is to be interpreted to encompass the specified materials or steps recited in the claim, “and those that do not materially affect the basic and novel characteristic(s)” of the claimed invention. See, *In re Herz*, 537 F.2d 549, 551-52, 190 USPQ 461, 463 (CCPA 1976) (emphasis in the original); see also MPEP § 2111.03.

[0030] “Nidovirus” as used herein refers to viruses within the order Nidovirales, including the families Coronaviridae and Arteriviridae. All viruses within the order Nidovirales share the unique feature of synthesizing a nested set of multiple subgenomic mRNAs. See M. Lai and K. Holmes, *Coronaviridae: The Viruses and Their Replication*, in *Fields Virology*, pg 1163, (4<sup>th</sup> Ed. 2001). Particular examples of Coronaviridae include, but are not limited to, toroviruses and coronaviruses.

[0031] “Coronavirus” as used herein refers to a genus in the family Coronaviridae, which family is in turn classified within the order Nidovirales. The coronaviruses are large, enveloped, positive-stranded RNA viruses. They have the largest genomes of all RNA viruses and replicate by a unique mechanism that results in a high frequency of recombination. The coronaviruses include antigenic groups I, II, and III. While the present invention is described primarily with respect to SARS coronavirus, the invention may be carried out with any coronavirus, such as transmissible gastroenteritis virus (TGEV), human respiratory coronavirus porcine respiratory coronavirus, canine coronavirus, feline enteric coronavirus, feline infectious peritonitis virus, rabbit coronavirus, murine hepatitis virus, sialodacryoadenitis virus,

porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, avian infectious bronchitis virus, and turkey coronavirus, as well as chimeras of any of the foregoing. See generally M. Lai and K. Holmes, “Coronaviridae: The Viruses and Their Replication,” in *Fields Virology*, (4<sup>th</sup> Ed. 2001).

[0032] A “nidovirus permissive cell” or “coronavirus permissive cell” as used herein can be any cell in which a coronavirus can at least replicate, including both naturally occurring and recombinant cells. In some embodiments the permissive cell is also one that the nidovirus or coronavirus can infect. The permissive cell can be one that has been modified by recombinant means to produce a cell surface receptor for the nidovirus or coronavirus.

[0033] A “heterologous RNA” as described herein can encode any protein, peptide, antisense sequence, ribozyme, etc., to be administered to a subject of this invention for any purpose. For example, the heterologous RNA can encode, and be expressed in the subject to produce, a protein or peptide. The protein or peptide may, for example, be an antigen or immunogen in embodiments where it is desired to produce antibodies in an animal subject, which antibodies can be collected and used for diagnostic and/or therapeutic purposes, or where it is desired to elicit an immune response to the protein or peptide in a subject.

[0034] A “structural protein” as used herein refers to a protein required for production of coronavirus particles of this invention, such as those encoded by the S, E, M and N genes, as well as any other structural proteins now known or later identified in the coronavirus and in particular in the SARS virus genome. In embodiments of this invention wherein the replicon RNA and/or helper RNAs lack a nucleotide sequence encoding a structural protein, the nucleotide sequence can be wholly or partly deleted, or the sequence can be present but in a mutated form, so that the net effect is that the replicon RNA and/or the helper RNA is effectively incapable of producing the necessary structural protein in functional form. Thus, for example, in an embodiment that recites a replicon RNA or helper RNA that “lacks a sequence encoding at least one coronavirus structural protein,” it is meant that the nucleotide sequence encoding the at least one coronavirus structural protein is deleted completely or in part from the replicon RNA or helper RNA or it is meant that the nucleotide sequence encoding the at least one coronavirus structural protein is present on the replicon RNA or helper RNA but in a form (e.g., mutated or otherwise altered) that cannot be expressed to produce a functional protein.

[0035] “Multiplication-defective” or “replication-defective” as used herein means that the replicon RNA contained within viral particles produced according to the present invention does not itself contain sufficient genetic information to allow for the production of new infectious viral particles.

[0036] As noted above, the present invention is based on the discovery of a full-length cDNA of the SARS coronavirus. As used herein, a “cDNA of a SARS coronavirus” or “infectious cDNA of a SARS coronavirus” is a nucleic acid molecule comprising the nucleotide sequence of a SARS coronavirus, from which RNA transcripts are produced that replicate and/or are infectious in vitro or in vivo. A SARS coronavirus cDNA of this invention can encode the

sequence of any SARS coronavirus isolate now known or later identified. The genomic sequences of some of the known SARS coronavirus isolates are set forth in Genbank and assigned Accession numbers AY278741 (SEQ ID NO:2, provided herein), AY274119, AY278554 and AY278554 and the entire contents of each of these sequences are incorporated by reference herein in their entireties as embodiments of this invention. (See also Rota et al. (2003) *Science* 300:1394; Marra et al. (2003) *Science* 300:1399; the entire contents of each of which are incorporated by reference herein for the teachings of the identification and characterization of the genomic sequence of a SARS coronavirus).

**[0037]** Two examples of a nucleic acid sequence encoding a cDNA of this invention are provided in the attached Sequence Listing as SEQ ID NO:1 and SEQ ID NO:3. As this invention encompasses all such cDNAs of all SARS coronavirus isolates, in one embodiment, the present invention provides an isolated nucleic acid comprising, consisting of and/or consisting essentially of, a nucleotide sequence selected from the group consisting of: a) SEQ ID NO:1 and/or SEQ ID NO:3 or a fragment of at least 25 contiguous nucleotide sequences of SEQ ID NO:1 and/or SEQ ID NO:3; b) a nucleotide sequence that is functionally equivalent to the nucleotide sequence of SEQ ID NO:1 and/or SEQ ID NO:3 but comprises different codons encoding the same amino acid sequences; c) a nucleotide sequence having at least 70% homology to the nucleotide sequence of SEQ ID NO:1 and/or SEQ ID NO:3; d) a nucleotide sequence having at least 95% homology to the nucleotide sequence of SEQ ID NO:1 and/or SEQ ID NO:3; e) a nucleotide sequence that hybridizes to the complement of SEQ ID NO:1 and/or SEQ ID NO:3 under stringent conditions; and f) a nucleotide sequence having complementarity (e.g., partial or complete) to any of the nucleotide sequences of (a)-(e) above.

**[0038]** The present invention further provides nucleic acid molecules comprising, consisting of and/or consisting essentially of, a fragment of at least 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 950 or 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000, 15,000, 20,000, 25,000, 26,000, 27,000, 28,000, 29,000, 29,500, etc. contiguous nucleotides (including any values within this range not specifically recited herein, e.g., 56 nucleotides or 6345 nucleotides) of the nucleotide sequence of SEQ ID NO:1, or a complement thereof. A fragment of this invention can be a fragment that hybridizes to a sequence that is unique to the cDNA of this invention. The production, identification and characterization of such fragments for desired properties as described herein is carried out according to protocols well known in the art.

**[0039]** An "isolated" nucleic acid molecule is one that is chemically synthesized (e.g., derived from reverse transcription) or is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein encoding sequences) that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an

"isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. (e.g., as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

**[0040]** In another embodiment of this invention, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence of SEQ ID NO:1 or a fragment thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence under conditions described herein, thereby forming a stable duplex.

**[0041]** The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO:1 due to degeneracy of the genetic code and thus encode the same proteins as those encoded by the nucleotide sequence of SEQ ID NO:1.

**[0042]** In particular embodiments, a nucleic acid of this invention has at least about 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more nucleic acid sequence homology with the sequences specifically disclosed herein. The term "homology" as used herein refers to a degree of similarity between two or more sequences. There can be partial homology or complete homology (i.e., identity). A partially homologous sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization to the target sequence can be examined using a hybridization assay (Southern or northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding can be tested by the use of a second target sequence, which lacks even a partial degree of complementarity (e.g., less than about 30% identity). In the absence of non-specific binding, the probe will not hybridize to the second non-complementary target sequence.

**[0043]** Alternatively stated, in particular embodiments, nucleic acids encoding a cDNA of a SARS coronavirus that hybridize under the conditions described herein to the complement of the sequences specifically disclosed herein can also be used according to the present invention. The term "hybridization" as used herein refers to any process by which a first strand of nucleic acid binds with a second strand of nucleic acid through base pairing.

**[0044]** The term "stringent" as used here refers to hybridization conditions that are commonly understood in the art to define the commodities of the hybridization procedure. High stringency hybridization conditions that will permit homologous nucleotide sequences to hybridize to a nucleic

otide sequence as given herein are well known in the art. As one example, hybridization of such sequences to the nucleic acid molecules disclosed herein can be carried out in 25% formamide, 5×SSC, 5× Denhardt's solution and 5% dextran sulfate at 42° C., with wash conditions of 25% formamide, 5×SSC and 0.1% SDS at 42° C., to allow hybridization of sequences of about 60% homology. Another example includes hybridization conditions of 6×SSC, 0.1% SDS at about 45° C., followed by wash conditions of 0.2×SSC, 0.1% SDS at 50-65° C. Another example of stringent conditions is represented by a wash stringency of 0.3 M NaCl, 0.03M sodium citrate, 0.1% SDS at 60-70° C. using a standard hybridization assay (see SAMBROOK et al., EDS., MOLECULAR CLONING: A LABORATORY MANUAL 2d ed. (Cold Spring Harbor, N.Y. 1989, the entire contents of which are incorporated by reference herein).

[0045] As is known in the art, a number of different programs can be used to identify whether a nucleic acid or amino acid has sequence identity or similarity to a known sequence. Sequence identity or similarity may be determined using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith & Waterman, *Adv. Appl. Math.* 2, 482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48,443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. 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 Drive, Madison, Wis.), the Best Fit sequence program described by Devereux et al., *Nucl. Acid Res.* 12, 387-395 (1984), preferably using the default settings, or by inspection.

[0046] An example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35, 351-360 (1987); the method is similar to that described by Higgins & Sharp, *CABIOS* 5, 151-153 (1989).

[0047] Another example of a useful algorithm is the BLAST algorithm, described in Altschul et al., *J. Mol. Biol.* 215, 403-410, (1990) and Karlin et al., *Proc. Natl. Acad. Sci. USA* 90, 5873-5787 (1993). A particularly useful BLAST program is the WU-BLAST-2 program that was obtained from Altschul et al., *Methods in Enzymology*, 266, 460-480 (1996). WU-BLAST-2 uses several search parameters, which are preferably set to the default values. The parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity. An additional useful algorithm is gapped BLAST as reported by Altschul et al. *Nucleic Acids Res.* 25, 3389-3402.

[0048] The CLUSTAL program can also be used to determine sequence similarity. This algorithm is described by Higgins et al. (1988) *Gene* 73:237; Higgins et al. (1989) *CABIOS* 5:151-153; Corpet et al. (1988) *Nucleic Acids Res.* 16: 10881-90; Huang et al. (1992) *CABIOS* 8: 155-65; and Pearson et al. (1994) *Meth. Mol. Biol.* 24: 307-331.

[0049] The alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences that contain either more or fewer nucleotides than the nucleic acids disclosed herein, it is understood that in one embodiment, the percentage of sequence identity will be determined based on the number of identical nucleotides in relation to the total number of nucleotide bases. Thus, for example, sequence identity of sequences shorter than a sequence specifically disclosed herein will be determined using the number of nucleotide bases in the shorter sequence, in one embodiment. In percent identity calculations, relative weight is not assigned to various manifestations of sequence variation, such as, insertions, deletions, substitutions, etc.

[0050] In certain embodiments of this invention, the nucleic acid of this invention can comprise a promoter that directs the production of an RNA transcript from the cDNA. This promoter can be active in vitro or in vivo to produce an RNA transcript from the SARS coronavirus cDNA that can replicate and/or is infectious. Non-limiting examples of the promoter of this invention include a T7 promoter, a SP6 promoter a T3 promoter, a CMV promoter, a MoMLV promoter, a metallothionein promoter, a glucocorticoid promoter, a SV40 promoter, a CaMV 35S promoter, a nopaline synthase promoter, and any other promoter that directs RNA transcription in vitro or in a cell.

[0051] Also provided herein is an RNA and a SARS coronavirus particle produced by the cDNA of this invention and a SARS coronavirus particle comprising the RNA produced from the cDNA of this invention. Further provided herein is a vector comprising the cDNA or RNA of this invention and a cell comprising the vector of this invention.

[0052] The present invention further provides a cDNA of a SARS coronavirus, wherein all or part of a nucleotide sequence of the cDNA is deleted and wherein the nucleotide sequence is selected from the group consisting of a nucleotide sequence encoding ORF1, ORF2, ORF3a, ORF3b, ORF4, ORF5, ORF6, ORF7a/b, ORF8a/b, ORF9a, ORF9b, any other ORF of a SARS coronavirus now known or later identified, and any combination thereof. It is also intended that the ORF sequence of this invention can be intact but altered to have the same effect as a total or partial deletion. Methods of deleting all or part of an ORF and/or altering an ORF of a cDNA of a SARS coronavirus of this invention and testing the resulting genotype and phenotype are set forth in the Examples and such methods are also routine to one of ordinary skill in the art. Examples of deletion mutants of this invention are provided in the Sequence Listing provided herein as SEQ ID NOs:4-9. These are sequences of subclone F having deletions in ORFX1 (SEQ ID NO:4), ORFX1 and X2 (SEQ ID NO:5), ORFX3 (SEQ ID NO:6), ORFX4 (SEQ ID NO:7) ORFX4 substituted with green fluorescent protein (GFP) (SEQ ID NO:8) and ORFX4 substituted with luciferase (SEQ ID NO:9).

[0053] Further provided herein is a cDNA of a SARS coronavirus, comprising a mutation in a nucleotide sequence selected from the group consisting of a nucleotide sequence encoding ORF1, ORF2, ORF3a, ORF3b, ORF4, ORF5, ORF6, ORF7a/b, ORF8a/b, ORF9a, ORF9b, ORF 10, ORF 13, ORF 14, any other ORF of a SARS coronavirus, and any combination thereof, wherein the mutation results in a nonfunctional gene product. Methods of producing the vari-



ous mutants of this invention and testing the resulting genotype and phenotype are set forth in the Examples provided herein and such methods are also routine to one of ordinary skill in the art.

[0054] In additional embodiments, the present invention provides a cDNA of a SARS coronavirus, wherein the order of nucleotide sequences encoding replicase, accessory ORFs and/or structural proteins S, E, M and N is rearranged in comparison to the order in wild type SARS coronavirus. In addition, or alternatively, one or more of the nucleotide sequences encoding replicase, accessory ORFs and/or structural proteins S, E, M and N can be present two or more times on the cDNA. Nonlimiting examples of these embodiments include cDNAs wherein the order of nucleotide sequences encoding structural proteins is: a) 5' S, N, E and M 3', b) 5' N, S, E and M 3', c) 5' E, M, S and N 3', d) 5' E, M, N and S 3', e) 5' S, N, E, M, N 3'. These examples are provided to show the order of the nucleotide sequences encoding the structural proteins S, E, M and N, relative to one another, with respect to the 5' and 3' ends of the cDNA molecule. However, these examples are not intended to be limiting in any way with respect to the positioning of the replicase, accessory ORFs and/or any other coding sequences present in the cDNA sequence. Thus, for example, the replicase, accessory ORFs and/or other coding sequences present in the cDNA sequence can be positioned anywhere (e.g., before, after, in multiple repeats before and/or after) relative to the each of the coding sequences of the S, E, M and N proteins. Furthermore, any of the coding sequences of the structural proteins, replicase, accessory ORFs and/or other coding sequences can be modified by mutation and/or deletion in this embodiment.

[0055] The present invention also provides a cDNA of a SARS coronavirus, comprising an attenuating mutation in a consensus sequence, ACGAAC, which can be, but is not limited to, a leader consensus sequence, an S(ORF2) consensus sequence, an ORF3a consensus sequence, an E consensus sequence, an M consensus sequence, an ORF6 consensus sequence, an ORF7 consensus sequence, an ORF8 consensus sequence, an N consensus sequence, and any combination thereof. The mutations of this invention can also include any mutation in any combination of the six consensus nucleotides and the surrounding flanking nucleotides (+/-50 nucleotides) that function as regulatory junctions to direct transcription of full length and subgenomic mRNAs. Two nonlimiting examples of mutations in the consensus sequence include "ACGAAC" to "ACCAAC" and "ACGAAC" to AGGAAG." Other examples of mutations of this invention are provided in Tables 1 and 2.

[0056] Further provided herein is a cDNA of a SARS coronavirus comprising a 29 bp insertion in ORF8a/b, or any other modification or alteration that reproduces a full length ORF8 protein sequence. In some embodiments, the full length ORF8 protein sequence provides an attenuating phenotype to the SARS coronavirus.

[0057] In additional embodiments, the present invention provides a cDNA of a SARS coronavirus wherein all or part of a nucleotide sequence of the cDNA is deleted and wherein the nucleotide sequence is selected from the group consisting of a nucleotide sequence encoding ORF1, ORF2, ORF3a, ORF3b, ORF4, ORF5, ORF6, ORF7a/b, ORF8a/b, ORF9a, ORF9b, any other ORF of a SARS coronavirus now

known or later identified, and any combination thereof and furthermore, wherein the order of nucleotide sequences of the cDNA encoding replication, accessory ORFs and/or structural proteins S, E, M and N is rearranged in comparison to the order in wild type SARS coronavirus. In some variations of this embodiment, it is also contemplated that the cDNA can further comprise an attenuating mutation in a consensus sequence of a nucleotide sequence of the cDNA selected from the group consisting of a leader consensus sequence, an S consensus sequence, an ORF3a consensus sequence, an E consensus sequence, an M consensus sequence, an ORF6 consensus sequence, an ORF7 consensus sequence, an ORF8 consensus sequence, an N consensus sequence and any combination thereof.

[0058] Furthermore, in additional embodiments, the present invention provides a SARS coronavirus replicon RNA comprising a coronavirus (e.g., SARS) packaging signal and a heterologous RNA sequence, wherein the replicon RNA lacks a sequence encoding at least one coronavirus structural protein.

[0059] Further provided herein is a DNA encoding a replicon RNA of this invention. In certain embodiments, the DNA can comprise a promoter to direct the transcription of the RNA, either in vitro or within a cell.

[0060] The present invention additionally provides an infectious, multiplication-defective, coronavirus particle, comprising a SARS coronavirus replicon RNA, wherein the replicon RNA comprises a coronavirus (e.g., SARS) packaging signal and a heterologous RNA sequence, and wherein the replicon RNA lacks a sequence encoding at least one coronavirus structural protein.

[0061] In some embodiments, the replicon RNA of this invention can comprise a nucleic acid sequence encoding at least one coronavirus structural protein, provided the replicon RNA does not comprise nucleic acid sequences functionally encoding all of the coronavirus structural proteins. In other embodiments, the replicon RNA can comprise a promoter. In yet other embodiments, the replicon RNA may or may not comprise a nucleic acid sequence encoding a replicase protein. It is also contemplated that the replicon RNA can lack a coronavirus packaging signal under circumstances wherein the replicon RNA is packaged into coronavirus particles nonspecifically or under conditions wherein it is contacted with a coronavirus structural protein comprising a nucleic acid binding site that facilitates packaging of the replicon RNA in the absence of a coronavirus packaging signal.

[0062] It is also an embodiment of this invention wherein the nucleic acid of the replicon RNA encodes, and/or the particle itself comprises, a coronavirus structural protein that is produced from nucleic acid of a coronavirus that can be, but is not limited to, SARS coronavirus, human respiratory coronavirus, mouse hepatitis virus, porcine transmissible gastroenteritis virus, porcine respiratory coronavirus, canine coronavirus, feline enteric coronavirus, feline infectious peritonitis virus, rabbit coronavirus, murine hepatitis virus, sialodacryoadenitis virus, porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, avian infectious bronchitis virus, turkey coronavirus and/or any other coronavirus now known or later identified, as well as any combination thereof, thereby resulting in chimeric coronavirus particles.

[0063] In embodiments wherein one or more of the structural proteins are from different coronaviruses in a particle and/or one or more structural proteins are encoded by the nucleic acid of the replicon RNA, the coronavirus structural protein encoded by the nucleic acid of the replicon RNA can be S, E, M, N or combinations thereof.

[0064] The coronavirus packaging signal of this invention can be a packaging signal of any coronavirus now known or later identified. For example, the packaging signal can be from a coronavirus that can be, but is not limited to, SARS coronavirus, human respiratory coronavirus, mouse hepatitis virus, porcine transmissible gastroenteritis virus, porcine respiratory coronavirus, canine coronavirus, feline enteric coronavirus, feline infectious peritonitis virus, rabbit coronavirus, murine hepatitis virus, sialodacryoadenitis virus, porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, avian infectious bronchitis virus, turkey coronavirus.

[0065] Further provided herein is a population of infectious, multiplication defective, coronavirus particles, wherein each particle comprises a SARS coronavirus replicon RNA, and wherein the replicon RNA comprises a coronavirus (e.g., SARS) packaging signal and a heterologous RNA sequence, and wherein the replicon RNA lacks a sequence encoding at least one coronavirus structural protein, wherein the population contains no detectable replication-competent coronavirus particles as determined by passage on coronavirus permissive cells in culture. Methods of detecting replication competent particles by passage on cell culture are standard in the art. These assays can also be carried out by passage of the replicon particles of this invention on cells constitutively expressing nucleic acid encoding the missing coronavirus structural protein(s), with the expected result of obtaining replication competent coronavirus particles.

[0066] It is also contemplated that the replicon RNA and/or replicon particles of this invention can comprise coronavirus RNA and/or structural proteins that comprise any or all of the gene order rearrangements, deletions and/or mutations described herein that can be present in the SARS coronavirus cDNA of this invention.

[0067] Another aspect of the present invention is a renetworked or rewired nidovirus genome and/or replicon RNA that results in a genetic trap for wild type viruses should recombination occur, e.g., when the genome or replicon RNA is contacted with wild type virus. Thus, in one embodiment, the present invention provides an isolated nucleic acid comprising, consisting essentially of and/or consisting of a nucleotide sequence encoding a Nidovirus genome or replicon RNA, wherein the genome or replicon RNA comprises, consists essentially of and/or consists of one or more of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural genes and further wherein wild type CS sequences are present in the CS of the TRS for the group specific ORF's 3a/b, ORF4, ORF5, ORF6, ORF7a/b, ORF8a/b and ORF9a/b. This remodeled genome or replicon RNA can be considered to be partially remodeled (e.g., PRG). Other mutations in the genome or replicon RNA may or may not be present, provided such other mutants do not affect the genetic trap function of the rewired (e.g., mutated) CSs.

[0068] Further provided herein is an isolated nucleic acid comprising, consisting essentially of and/or consisting of a nucleotide sequence encoding a Nidovirus genome or replicon RNA, wherein the genome or replicon RNA comprises, consists essentially of and/or consists of one or more of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural genes and further comprising one or more of the same mutations in the CS of the TRS located upstream of open reading frame (ORF) 3a/3b, ORF 4, ORF 5, ORF6, ORF 7a/7b, ORF 8a/b and ORF 9a/b and downstream of the leader RNA. This remodeled genome or replicon RNA can be considered to be completely remodeled (e.g., CRG). Other mutations in the genome or replicon RNA may or may not be present, provided such other mutants do not affect the genetic trap function of the rewired (e.g., mutated) CSs.

[0069] The isolated nucleic acids described herein can be from a Nidovirus that is a severe acute respiratory syndrome (SARS) coronavirus having the CS of ACGAAC. In this embodiment, the mutation can be a single mutation in the CS sites (e.g., ACGGAC, etc.), two mutations in the CS sites (e.g., ACGGAT, CCGGAC, CCGAAT, etc.), three mutations in the CS sites (e.g., CCGGAT, CCGCGC, CGCAAC, etc.), four mutations in the CS sites (e.g., CCGGAT, AGCGAT, etc.), five mutations in the CS sites (CGCGAT, CCCGTT, etc.) and six mutations in the CS sites (CGCGTT, TGCGGT, etc.). It is contemplated in this invention and applicants are in possession of the embodiment wherein the CS sequence is mutated according to any one of 4<sup>6</sup> possible combinations of sequence variations, which combinations can be readily calculated and identified according to methods standard in the art. Factors governing site selection for mutation are based on 1) a unique sequence element that is not repeated elsewhere in the genome and 2) a mutant CS that functions as a regulatory start site when coupled with compensating changes at the leader CS site. Thus, the mutation(s) can be any possible combination of changes in the body CSs and in the leader CS and a particular mutation or combination of mutations is not critical as long as the same mutations are present in the body CSs as in the leader CS. In other words, the actual mutant CS sequence is not critical, the major factor being that CS sites must allow for communication via efficient base-pairing for discontinuous transcription of subgenomic RNAs, thus the same mutation(s) is present in the leader CS as in the body CS(s).

[0070] In further embodiments, the Nidovirus of this invention can be a group I coronavirus having the CS of CUAAAC and the mutation can be a single mutation (e.g., GUAAAC, etc.), two mutations (e.g., GCAAAC, etc.), three mutations (e.g., CGAAAG, etc.), four mutations (e.g., GCTAAAG, etc.), five mutations (e.g., GCTTAG, etc.) and/or six mutations (GCTTGG, etc.). It is contemplated as part of this invention and applicants are in possession of a total of 4<sup>6</sup> possible combinations of sequence variation in the CS of this invention, as could be identified and produced according to standard methods. Factors to consider in selecting mutations include whether the mutation(s) are unique and able to interact with an identical leader CS site to drive expression of subgenomic mRNAs.

[0071] The Nidovirus of this invention can also be a group II coronavirus having the CS of TCTAAAC and the mutation can be a single mutation (e.g., CCTAAAC, etc.), two

mutations (e.g., CCGAAAC, etc.), three mutations (e.g., CGTAAAG, etc.), four mutations (e.g., CCGAAGG, etc.), five mutations (e.g., CGTCCGC, etc), six mutations (e.g., CGGATTG, etc) and/or seven mutations (e.g., CGGCCTG, etc). It is contemplated as part of this invention and applicants are in possession of a total of  $4^7$  possible combinations of sequence variation in the CS of this invention, as could be identified and produced according to standard methods. As noted herein for other nidoviruses, the principle requirements for mutation selection are that the sequence not be located elsewhere in the genome and that it function to regulate subgenomic transcription when paired with a leader TRS of like sequence.

[0072] In yet further embodiments, the Nidovirus of this invention can be a group III coronavirus having the CS of CUUAACAA and the mutation can be a single mutation (e.g., CUUAAGAA, etc.) two mutations (e.g., GUUAA-GAA, etc.) three mutations (GUUGAGAA, etc.), four mutations (e.g., GUUUCAG, five mutations (e.g., CAAG-GCAA, TCCAAGAT, etc.), six mutations (e.g., GUUCCTC, etc.), seven mutations (e.g., GCCTAGCG, etc.) and/or eight mutations (e.g., GCCTGGCT, etc.). It is contemplated as part of this invention and applicants are in possession of a total of  $4^8$  possible combinations of sequence variation in the CS of this invention, as could be identified and produced according to standard methods.

[0073] The present invention further provides an embodiment wherein the Nidovirus is a torovirus having a CS regulatory sequence of UUUAGA and the mutation is a single mutation (e.g., GUUAGA, etc.) two mutations (e.g., GUUGGA, etc.), three mutations (e.g., GUUGCA, etc.), four mutations (e.g., GCUCCA, etc.) five mutations (e.g., GCCACT, etc.) and/or six mutations (e.g., GCCTCT, etc.). It is contemplated as part of this invention and applicants are in possession of a total of  $4^6$  possible combinations of sequence variation in the CS of this invention, as could be identified and produced according to standard methods. As noted herein for other nidoviruses, the exact mutation(s) are dependent upon the uniqueness of this sequence in the different Torovirus genome CS regions coupled with its ability to regulate subgenomic transcription via a matching mutation in the leader CS region.

[0074] The nidovirus of this invention can also be an arterivirus having a CS of UCNUAACC, U(A/G)(U/A)AACC, or UUAACC and the mutation can be a single mutation (e.g., CUAACC, etc.) two mutations (e.g., CCAACC, etc.), three mutations (e.g., CCAAGC, etc.) four mutations (e.g., CCAGGC, etc.) five mutations (e.g., CCAGGT, etc.) and/or six mutations (e.g., GGTTAG, etc.), including a cluster of mutations in a six nucleotide segment of these CSs. It is contemplated as part of this invention and applicants are in possession of a total of  $4^9$  possible combinations of potential mutant CS sites available for use in this invention. The mutated CS sequence should be unique to that particular arterivirus and able to interact with the identically mutated leader CS site located at the 5' end of the genome.

[0075] The present invention further provides the isolated nucleic acids described herein having all of the body and leader CS sites reengineered by mutation (e.g., CRG), wherein the nucleic acid further comprises a secondary CS trap that is preferentially recognized in RNA recombinant

viruses. These secondary CS trap sites are engineered wild type CSs located within a structural protein gene that lead to expression of one or more subgenomic mRNAs that do not encode a full length structural protein when present in a recombinant virus.

[0076] Thus, a CRG genome or replicon RNA of this invention can further comprise, consist essentially of and/or consist of a wild type CS engineered within one or more structural protein genes. As one example, the genome or replicon RNA can be of a SARS coronavirus having a wild type CS sequence of ACGAAC. In addition to mutations in the body CSs and leader CS, additional mutations would be introduced into the N, M, E and/or S gene(s) in any combination to provide an artificial CS having the same nucleotide sequence as the wild type CS, ACGAAC. This artificial CS is introduced into the structural gene sequence as described herein in Example 6.

[0077] Thus, in one embodiment, the present invention provides an isolated nucleic acid comprising, consisting essentially of and/or consisting of a nucleotide sequence encoding a Nidovirus genome or replicon RNA, wherein the genome or replicon RNA comprises, consists essentially of and/or consists of one or more of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural genes and further comprising one or more of the same mutations in the CS of the TRS located upstream of open reading frame (ORF) 3a/3b, ORF 4, ORF 5, ORF6, ORF 7a/7b, ORF 8a/b and ORF 9a/b and downstream of the leader RNA and further comprising one or more than one artificial CS having the nucleotide sequence of the wild type CS of the nidovirus, in one or more of the structural protein genes (e.g., N, M, E and/or S).

[0078] Additionally provided is an isolated nucleic acid comprising, consisting essentially of and/or consisting of a nucleotide sequence encoding a Nidovirus genome or replicon RNA, wherein the genome or replicon RNA comprises, consists essentially of and/or consists of one or more of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural genes S, E, M and N and further wherein wild type CS sequences are present in the CS of the TRS for the group specific ORFs 3a/b, ORF4, ORF5, ORF6, ORF7a/b, ORF8a/b and ORF9a/b and further comprising one or more than one artificial CS having the nucleotide sequence of the wild type CS of the nidovirus, in one or more of the structural protein genes (e.g., N, M, E and/or S).

[0079] The present invention further provides a nidovirus particle as well as a population of nidovirus particles comprising any of the nucleic acids of this invention

[0080] The nidovirus and/or nidovirus particle of this invention can be a coronavirus, a torovirus, an arterivirus and chimeras thereof, as are known in the art.

[0081] A coronavirus of this invention can be, but is not limited to, transmissible gastroenteritis virus (TGEV), human respiratory coronavirus, human coronavirus, porcine respiratory coronavirus, porcine epidemic diarrhea virus, respiratory bovine virus, canine coronavirus, bat SARS-CoV, human coronavirus NL63, human coronavirus HKU1,

human coronavirus OC43, human coronavirus 229E, feline enteric coronavirus, bat coronaviruses, feline infectious peritonitis virus, rabbit coronavirus, murine hepatitis virus, sialodacryoadenitis virus, porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, avian infectious bronchitis virus, turkey coronavirus, and a chimera of any combination of these viruses.

[0082] An arterivirus of this invention can be but is not limited to, equine arteritis virus, lactate dehydrogenase-elevating virus, simian hemorrhagic fever virus, porcine reproductive and respiratory disease virus, human arteriviruses, and chimeras of any combination of these viruses containing, e.g., replicase protein genes from one arterivirus fused with structural ORFs and group specific ORFs of other arteriviruses.

[0083] A torovirus of this invention can include, but is not limited to, bovine torovirus, equine torovirus, human torovirus, porcine torovirus and a chimera of any combination of these viruses.

[0084] An example of a chimera of any combination of these nidoviruses is a chimeric virus comprising-replicase protein genes from one nidovirus fused with structural ORFs and group specific ORFs of other nidoviruses

[0085] Additionally provided herein is a composition comprising the nucleic acids, nidovirus particles and/or population of nidovirus particles as described herein and a pharmaceutically acceptable carrier.

[0086] In further embodiments, the present invention provides a method of eliciting an immune response in a subject, comprising administering to, delivering to, and/or introducing into the subject an effective amount of the nucleic acids, viruses, particles, compositions and/or populations of this invention.

[0087] Also provided herein is a method of treating and/or preventing a Nidovirus infection in a subject, comprising administering/delivering/introducing into to the subject an effective amount of the nucleic acids, viruses, particles, compositions and/or populations of this invention.

[0088] Methods are also provided herein for producing a nidovirus particle comprising a replicon RNA or a nucleic acid comprising a nucleotide sequence encoding a Nidovirus genome, wherein the genome or replicon RNA comprises one or more of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural genes S, E, M and N and further comprising a wild type CS sequence in a TRS for ORFs 3a/b, ORF6, ORF7a/b and ORF8a/b, comprising introducing the replicon RNA or nucleic acid into a nidovirus-permissive cell under conditions whereby nidovirus particles are produced.

[0089] Further provided herein is a method of producing a nidovirus particle comprising a replicon RNA or a nucleic acid comprising a nucleotide sequence encoding a Nidovirus genome, wherein the replicon RNA or genome comprises one or more of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural genes S, E, M and N and further comprising one or more of the same mutations in the CS of the TRS located upstream of open reading frame (ORF) 3a/3b,

ORF6, ORF 7a/7b, ORF 8a/b and ORF 9a/b, comprising introducing the nucleic acid or replicon RNA into a nidovirus-permissive cell under conditions whereby nidovirus particles are produced.

[0090] In addition, the present invention provides a method of producing a nidovirus particle comprising a replicon RNA or a nucleic acid comprising a nucleotide sequence encoding a Nidovirus genome, wherein the replicon RNA or genome comprises one or more of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural protein genes and further comprising one or more of the same mutations in the CS of the TRS located upstream of open reading frame (ORF) 3a/3b, ORF6, ORF 7a/7b, ORF 8a/b and ORF 9a/b and further comprising one or more than one artificial CS having the nucleotide sequence of the wild type CS of the nidovirus in one or more of the structural protein genes, comprising introducing the nucleic acid or replicon RNA into a nidovirus-permissive cell under conditions whereby nidovirus particles are produced.

[0091] Methods are also provided herein for producing a nidovirus particle comprising a replicon RNA or a nucleic acid comprising a nucleotide sequence encoding a Nidovirus genome, wherein the genome or replicon RNA comprises one or more of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural protein genes (e.g., S, E, M and/or N) and further comprising a wild type CS sequence in a TRS for ORFs 3a/b, ORF6, ORF7a/b and ORF8a/b and further comprising one or more than one artificial CS having the nucleotide sequence of the wild type CS of the nidovirus in one or more of the structural protein genes (e.g., S, E, M and/or N), comprising introducing the replicon RNA or nucleic acid into a nidovirus-permissive cell under conditions whereby nidovirus particles are produced.

[0092] In methods wherein nidovirus particles are made that comprise a replicon RNA of this invention, a "nidovirus-permissive cell" is a cell that contains transcripts encoding the structural proteins that are not encoded for on the replicon RNA, as described herein. For example, a replicon RNA comprising a remodeled TRS CS site in the leader region and in the N gene can be packaged by transfection into cells containing transcripts encoding the remaining structural proteins, S, E and M.

[0093] The present invention further provides nidovirus particles produced by the methods described herein.

[0094] Also provided herein is a method of producing a nidovirus genome or replicon RNA comprising reengineered CS sequences for the purpose of preventing recombination repair of a live attenuated or replicon viral construct and/or preventing repair of attenuating alleles and/or for stabilizing attenuating mutations, comprising producing the nucleic acids of this invention. Additionally provided herein is a method of preventing recombination repair or repair of attenuating alleles of a live attenuated or replicon construct to be used in a therapeutic and/or immunological method comprising producing the nucleic acids of this invention. Thus, the present invention provides therapeutic and immunological compositions that are engineered to prevent recombination repair or repair of attenuating alleles if the

nucleic acids of the compositions are contacted with wild type virus under conditions whereby recombination can occur. Further provides are therapeutic and immunological compositions that are engineered to stabilize attenuating mutations in the nucleic acid therein.

[0095] The nucleic acids, viruses, vectors, particles and populations are intended for use as therapeutic agents and immunological reagents, for example, as antigens, immunogens, vaccines, and/or nucleic acid delivery vehicles. Thus, in various embodiments, the present invention provides a composition comprising the nucleic acid, virus, vector, particle, and/or population of this invention in a pharmaceutically acceptable carrier. The compositions described herein can be formulated for use as reagents (e.g., to produce antibodies) and/or for administration in a pharmaceutical carrier in accordance with known techniques. See, e.g., Remington, *The Science And Practice of Pharmacy* (latest edition). In the manufacture of a pharmaceutical composition according to embodiments of the present invention, the composition of this invention is typically admixed with, inter alia, a pharmaceutically acceptable carrier. By "pharmaceutically acceptable carrier" is meant a carrier that is compatible with other ingredients in the pharmaceutical composition and that is not harmful or deleterious to the subject. The carrier may be a solid or a liquid, or both, and is preferably formulated with the composition of this invention as a unit-dose formulation. The pharmaceutical compositions are prepared by any of the well-known techniques of pharmacy including, but not limited to, admixing the components, optionally including one or more accessory ingredients. Exemplary pharmaceutically acceptable carriers include, but are not limited to, sterile pyrogen-free water and sterile pyrogen-free physiological saline solution. Such carriers can further include protein (e.g., serum albumin) and sugar (sucrose, sorbitol, glucose, etc.)

[0096] The pharmaceutical compositions of this invention include those suitable for oral, rectal, topical, inhalation (e.g., via an aerosol) buccal (e.g., sub-lingual), vaginal, parenteral (e.g., subcutaneous, intramuscular, intradermal, intraarticular, intrapleural, intraperitoneal, intracerebral, intraarterial, or intravenous), topical (i.e., both skin and mucosal surfaces, including airway surfaces) and transdermal administration. The compositions herein may also be administered via a skin scarification method, or transdermally via a patch or liquid. The compositions may be delivered subdermally in the form of a biodegradable material that releases the compositions over a period of time. The most suitable route in any given case will depend, as is well known in the art, on such factors as the species, age, gender and overall condition of the subject, the nature and severity of the condition being treated and/or on the nature of the particular composition (i.e., dosage, formulation) that is being administered.

[0097] Subjects to whom the viruses, vectors, particles, populations and/or other compositions of this invention can be administered according to the methods described herein can be any subject, generally vertebrates, for which the particles, populations and/or compositions are infectious, including but not limited to, birds and mammals such as pigs, mice, cows, and humans.

[0098] As used herein, an "effective amount" refers to an amount of a compound or composition that is sufficient to

produce a desired effect, which can be a therapeutic, prophylactic and/or beneficial effect.

[0099] Thus, the present invention provides a method of inducing an immune response in a subject, comprising administering to the subject an effective amount of a virus, vector, particle, population and/or composition of this invention.

[0100] The present invention also provides a method of treating and/or preventing a SARS coronavirus infection in a subject, comprising administering to the subject an effective amount of a virus, vector, particle, population and/or composition of this invention.

[0101] Also as used herein, the terms "treat," "treating" and "treatment" include any type of mechanism, action or activity that results in a change in the medical status of a subject, including an improvement in the condition of the subject (e.g., change or improvement in one or more symptoms and/or clinical parameters), delay in the progression of the condition, prevention or delay of the onset of a disease or illness, etc.

[0102] One example of an effective amount is from about  $10^4$  to about  $10^{10}$ , preferably  $10^5$  to  $10^9$ , and in particular  $10^6$  to  $10^8$  infectious units (IU, as measured by indirect immunofluorescence assay), or virus particles, per dose, which can be administered to a subject, depending upon the age, species and/or condition of the subject being treated.

[0103] In some embodiments of the present invention, the compositions can be administered with an adjuvant. As used herein, "adjuvant" describes a substance, which can be any immunomodulating substance capable of being combined with the polypeptide or nucleic acid vaccine to enhance, improve or otherwise modulate an immune response in a subject without deleterious effect on the subject.

[0104] Non-limiting examples of adjuvants that can be used in the vaccine of the present invention include the RIBI adjuvant system (Ribi Inc., Hamilton, Mont.), alum, mineral gels such as aluminum hydroxide gel, oil-in-water emulsions, water-in-oil emulsions such as, e.g., Freund's complete and incomplete adjuvants, Block copolymer (CytRx, Atlanta Ga.), QS-21 (Cambridge Biotech Inc., Cambridge Mass.), SAF-M (Chiron, Emeryville Calif.), AMPHIGENT™, adjuvant, saponin, Quil A or other saponin fraction, monophosphoryl lipid A, and Avridine lipid-amine adjuvant. Non-limiting examples of oil-in-water emulsions useful in the vaccine of the invention include modified SEAM62 and SEAM 1/2 formulations. Modified SEAM62 is an oil-in-water emulsion containing 5% (v/v) squalene (Sigma), 1% (v/v) SPANT™ 85 detergent (ICI Surfactants), 0.7% (v/v) TWEEN™ 80 detergent (ICI Surfactants), 2.5% (v/v) ethanol, 200 µg/ml Quil A, 100 µg/ml cholesterol, and 0.5% (v/v) lecithin. Modified SEAM 1/2 is an oil-in-water emulsion comprising 5% (v/v) squalene, 1% (v/v) SPANT™ 85 detergent, 0.7% (v/v) Tween 80 detergent, 2.5% (v/v) ethanol, 100 µg/ml Quil A, and 50 µg/ml cholesterol. Other immunomodulatory agents that can be included in the vaccine include, e.g., one or more interleukins, interferons, or other known cytokines.

[0105] SARS coronavirus vectors also provide a system for the incorporation and expression of one or more heterologous nucleic acids, as coronaviruses contain a polycistronic genome organization and synthesize multiple subge-

nomic-length mRNAs (Enjuanes and van der Zeijst (1995) In: S. G. Siddell (ed.), *The Coronaviridae*. Plenum Press, New York, N.Y., p. 337-376).

[0106] In certain embodiments, the present invention describes the assembly of recombinant transmissible virus and replicons that express heterologous nucleic acids, which can be used to deliver such nucleic acids and/or make vaccines/immunogenic compositions against homologous and heterologous pathogens (Agapov et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:12989-12994; Balasuriya et al. (2000) *J. Virol.* 74:10623-10630; Berglund et al. (1998) *Nat. Biotechnol.* 16:562-565; Bredenbeek et al. (1993) *J. Virol.* 67:6439-6446; DiCiommo and Bremner (1998) *J. Biol. Chem.* 273:18060-18066; Dollenmaier et al. (2001) *Virology* 281:216-230; Dubensky et al. (1996) *J. Virol.* 70:508-519; Hevey et al. (1998) *Virology* 251:28-37; Johanning et al. (1995) *Nucleic Acids Res.* 23:1495-1501; Khromykh (2000) *Curr. Opin. Mol. Ther.* 2:555-569; Khromykh and Westaway (1997) *J. Virol.* 71:1497-1505; Liljestrom and Garoff (1991) *Bio/Technology* 9:1356-1361; Percy et al. (1992) *J. Virol.* 66:5040-5046; Porter, et al. (1993) *J. Virol.* 67:3712-3719; Pushko et al. (2000) *Vaccine* 19:142-153; Schultz-Cherry et al (2000) *Virology* 278:55-59; Varnavski and Khromykh (1999) *Virology* 255:366-375; Varnavski et al. (2000) *J. Virol.* 74:4394-4403).

[0107] The use of replicons as a vaccine delivery system offers a number of important advantages over the use of live, attenuated virus vaccines, which are capable of independent spread and recombination with wild-type virus populations. Replicon vectors are an inherently safer alternative to the use of live, attenuated virus vaccines due to the lack of progeny virus production. In addition, high-level expression of heterologous nucleic acids can result in the use of a relatively low dose of virus replicon particles (VRPs) for vaccination and immune induction. Moreover, gene order rearranged and/or otherwise attenuated replicon particles will be inherently more stable and less pathogenic than attenuated wild-type strains.

[0108] Thus, the present invention also provides a method of introducing a heterologous RNA into a subject, comprising administering to the subject an effective amount of the particles and/or the populations and/or compositions of this invention comprising these particles or populations. The heterologous RNA can encode any protein or peptide or antisense sequence or ribozyme and can be administered to impart any type of effect (e.g., immunological or therapeutic, etc.).

[0109] The production of virus replicon particles is well known in the art for a variety of virus systems, including coronaviruses (see, e.g., Curtis et al. (2002) *J. Virol.* 76:1422-1434; PCT Publication No. WO 02/086068, the entire contents of each of which are incorporated by reference herein). The present invention can also be implemented in any of a variety of ways, including by techniques, compositions and formulations known in the art (see, e.g., U.S. Pat. No. 6,593,311 to Baric et al.; U.S. Pat. No. 6,156,558 to Johnston et al.; and U.S. Pat. No. 5,639,650 to Johnston et al.; U.S. Pat. No. 6,342,372 to Dubensky et al.) modified in light of the teachings set forth herein. Applicants specifically intend that the disclosures of all United States patent references and patent publications cited herein be incorporated herein by reference in their entirety.

[0110] The synthesis of large RNA transcripts (~27 to 29 kb) in vitro is problematic, and the electroporation of such large RNA constructs, even in the presence of enhancing N transcripts, has also proven difficult, resulting in a 1% transfection efficiency. Therefore, transfecting cells with helper packaging constructs and subsequently passing the coronavirus VRPs in the presence of VEE-(E) VRPs can address this issue. In this way, VRPs can be amplified and high concentrations may amplify replicon titers for future applications. In addition, the use of a DNA launch platform, such as with a cytomegalovirus promoter, may be used to overcome any problems associated with an RNA launch system.

[0111] The strategy presented herein for the assembly of SARS replicon constructs was based on a similar strategy for the construction of TGEV cDNA, employing six cDNA subclones that span the entire length of the SARS genome (see Yount et al. *J. Virol.* 74:10600-10611 (2000)). Each fragment is flanked by restriction sites that leave unique interconnecting junctions of 3 or 4 nt in length. These sticky ends are not complementary to most other sticky ends generated with the same enzyme at other sites in the DNA, allowing for the systematic assembly of SARS cDNAs by in vitro ligation.

[0112] The infectious, replication defective, coronavirus particles can be prepared according to the methods disclosed herein in combination with techniques known to those skilled in the art. As one example, the method can comprise a) introducing into a coronavirus-permissive cell 1) a SARS coronavirus replicon RNA comprising a coronavirus packaging signal and a heterologous RNA, a first helper RNA encoding at least one coronavirus structural protein and a second (and possibly third, fourth, etc.) helper RNA encoding at least one coronavirus structural protein that is different from that encoded by the first helper RNA; b) producing the coronavirus particles in the cell; and c) optionally collecting the particles from the cell. The step of introducing the replicon RNA and helper RNA(s) into the coronavirus-permissive cell can be carried out according to any suitable means known to those skilled in the art. For example, uptake of the RNA into the cell can be achieved by any suitable means, such as for example, by treating the cells with DEAE-dextran, treating the cells with "LIPOFECTIN™," and/or by electroporation, with electroporation being the currently preferred means. These techniques are well known in the art. See e.g., U.S. Pat. No. 5,185,440 to Davis et al., and PCT Publication No. WO 92/10578 to Bioption AB, the disclosures of which are incorporated herein by reference in their entirety. Alternatively, a DNA encoding the replicon RNA and/or a DNA encoding the helper RNA(s) can be introduced into the cell according to known methods and the DNA can be transcribed into RNA within the cell.

[0113] The present invention also provides methods for producing SARS coronavirus replicon particles, as well as helper RNAs and helper cells employed in the production. Thus, in further embodiments, the present invention provides a helper cell for producing an infectious, multiplication-defective, coronavirus particle, comprising: (a) a SARS coronavirus replicon RNA comprising a SARS coronavirus packaging signal and a heterologous RNA sequence, wherein said replicon RNA further lacks a sequence encoding at least one coronavirus structural protein; and (b) at least one separate helper RNA encoding the at least one

structural protein absent from the replicon RNA, said helper RNA lacking a coronavirus packaging signal; wherein the combined expression of the replicon RNA and the helper RNA produces an infectious, multiplication-defective coronavirus particle.

[0114] As noted above, the replicon RNA can further comprise a sequence encoding at least one of the coronavirus structural proteins, provided that the replicon RNA does not comprise nucleic acid encoding all of the coronavirus structural proteins.

[0115] In the helper cells of this invention, the helper RNA can comprise a nucleic acid sequence encoding a coronavirus structural protein that can be E, M, N, S, or any combination thereof, provided that the helper RNA does not comprise nucleic acid encoding all of the coronavirus structural proteins. The nucleic acid encoding the coronavirus structural protein can be from a coronavirus that can be SARS coronavirus, human respiratory coronavirus, mouse hepatitis virus, porcine transmissible gastroenteritis virus, porcine respiratory coronavirus, canine coronavirus, feline enteric coronavirus, feline infectious peritonitis virus, rabbit coronavirus, murine hepatitis virus, sialodacryoadenitis virus, porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, avian infectious bronchitis virus, turkey coronavirus and/or any other coronavirus now known or later identified, as well as combinations thereof, thereby resulting in chimeric coronavirus particles.

[0116] Furthermore, the helper RNA can comprise nucleic acid of any other non-coronavirus, including, but not limited to, vaccinia virus, alphavirus, flavivirus, vesicular stomatitis virus, poxvirus, adenovirus, herpesvirus, paramyxovirus, parvovirus, papovavirus, adeno-associated virus, and retrovirus. The helper RNA can also be a vector of non-viral origin that provides nucleic acid encoding the coronavirus structural protein(s) not present on the replicon RNA in the helper cell.

[0117] In certain embodiments, the replicon RNA and/or the helper RNA contains at least one attenuating gene order rearrangement among the 3A, 3B, HP, S, E, M and N genes. Helper or replicon RNAs (and their corresponding DNAs) that contain two or more genes optionally but preferably include a gene order rearrangement to attenuate (e.g., reduce the virulence) as compared to a corresponding wild-type virus that does not contain such a gene order rearrangement (i.e., comparing a virus with all of the necessary genes and the order rearrangement with a wild-type virus). Depending upon the number of genes within the helper or replicon RNA, it may contain two, three, or four or more gene order rearrangements. The wild-type gene order, from 5' to 3', is: S, 3A, 3B, E, M, N, and HP. For example, modified orders for the replicon RNA, when the helper RNA(s) contains the E gene, may include: S, 3B, M, N, HP and 3A; 3A, 3B, M, N, HP and S; S, 3A, 3B, N, Hp, and M; etc. In other embodiments, the E gene may be provided alone on the replicon RNA, and the helper RNA(s) may contain the genes described above in the orders given above. Modified orders for the replicon RNA, when the helper RNA(s) contains the M and N (in natural or reverse order) genes, may include: 3A, 3B, E, S and HP; S, E, 3A, 3B, and HP; S, 3A, 3B, E, and HP; S, E, HP, 3A, and 3B, etc. In other embodiments, the replicon RNA may contain the M and N genes (in natural or reverse orders), and the remaining genes may be provided

on the helper RNA(s) in orders such as given above. In still other embodiments, where the helper RNA(s) contains the S gene, the replicon RNA may contain the remaining genes in the order 3B, 3A, E, M, N, and HP; 3A, 3B, E, N, M, and HP; 3B, 3A, E, N, M, and HP; etc. Again, the replicon RNA may contain the S gene, and the helper RNA(s) may contain the remaining genes in the orders given above. The 3A, 3B, and HP genes are nonessential and some or all may be deleted, or they may be included in an alternate order to serve as attenuating mutations. The genes may be divided among multiple helper RNAs, some or all of which contain gene order rearrangements. The foregoing examples are merely illustrative, and numerous additional variations will be readily apparent to those skilled in the art.

[0118] In other embodiments, an attenuating mutation can be introduced by deleting one or more of the nonessential genes 3A, 3B, and HP.

[0119] In additional embodiments of the helper cells of this invention, the helper RNA(s) and/or the replicon RNA can comprise a promoter.

[0120] Additionally provided herein as an embodiment of this invention is a DNA encoding a helper RNA of this invention and a helper cell comprising said helper RNA-encoding DNA, as well as a DNA encoding a replicon RNA of this invention and a helper cell comprising said replicon-encoding DNA. The DNA can be present in the cell transiently or in a stably transformed state. The DNAs of this invention can further comprise a promoter to direct the transcription of the helper RNA and the replicon RNA, respectively, in the helper cell.

[0121] A further embodiment of this invention is a method of making infectious, multiplication-defective, coronavirus particles, comprising: a) providing the helper cells of this invention; and b) producing coronavirus particles in the helper cell. Optionally, the particles can be collected from the cells.

[0122] In certain embodiments, the helper cell can be provided by introducing the replicon RNA and/or the helper RNA into the helper cell by electroporation. However, the replicon RNA and/or helper RNA, as well as their respective DNAs can be introduced into the helper cell according to any methods known in the art for introducing nucleic acid into a cell. The nucleic acids can be present in the helper cell transiently or as stable transformants.

[0123] The present invention additionally provides infectious coronavirus particles produced by the methods of this invention.

[0124] The step of producing the infectious viral particles in the helper cells can also be carried out using conventional techniques. See e.g., U.S. Pat. No. 5,185,440 to Davis et al., PCT Publication No. WO 92/10578 to Bioption AB, and U.S. Pat. No. 4,650,764 to Temin et al. (although Temin et al. relates to retroviruses rather than coronaviruses). The infectious viral particles may be produced by standard cell culture growth techniques.

[0125] The step of collecting the infectious coronavirus particles can also be carried out using conventional techniques. For example, the infectious particles can be collected by cell lysis, or collection of the supernatant of the cell culture, as is known in the art. See e.g., U.S. Pat. No.

5,185,440 to Davis et al., PCT Publication No. WO 92/10578 to Bioption AB, and U.S. Pat. No. 4,650,764 to Temin et al. (although Temin et al. relates to retroviruses rather than coronaviruses). Other suitable techniques will be known to those skilled in the art. Optionally, the collected infectious coronavirus particles can be purified if desired. Suitable purification techniques are well known to those skilled in the art.

[0126] In some embodiments, VEE replicon vectors can be used to express coronavirus structural genes in producing combination vaccines. Dendritic cells, which are professional antigen-presenting cells and potent inducers of T-cell responses to viral antigens, are preferred targets of VEE and VEE replicon particle infection, while SARS coronavirus targets the mucosal surfaces of the respiratory and gastrointestinal tract. As the VEE and SARS replicon RNAs synergistically interact, two-vector vaccine systems are feasible that may result in increased immunogenicity when compared with either vector alone. Combination prime-boost vaccines (e.g., DNA immunization and vaccinia virus vectors) have dramatically enhanced the immune response (notably cellular responses) against target papillomavirus and lentivirus antigens compared to single-immunization regimens (Chen et al. (2000) *Vaccine* 18:2015-2022; Gonzalo et al. (1999) *Vaccine* 17:887-892; Hanke et al. (1998) *Vaccine* 16:439-445; Pancholi et al. (2000) *J. Infect. Dis.* 182:18-27). Using different recombinant viral vectors (influenza and vaccinia) to prime and boost may also synergistically enhance the immune response, sometimes by an order of magnitude or more (Gonzalo, et al. (1999) *Vaccine* 17:887-892). Thus, the present invention also provides methods of combining different recombinant viral vectors (e.g., VEE and SARS) in prime boost protocols.

[0127] The examples, which follow, are set forth to illustrate the present invention, and are not to be construed as limiting thereof. In the following examples, mM means millimolar, µg means microgram, ml means milliliter, µl means microliter, V means volt, µF means microfarad, cm means centimeter, h means hour, ORF means open reading frame, GFP means green fluorescent protein, PBS means phosphate-buffered saline, M means molar, s means second, nt means nucleotide, and min means minute.

## EXAMPLES

### Example 1

#### Production of Full-Length Infectious cDNA of SARS Coronavirus

[0128] Virus and Cells. The Urbani, Tor-2 and Tor-7 Canadian strains of SARS-CoV were propagated on VeroE6 cells in Eagle's MEM supplemented with 10% fetal calf serum and kanamycin (0.25 µg/ml) and gentamycin (0.05 µg/ml) at 37° C. in a humidified CO<sub>2</sub> incubator (Tor isolates were kindly provided by H. Feldmann). For virus growth, cultures of VeroE6 cells were infected at a multiplicity of infection (MOI) of 5 for 1 hr and samples were titered by plaque assay. At 1 hr postinfection, some cultures were treated with the cysteine protease inhibitor E64-d ({2S, 3S}-trans-epoxysuccinyl-L-leucylamido-3-methylbutane ethyl ester) at a concentration of 500 µg/ml. Virus plaques were visualized by neutral red staining at 2 days postinfection.

[0129] Strategy for Cloning the SARS-CoV cDNAs. Reverse transcription was performed using Superscript II™, oligodeoxynucleotide primers and intracellular RNA from SARS infected cultures [17, 18]. The cDNA was denatured for 2 min at 94° C. and amplified by PCR with Expand Long TAQ polymerase (Boehringer Mannheim Biochemical) for 25 cycles at 94° C. for 30 sec, 58° C. for 25-30 sec and 68° C. for 1-7 min. The amplicons were cloned into Topo II TA (Invitrogen) (SARS subclones D-F) or in pSMART vectors (Lucigen) (SARS subclones A-C). All cDNAs were assembled as consensus sequences based on independent sequence analysis of 4-7 sibling clones and the reported Urbani sequence [8]. The following primers were used in the isolation of the SARS A subclone (forward: tactaatagactcactatagatattaggtttttacctaccagg-1; reverse: acaccatagtcacgatgcc-4452), SARS B subclone (forward: gcctatatgcatggatgtagat-4359; reverse: tgaaccgccacgtggctaaacc-8727), SARS C subclone (forward: agccagcgtggcggttcacac-8710; reverse: aggcctcttggcagtgccataag-12,085), SARS D subclone (forward: actgcccaagatgcctatgagc-12,070; reverse: cagccaggaggcagacttcacaacc-18,939), SARS E subclone (forward: gtctgccctctggctgataagttccag-18,923; reverse: gagcagccgtgtaggcagcaat-24066) and SARS F subclone (forward: attgctgctacacggctgctc-24,045; reverse: (tt)<sub>7</sub>gtcattctcctaagaagc-29,710).

[0130] To repair sibling clones, primer pairs were designed that contained a Class IIS restriction enzyme (e.g., AarI). Using high fidelity PCR, the consensus portions of different sibling clones were amplified, digested with AarI and ligated into plasmid. The AarI junctions were designed to seamlessly link consensus fragments, resulting in the production of a full-length cDNA [17]. Using an automated ABI DNA sequencer, 2-3 candidate DNAs were sequenced to identify the consensus clone.

[0131] Systematic Assembly of a Full-Length SARS-CoV cDNA. The SARS A through F inserts were restricted, separated through 0.8% agarose gels, visualized with a Darkreader Lightbox (Claire Chemical), excised, and purified using the Qiaex II DNA purification kit. The SARS A+B, C+D and E+F subclones were ligated overnight and isolated [17, 18]. The SARS AB+CD+EF cDNAs were ligated overnight at 4° C., phenol/chloroform extracted and precipitated under isopropyl alcohol. Full-length transcripts were generated in vitro as described by the manufacturer (Ambion, mMessage mMachine) with certain modifications [17]. For SARS N transcripts, 1 µg of plasmid DNA encoding the N gene (primer: 5'nngcctcgatggccatttagtgacacatagatgtctgataatggacccaac-3' and reverse primer 5'-nnntttttttttttttttttttttatgcctgagtgatcagcag-3') was transcribed by SP6 RNA polymerase with a 2:1 ratio of cap analog to GTP.

[0132] Transfection of Full-Length Transcripts. RNA transcripts were added to 800 µl of the BHK cell suspension (8.0×10<sup>6</sup>) in an electroporation cuvette and three electrical pulses of 850 V at 25 µF were given with a Gene Pulser II electroporator (BioRad) [17, 18]. The transfected BHK cells were seeded with 1.0-2.0×10<sup>6</sup> uninfected Vero E6 cells in a 75 cm<sup>2</sup> flask and incubated at 37° C. for 2 days. Virus progeny were then passaged in Vero E6 cells at ~30 hr intervals and purified by plaque assay.

[0133] Detection of Marker Mutations Inserted in icSARS-CoV. Intracellular RNA was isolated from either



wild type or icSARS-CoV infected cells at 24 hrs postinfection. After RT-PCR, a 1668 nt amplicon (nt position 1007-2675) was obtained, spanning the BglI site at position 1572 that had been ablated in the icSARS-CoV component clones, but not wild type SARS-CoV. Other PCR products included a 799 nt amplicon spanning the SARS-CoV B/C junction (nt position 8381-9180), a 544 nt amplicon (nt position 11,721-12,265) spanning the SARS-CoV C/D junction, a 652 nt amplicon spanning the SARS-CoV D/E junction, and a 1594 nt amplicon (nt position 23,665-25,259) spanning the SARS-CoV E/F junction. The 1594 nt SARS E/F junction-containing amplicon was subcloned and sequenced.

**[0134]** RT-PCR of Leader-containing Transcripts. Leader containing amplicons were obtained from wild type and icSARS-CoV infected cells using primers at the 3' end of the genome (5'-ttttttttttttttttttgtcattctcctaagaagc-3') and in the SARS leader RNA sequence (5'-aaagccaaccaacctgcatc-3'). Leader-containing amplicons were excised from gels, subcloned into TopoII vectors and sequenced using appropriate primers.

**[0135]** Assembly of Coronavirus Full Length cDNAs. Rapid response and control of exigent emerging pathogens requires an approach to rapidly generate full-length infectious cDNAs that afford genetic control over the genome. Full-length infectious cDNAs were isolated for TGEV, HCoV-229E, IBV and MHV strain A59 [17-21]. The approach described herein has been to isolate a panel of cDNAs spanning the entire coronavirus genome, which can be systematically and directionally assembled into a genome-length cDNA by in vitro ligation [17, 18].

**[0136]** Patterned after the strategy devised for TGEV [18], the SARS genome was cloned by RT-PCR as six contiguous subclones linked by unique Bgl I restriction endonuclease sites. Bgl I is a class IIS restriction endonuclease that cleaves the symmetrical sequence GCCNNNN<sup>+</sup>NGGC, but leaves 64 different asymmetrical ends. Consequently, pairs of contiguous subclones encoded junctions that allow unidirectional assembly of intermediates into a full-length cDNA. Two Bgl I junctions were derived from sites encoded within the SARS-CoV genome at positions nt 4373 (A/B junction) and nt 12,065 (C/D junction) [8-10]. A third Bgl I site at position nt 1557 was removed and new Bgl I sites were inserted by the introduction of silent mutations into the SARS-CoV sequence at positions nt 8700 (B/C junction), nt 18,916 (D/E junction) and at nt 24,040 (E/F junction). These mutations are shown in the Sequence Listing as SEQ ID NOs:10-14. SARS-CoV sequence toxicity was circumvented by disruption of toxic domains and the use of stable cloning vectors [17]. The resulting cDNAs include SARS A (nt 1-4436), SARS B (nt 4344-8712), SARS C (nt 8695-12,070), SARS D (nt 12,055-18,924), SARS E (nt 18,907-24,051) and SARS F (nt 24,030-29,736) subclones. The SARS A subclone contains a T7 promoter and the SARS F subclone terminates in 21 Ts, allowing for in vitro transcription of capped, polyadenylated transcripts.

**[0137]** Numerous mutations were noted in each of the four to seven sibling subclones encoding a given SARS cDNA. To rapidly assemble a consensus clone, the "No See'm" approach was used with another group of class IIS restriction endonucleases that cut at asymmetric sites and leave asymmetric ends. These enzymes cut strand specific and

allow the seamless ligation of cDNAs with the loss of the restriction site used to join two component clones [17]. As illustrated with the SARS F sibling clones, primer pairs were designed that contained terminal Aar I (CACCTGCNNNN<sup>+</sup>NNNN) sites that flanked each of the various consensus portions of different sibling clones. In some instances (amplicons 3 and 2 in sibling clones 1 and 4, respectively), primers were also designed to repair specific mutations located near the ends of a given amplicon. The combination of high fidelity PCR, oligonucleotide primer repair and the seamless ligation of sequence fragments [17], rapidly generated Urbani consensus cDNAs for each of the SARS A, B, C, D, E and F subclones. Silent changes retained in the full length construct included an A to G change at nt position 6460, a T to C change at nt position 14,178, a T to C change at nt position 15,740, a C to T change at nt position 19,814, an A to G change at nt position 20,528 and a T to C change at nt position 20,555. These mutations are shown in the Sequence Listing as SEQ ID NOs:15-21.

**[0138]** Rescue of Molecularly Cloned SARS-CoV. To build full length SARS-CoV cDNAs, individual subclones were digested with the appropriate restriction enzymes, ligated together in vitro and used as template for in vitro transcription with the T7 RNA polymerase. Since N transcripts enhance RNA transfection efficiencies of TGEV and MHV [17, 22], and are essential for IBV transcript infectivity [20], SARS-CoV full-length transcripts were either electroporated into cells alone or mixed with SARS-CoV N transcripts. Within 48 hrs post transfection, SARS-CoV infected cells were detected by fluorescent antibody staining. Infectious virus (icSARS-CoV) titers approached 1.0x 10<sup>6</sup> PFU/ml at 48 hrs postinfection in the mixed transcript transfected cultures. Recombinant viruses were also detected in cultures transfected with genome-length SARS transcripts alone, but titers were reduced. As described with MHV [17], SARS N transcripts may enhance infectivity of SARS full-length transcripts, but are not essential. The molecular cloned viruses were plaque purified in VeroE6 cells and produced similar sized plaques as wild type SARS Urbani.

**[0139]** icSARS CoV Marker Mutations. Rescued icSARS-CoV, but not wild type SARS-CoV should contain several Bgl I sites that were engineered as junctions between the SARS B/C, D/E and E/F subclones and lack the Bgl I site at nt position 1557. Intracellular RNA was isolated from infected cultures, RT-PCR amplified using primer pairs flanking these various sites and subjected to restriction fragment length polymorphism (RFLP) analysis with Bgl I. Clearly, icSARS-CoV contained the marker mutations inserted within and between the component clones. To confirm these findings, selected amplicons were cloned and sequenced, demonstrating that the icSARS-CoV originated from transcripts derived from the full-length cDNA construct.

**[0140]** Phenotype of Rescued icSARS-CoV. Cultures of cells were infected with an MOI of 5. In untreated cultures, intracellular RNA and virus titers were determined as described herein. At 1 hr, cultures were treated with E64-d at a concentration of 500 µg/ml and virus titers determined by plaque assay in VeroE6 cells.

**[0141]** The recombinant icSARS CoV isolate replicated as efficiently as wild type Urbani, but less efficiently than the

Canadian isolates, Tor-2 and Tor-7 SARS-CoV. These data demonstrate that the introduced alterations were not debilitating to virus growth in culture as both replicated to titers of about  $10^7$  within ~24-48 hrs postinfection. To further characterize the transcription strategy of SARS-CoV, intracellular RNA was isolated from Urbani wild type and icSARS-CoV-infected cultures. Following RT-PCR amplification of leader containing amplicons, sequence analysis indicated that wild type and icSARS-CoV subgenomic transcripts originated at identical CS sites, defined by the core sequence ACGAAC, as illustrated by leader-containing transcripts encoding X1, E, M, X3 and X4. SARS S and N encoding transcripts initiated subgenomic transcription at similar CS sites. This sequence represents a truncation of the AAACGAAC CS site that had been predicted by Rota et al., 2003 and is different from the group I, II and III coronavirus CS sequences, CUAAAC (TGEV), UCTAAAC (MHV), and CUUAACAA (IBV), respectively. Although previous studies had suggested that the SARS E protein and ORF X3 might be expressed from polycistronic mRNA, these findings indicate that independent transcripts are initiated at the core CS sequence ACGAAC noted at nt positions 26,109 for E transcripts and 26,913 for ORF X3 transcripts, respectively.

**[0142]** In vitro inhibition of SARS-CoV Replication. Given the high virulence of SARS-CoV infection in humans, antiviral drugs will be required to reduce the scope and severity of disease. In MHV, the cysteine proteinase inhibitor E64-d blocks replicase polyprotein processing and thereby inhibits viral RNA synthesis and virus growth [11]. To determine if icSARS-CoV was susceptible to the inhibitory effects of E64-d, growth analysis was performed in the presence and absence of 500  $\mu\text{g/ml}$  of drug. In the absence of E64-d, wild type and icSARS-CoV grew to equivalent titers of  $\sim 1.0 \times 10^7$  PFU/ml after 24-48 hrs postinfection. Treatment of cells with a single dose of E64-d at 1 hr pi resulted in almost complete elimination of viral CPE and viral antigen expression and a significant ~3-4 log reduction in virus yield for both wild type SARS-CoV and icSARS-CoV at 24 and 48 hrs postinfection.

#### Example 2

##### Development of SARS Virus Single Hit Replicon Vector Vaccines

**[0143]** SARS Replicon Particle Assembly. Viral replicon particles are single hit vectors that are incapable of spreading beyond the site of initial infection and are novel vaccine candidates for human and animal diseases. The SARS single hit replicon particles will be produced in the helper cells as described herein. Specifically, the SARS E, M, S and/or N structural proteins will be produced in helper cells from helper RNA or DNA, e.g., VEE VRPs, noncytotoxic Sindbis virus replicons, BAC or vaccinia vectors, CMV eukaryotic expression plasmids, etc., using standard protocols in order to provide SARS coronavirus structural proteins in trans for assembly of coronavirus replicon particles in helper cells. The SARS E, M, S and/or N coding sequences will be deleted from the replicon RNA and one or more heterologous coding sequences will be inserted into the replicon RNA. The replicon RNA will be packaged into coronavirus particles in the helper cells. Populations of the particles produced in the helper cells as described herein can be tested

for the ability to maintain only a single round of infection by passage on coronavirus-permissive cells.

**[0144]** Cell Preparation: VERO E6 Cells Expressing SARS E protein or M glycoprotein. Sindbis noncytotoxic replicon vectors (pSinRep) induce persistent infections in VERO and BHK cells, and constitutively express resistance to puromycin and a foreign gene of interest. These vectors have been used to constitutively express human CEA receptors in BHK cells, converting these cells to susceptible hosts for the MHV host range mutant V51. SARS E and M coding sequences will be inserted into pSinRep or CMV expression vectors, VERO E6 cells will be transfected and selected with puromycin (or G-418 for CMV vectors) for clones that stably express high concentrations of selected SARS structural proteins. The SARS E protein and M glycoprotein will also be expressed from VEE VRPs, and inoculated into mice for the production of monospecific antisera against each of these proteins. VEE VRPs have been used to produce high titer antiserum against Norovirus antigens, as well as the TGEV M and N proteins. Using flow cytometry and polyspecific antisera directed against the SARS E protein and M glycoprotein, enrichment for cells expressing high or low levels of the desired protein will be carried out. As VERO E6 cells are susceptible hosts for SARS replication, these cells will allow for additional rounds of infection with SARS VRPs with subsequent packaging and release of progeny VRPs.

**[0145]** SARS Vector Replicon Design: First, a determination will be carried out regarding whether any of the group specific ORFs (X1-X5) can be deleted and replaced with either luciferase or GFP. Specifically, the focus will be on ORFs X1, X4 and X5 because of their position in the genome and the possibility that they encode luxury functions that are not essential for replication in vitro. The overall goal is to identify a group specific ORF that can be deleted and replaced with an indicator gene that allows for easy visualization of replication and gene expression. Such mutants will also be of value for drug screening. Essential points in these experiments are to leave the CS and surrounding flanking sequences intact, so that foreign gene expression is maximized. A second issue is that foreign genes may encode noncanonical CS sites that are transcriptionally active. To circumvent this problem, noncanonical "TAAACGAAC" CS sites in GFP or luciferase will be altered at the 3<sup>rd</sup> codon to prevent spurious expression of cryptic subgenomic RNAs. Growth curves will be performed to determine if rescued viruses replicate at equivalent levels with wild type SARS coronavirus and cultures will be examined for GFP or luciferase expression. The most viable of the ORFX1-X3 constructs will be used for future studies.

**[0146]** SARS replicon genomes will be assembled by systematically deleting the E and/or M genes using the No See'm Strategy described in Yount et al. (2002). These constructs will be built within a backbone that encodes the luciferase gene as described above. In the case of M, the M ORF will be deleted and GFP or luciferase will be inserted, leaving the M CS site intact. In the latter two instances, this approach will allow for coordinated and equivalent levels of expression of upstream genes, while inserting a heterologous foreign nucleic acid for easy quantitation of VRP titers.

**[0147]** SARS replicon constructs will be assembled as described, except that Not 1 sites will be encrypted at the 5'

end of Clone A and 3' end of Clone F. Following assembly, the DNA will be restricted with Not I and inserted into BAC vectors. This allows for BAC propagation in bacteria and circumvents tedious reconstructions of SARS replicon constructs. The stability of these BAC vectors will be evaluated by serial passage and sequence analysis.

**[0148]** Stable VRP Single Hit Expression Vectors. Successful development of stable single hit SARS replicon vectors will be demonstrated by standard approaches. To titer VRP stocks, quantitative methods have been developed that rely on serial dilution and counting cells expressing nucleic acid encoding GFP or viral and heterologous antigens by fluorescent antibody staining in a number of randomly chosen but defined ( $\mu\text{m}^2$ ) fields. Titers are estimated by counting the number of fluorescent cells in a known area, statistically determining the mean and then adjusting for total area of the plate. Alternatively, viral VRP stocks can be quantified by endpoint PCR or quantitative PCR, potentially compromised by an unknown particle/PFU ratio. It is expected that the SARS M glycoprotein and E protein will be essential for assembly and release, based upon experience with TGEV. However, it has been suggested that SARS is more closely akin to the group II coronaviruses like MHV, so it is also possible that SARS E protein may be nonessential for packaging and release.

**[0149]** A number of VRP phenotypes will be analyzed. The SARS VRPs lacking E protein should be single hit vectors in VERO E6 cells, but should serially passage in VERO E6 cells expressing E in trans. Similarly, SARS VRPs lacking M glycoprotein should be packaged and released from VERO E6 cells expressing M, but not E. SARS VRP protein expression will be determined by Western Blot assays, to determine if E or M excision alters expression levels of other SARS structural and nonstructural ORFs as compared with wild type. This will be accomplished with antiserum generated from the VEE VRPs inoculated into mice. GFP or luciferase expression will be measured by fluorescence and Western Blot assays, using commercially available antibodies and screens. The ability of different packaging cell lines (e.g., alphavirus vectored, CMV vectored) to support the production of high titered SARS VRPs will be compared. RNA recombination is always a concern with a vectored packaging cell lines, so progeny VRPs will be isolated and passaged several fold in a packaging cell line and passaged onto control cells lacking the appropriate structural protein in trans.

**[0150]** SARS-CoV Recombinant Protein Assays. Venezuelan equine encephalitis virus (VEE) is a member of the alphavirus genus of the family *Togaviridae*. The virus consists of an icosahedral nucleocapsid composed of 240 copies of the capsid protein (C) surrounded by an envelope containing spikes formed from two glycoproteins, E1 and E2. The VEE genome is a positive sense single stranded RNA of 11.5 kb, which replicates through a minus strand intermediate. A subgenomic 26S mRNA, transcribed from the 26S promoter resident internally on the minus strand, encodes the structural proteins and is synthesized at  $\sim 10\times$  molar excess relative to genome. Four nonstructural replicase proteins (nsp1-4) mediate all transcription steps, which occur in the cell cytoplasm, and virus budding is at the plasma membrane.

**[0151]** The VEE genome, when introduced into a cell, can be viewed as a highly efficient machine for the production of

large amounts of its own structural proteins through the transcription of a subgenomic mRNA encoding these proteins. If a nucleic acid of interest is substituted for the structural protein genes, then the VEE replicative machine expresses high levels of that gene product. In the VEE replicon vaccine vectors, an immunizing gene is substituted for the structural protein genes. Upon electroporation of replicon RNA into cells, the VEE replication machinery produces high levels of the subgenomic mRNA and vectored gene product. However, as the structural protein genes are not present in the replicon RNA, no new virus particles are released. One can package the replicon into "replicon particles" by supplying the structural proteins in trans from helper RNAs. The replicase proteins encoded on the replicon RNA provide the machinery for replicating the helper RNAs, while the structural proteins encoded by the helpers encapsidate and envelope the replicon RNA. Only the replicon RNA is packaged into VEE replicon particles (VRP), because only the replicon RNA contains the cis-acting packaging signal. When VRPs infect another cell, either in culture or in vivo, the replicon RNA synthesizes high levels of the gene product, but no new replicon particles are formed due to the absence of the structural protein genes. Therefore, these vectors are not cell to cell transmitted.

**[0152]** Alphavirus Vectored Expression of Recombinant SARS-CoV Proteins. The SARS S, E, N, M, and the group specific ORFs were cloned into VEE and packaged into VRPs. Following infection with VRP-S and VRP-N, cultures were radiolabeled with  $^{35}\text{S}$ -Met from 12-16 hrs post infection, resulting in the expression of a SARS  $\sim 46$  kDa N protein and a  $\sim 177/205$  kDa S glycoprotein. VRP-E and VRP-M also expressed appropriately sized products using these conditions. Various VRP-group specific ORFs have also been inoculated into mice and cells. These data suggest 1) VRP immunization induced significant antibodies to authentic S protein expressed from three different SARS isolates or with the rescued molecularly cloned icSARS-CoV (Similar results have been obtained with VRP-E, VRP-M and VRP-N antiserum), 2) patient serum was reactive with the S protein expressed by the VRPs and by icSARS-CoV, and 3) SARS ORF3a is immunogenic and detected by convalescent patient serum. The SARS-S gene, designed to produce either membrane-bound or secreted forms, is also available and antiserum against the group specific ORFs is being raised in mice. SARS-CoV PRNT<sub>90</sub> titers were  $\sim 1:905$  with VRP-S vs.  $<1:80$  with VRP-HA (flu HA).

**[0153]** VEE replicon particles (VRPs) expressing the SARS S protein have been constructed and antigen derived from cells infected with these vectors was initially evaluated as a target for EIA assays. Vero cells were infected with VRP-SARS(S) at a MOI of 6.0 for 15 hours and the cells were lysed with 0.5% NP-40. The optimal antigen concentration was found to be a 1:300 dilution. Using this assay, the human convalescent serum was shown to have an endpoint (the last dilution with an O.D. reading  $\geq 0.200$ ) anti-SARS virus IgG titer of 1:3200, while serum from two uninfected control individuals had titers of  $<1:200$ . Sera from mice that received a primary inoculation of VRP-SARS(S) in the rear footpads, followed by a boost at week 5 post primary inoculation, were evaluated at two weeks post boost and ranged from 1:1600 to 1:6400 (ave.=3300, n=12), while control animals that received VRP expressing the influenza virus A/PR/8 HA gene had titers of 1:100-1:200 (ave.=150,

n=2). Similar assays were developed for other SARS structural protein, providing a unique panel of reagents to measure immune responses following SARS-CoV infection in rodents.

[0154] To perform CTL assays and other immunologic experiments, the SARS S and N proteins were inserted into vaccinia virus vectors.

### Example 3

#### Deletion of SARS Virus Group Specific ORFs

[0155] The SARS virus genome encodes several group specific ORFs at the 3' end of the genome including ORFs 3a/b (X1, X2), 6 (X3), 7a/b (X4), 8a/b (X5) and 9b. By analogy with other coronaviruses, it is believed that one or more of these group specific ORFs encode luxury functions that do not specifically contribute to virus replication in vitro but disable virus pathogenesis in vivo. This invention includes the systematic and combined deletion or knocking out of expression of these various ORFs and rescue of the molecularly cloned viruses containing these genetic lesions. All recombinant viruses are obtained following transfection in Vero cells similar to those certified for cultivation and production of human virus vaccines.

[0156] Protocols are carried out wherein the group specific nucleic acid sequences designated ORFs 3a/b, 6, 7a/b, 8 and 9a are deleted. Initially, each of the ORFs is systematically deleted, either by deletion of individual ORFs (ORF3a-X1, ORF6-X3, ORF7a, ORF7b, ORF8a, ORF8b), entire TRS cistrons (e.g., ORF3a/b; ORF7a/b-X4; ORF8a/b-X5), or by mutations that specifically ablate ATG start codons and introduce premature stop codons in an overlapping gene (ORF3b-X2, ORF9b). For deletions, a particular gene and its corresponding CS site (when appropriate) is excised, and the flanking upstream and downstream ORFs seamlessly stitched together. For example, two amplicons are isolated using primer pairs that are designed to contain external Bsa1 sites that are lost but leave complementary four nucleotide asymmetric sites that direct seamless ligation of the SARS CoV ORF 6 and ORF 8 CS sites, simultaneously deleting the entire ORF 7a/b sequence. The ATG start is knocked out and a premature stop is inserted into ORF3a (X1) and ORF6 (X3) because virus phenotypes are not confounded by changes in transcription.

[0157] SARS-CoV from civet cats and human patients early in the epidemic contained a 29 bp insertion in ORF8a/b that resulted in a single ORF 8 protein. As it has been suggested that this deletion may have enhanced SARS-CoV pathogenicity in humans and perhaps transmission, the Civet cat full length ORF8 by fusing ORF8a and 8b will be reconstructed by insertion of the 29 bp sequence identified in this virus, with the expectation that the addition of a full-length civet cat ORF8 will attenuate pathogenesis.

[0158] In mRNA3 transcripts encoding ORF 3a, ORF 3b is encoded as an overlapping out of frame 124 amino acid ORF that cannot be deleted without removing a significant portion of ORF 3a. If both ORFs encode luxury functions in vitro, recombinant viruses lacking each ORF will be rescued separately. While deletion of ORF 3a is straightforward, ORF 3b also contains ten in frame ATG start codons, thus selective mutagenesis of ORF3b start codons will only N-terminally truncate ORF X3b (if internal ATGs translate

efficiently). Consequently, premature ORF3b stop codons will be introduced into the sequence: <sup>425735</sup>CAGTGTCA-CAGAT" to "CAGTGTGACAGAT." This alteration introduces a stop codon that truncates ORF3b after 24 amino acids, but doesn't change the protein sequence of ORF3a. A global approach is to combine this termination codon with additional mutations that destroy the ORF3b ATG start codon at residues 1 and 2, but leave the ORF 3a amino acid sequence intact. Changes will be introduced into the SARS CoV F subclone, which will be re-sequenced and then reassembled into full-length cDNAs. A similar strategy will be used to knockout ORF9b in mRNA9, which also encodes the N protein (ORF9a). Following in vitro transcription, full-length transcripts will be mixed with N transcripts (to enhance infectivity), introduced into cells by electroporation. Rescued viruses will be quantified by plaque assay.

[0159] Deletion of SARS-CoV Group Specific Genes. The SARS-CoV ORF3a (X1), ORF 6 (X3) and ORF 7a/b (X4) domains have been deleted and recombinant viruses have been isolated that replicated in Vero cells. Moreover, ORF7a/b has been replaced with luciferase under control of the 7a/b TRS and is expressed in icSARS-CoVΔX4 infected cells. Following transfection, leader containing transcripts were identified, and ORF3a/ORF6/ORF 7a/b deletions were confirmed by RT-PCR and sequencing. Passage of progeny virions revealed abundant viral protein expression in icSARS-CoV, icSARS-CoV ΔX3 and ΔX4-infected cultures, although protein expression in icSARS-CoV ΔX1 infected cultures was less robust. At this time it is not possible to determine if this reflected a lower initial transfection efficiency or lower rates of virus replication. The SARS-CoV X1 and X3 ORFs encode interferon antagonist genes. It seems likely that virus growth might be boosted in certain cell types by providing IFN antagonist genes in trans, should the ΔX1 molecularly cloned variant be highly attenuated in replication. Based on findings with other highly pathogenic viruses, the deletion of IFN antagonists should attenuate virus replication and pathogenesis in vitro. More importantly, the data demonstrate the feasibility of using 1) the SARS-CoV cDNA reverse genetics to modify the SARS-CoV genome, 2) that several group specific genes of SARS-CoV encode luxury functions for in vitro growth (ORF3a, ORF3b, ORF6 and ORF7a and ORF7b) and 3) it can be determined if deletion of SARS-CoV IFN antagonist genes and group specific genes attenuate pathogenesis in vivo in the mouse, ferret and eventually primate model for SARS-CoV pathogenesis. Deletion of ORF3a/b (X1 and X2 protein) also results in the rescue of viable progeny viruses, supporting the notion that deletion of multiple accessory (group specific) ORFs still allow for the recovery of robust viruses. All deletion mutants tested replicate to titers in excess of 10<sup>7</sup> PFU/ml in 32 hrs. We have also produced a SARS-CoV isolate expressing GFP from the ORF7a/b location. Recombinant viruses encoding either the Luciferase or GFP indicator molecules provide a rapid screen for identifying compounds that block SARS-CoV replication and gene expression. Reduced expression of either luciferase or GFP in the presence of drug allows for rapid identification of compounds with antiviral activity.

[0160] SARS-CoV Minimal Genome Recombinants/Multiple Accessory Gene Knockouts. In addition to deleting each of the group specific ORFs, the nonessential ORFs can be deleted in combination. The goal is to produce a minimal SARS genome lacking as many of the group specific ORFs

as possible (SARS  $\Delta$ 3a/b, 6, 7a/b, 8a/b), while retaining adequate levels of virus replication in vitro. In the case of MHV, deletion of group specific ORFs significantly attenuated MHV pathogenesis in mice and recombinants lacking combinations of group specific genes were generally more attenuated than isolates lacking a single group specific ORF, but not always. No experiments were conducted to determine if mutant-infected animals were protected against wild type challenge.

**[0161]** All TGEV and MHV group specific ORFs were nonessential for in vitro replication and at least ORF3a (X1), ORF3a/3b (X1/X2), ORF6 (X3) and ORF7a/b (X4) are nonessential for SARS-CoV replication. Thus, most, if not all, of the SARS virus group specific ORFs are believed to be capable of being deleted, resulting in a minimal genome that should be attenuated in vivo. Following mutagenesis of the SARS F subclone, assembly of the cDNAs and electroporation of transcripts, recombinants will be rescued from transfected cultures.

**[0162]** Other attenuating mutations. Mutations can be inserted virtually anywhere in the SARS-CoV genome including the replicase, the replicase ribosome frame shifting site, 5' and 3' transcriptional regulatory sequences, and alterations in the SARS-CoV structural genes as examples.

**[0163]** Characterization of Rescued Viruses. Rescued viruses will be harvested between 24 to 36 hrs post transfection and plaque purified prior to isolating low passage stocks in Vero E6 cells. During production of the recombinant virus stocks, passage numbers will be kept to a minimum so that potential 2<sup>nd</sup> site compensatory changes that restore virus replication and fitness do not have time to evolve and severely complicate an accurate interpretation of the effects of particular changes on SARS replication fitness.

**[0164]** Genotype Analysis. Plaque purified molecularly cloned viruses will be inoculated into Vero E6 cells and intracellular RNA isolated using Trizol reagents. Using RT-PCR and primer-pairs that flank the various group specific ORFs, the presence or absence of group specific ORFs will be determined in recombinant viruses by size analysis of PCR product and by sequence analysis of amplicons. The S gene through the 3' end of the genome will also be cloned and sequenced to identify potential 2<sup>nd</sup> site alterations that may have evolved in response to group specific gene deletion. As the gene order has been restructured by deleting various group specific ORFs, it is possible that these rearrangements may restructure TRS networks and the function of individual CS sites involved in the synthesis of a particular subgenomic transcript (most likely a gene upstream or downstream of the deleted ORF). To determine if subgenomic RNA synthesis originated from the appropriate CS sites, leader containing amplicons that flank the various group specific ORF deletions will be RT-PCR cloned and sequenced.

**[0165]** Phenotype Analysis. There are no precise in vitro correlates to in vivo SARS pathogenesis, so a variety of phenotypic characteristics will be measured to provide a global view of rescued-virus fitness in cell culture. Virus growth will be evaluated in culture, as well as plaque morphology. Rescued molecularly cloned viruses will be analyzed for SARS protein expression by immunoprecipitation or Western blots using antiserum from mice inoculated with alphavirus VRPs encoding the various SARS

virus structural and group specific ORFs. By immunoblot, the expression of the structural and various group specific nonstructural proteins that are produced following gene deletion will be measured. Northern Blot analysis will be used to identify the viral mRNAs and determine if selected gene deletion(s) alter the molar ratio of the upstream and downstream viral-mRNAs. In selected instances, cultures will be radiolabeled with <sup>3</sup>H-uridine and the labeling kinetics of viral mRNAs and replicative forms RNAs analyzed. These experiments will also determine 1) deletion effects on RNA expression patterns, and 2) deletion effects on relative molar ratios of mRNA. These experiments will also determine if full length and subgenomic length replicative form RNAs are present in SARS virus infected cells, as shown with other group I and II coronaviruses.

**[0166]** Stability of Rescued Virus with Passage: Compensatory Evolution. TGEV deletion and gene rearranged viruses have been shown to rapidly evolve 2<sup>nd</sup> site compensatory changes that restore virus growth fitness in vitro. It is believed that the 2<sup>nd</sup> site changes will subtly alter coronavirus gene networks and protein-protein interactions to restore virus growth in vitro, yet enhance attenuation in animal models by subtly changing the affinity of these highly orchestrated interactions that influence replication in the human host. To test this hypothesis, the most debilitated rescued viruses that have deletions in one or more group specific ORFs will be tested. Recombinant viruses will be inoculated onto Vero E6 cells and progeny viruses passaged into fresh cultures at 24 hrs intervals. After 15-20 passages, progeny virus will be plaque purified and compared to wild type virus by growth curves and gene expression patterns. The corresponding SARS E and F genome fragments will be cloned and sequenced. Consensus SARS E and F subclones will be assembled into the full length cDNA to identify the exact 2<sup>nd</sup> site mutations that restore replication fitness. In the case of multiple alterations, the mutation that confers the strongest fitness recovery phenotype will be tracked. Although replicase mutations may also restore growth, preliminary data suggests that most of the compensatory changes will cluster at the 3' end of the genome.

**[0167]** Rewiring the Coronavirus Genome. Mutation in the SARS Transcription Regulatory Sequence Attenuates Replication and Recombination. Live virus vaccine development provides an approach for identifying virulence alleles and pathogenic determinants, providing a template our understanding viral pathogenesis. Live viruses that lack several of the SARS-CoV group specific ORFs have been rescued, and are being tested for whether such viruses are attenuated.

**[0168]** The coronavirus TRS includes the highly conserved CS and flanking sequences, which regulate the efficiency of coronavirus transcription. It is believed that 1) mutations in the leader and body CS will attenuate SARS-CoV gene expression, and 2) SARS-CoV CS regulatory networks can be rewired, making these rescued viruses highly resistant to RNA recombination repair at the 3' end of the genome. The goal is to develop viruses that have significantly different TRS regulatory networks that upon recombination create incompatibility networks that disrupt expression of subgenomic mRNAs. These "rewired" SARS-CoV will be highly resistant to recombination repair from wild type viruses and other coronaviruses. These experiments will enhance the feasibility of safe SARS-CoV attenu-

ated seed stocks for killed vaccines and serve as a model for engineering recombination resistant viruses.

**[0169]** Leader/body CS Compensatory Mutations. Previous studies with arteriviruses and TGEV indicated that mutations in the leader CS globally suppress transcription of all subgenomic mRNAs, unless the corresponding mutations are also duplicated into the body CS. Rewiring the Nidovirus transcription regulatory network provides a novel approach to selectively attenuate gene expression of certain genes while maintaining efficient expression of other genes. The current hypothesis is that the actual CS sequence is not so critical, the major factor being that CS sites must allow for communication via efficient base-pairing for discontinuous transcription of subgenomic RNAs. It is believed that double and triple compensatory mutations in the leader/body CS will allow for viable viruses with robust gene expression, but be highly resistant to recombination repair.

**[0170]** Selected mutations (CS mutations 1 and 2) will be inserted into the 5' leader CS and body CS of the N structural gene (Tables 1 and 2), in essence producing SARS-CoV single hit replicon RNAs as described herein. N CS sites are targeted initially, as N subgenomic transcripts are closest to the 3' end of the genome and expressed most abundantly, providing a reliable signal for detection. N transcript expression from single hit TGEV replicons has been detected under similar conditions. The CS mutational spectra were chosen as: 1) the sequence is unique in the SARS-CoV genome and 2) it differs from CS sites present in other group I-III coronaviruses. Cultures will be transfected with recombinant virus RNA genomes and subgenomic transcription will be measured by quantitative RT-PCR using primers in the leader sequence and the body sequence of several structural genes, including N and genomic RNA. It is expected that robust N transcript expression will be demonstrated, with little if any expression of other subgenomic mRNAs. It may prove necessary to test other potential leader/body CS sequences that effectively optimize subgenomic expression of N, but not the other SARS-CoV transcripts. This will be tested empirically should the initial mutations prove unsatisfactory for driving mRNA expression (as determined by the relative ratio of genome to mRNA expression).

**[0171]** Using the most robust CS sites identified in the initial experiment, the remainder of the SARS-CoV CS sites will be converted to the new consensus sequence. Recombinant viruses (csSARS-CoV) will be assembled and transcripts electroporated into cells. Such viruses will be viable in culture and should express normal ratios of mRNA, RF RNA and protein.

**[0172]** At 24-36 hrs post transfection, rescued viruses will be plaque purified and stocks grown in Vero E6 cells as previously described. Virus passage will be minimized to prevent the emergence of 2<sup>nd</sup> site compensatory changes that restore virus growth fitness in culture. If recombinant viruses grow inefficiently, virus will be passaged 15-20x and mutants with increased growth kinetics and gene expression patterns will be identified by comparison with wild type (e.g., growth curves, structural protein expression, and northern blots). Given that reversion to wild type sequence in the leader/body CS sites is unlikely because of the requirement for multiple mutations (~18-27 changes in total genome CS sites), revertants will likely contain compensatory changes that reside in replicase proteins that interact

with SARS-CoV TRS sites or in the flanking TRS sites that regulate discontinuous transcription (less likely as several would be required). These will be identified by sequence analysis followed by reintroduction of specific mutations into the full length cDNA construct.

**[0173]** Genotype Analysis. Plaque purified molecularly cloned viruses will be inoculated onto Vero E6 cells and intracellular RNA isolated using Trizol reagents. Using RT-PCR and primer-pairs that flank selected leader and body TRS mutations, amplicons will be sequenced to confirm that rescued viruses contain the expected alterations. Using RT-PCR, leader-containing amplicons representing each of the downstream transcripts that are synthesized during infection will also be cloned and sequenced. However, it is possible that altered CS mutants might initiate subgenomic transcripts from noncanonical sites. Any such aberrant leader-containing amplicons will be identified by sequencing. Depending upon the level of expression, the aberrant CS sequence in newly emerged TRS sites will be mutated to knock out indiscriminant subgenomic transcription.

**[0174]** Phenotype Analysis. Virus growth will be evaluated in culture and by plaque morphology. Rescued molecularly cloned viruses will be analyzed for SARS protein expression by Western Blot using antiserum from mice inoculated with alphavirus replicon particles encoding the various SARS structural and selected group specific ORFs. The ratios of structural proteins that are expressed in the rescued viruses will be determined by Western blot. Northern blot analysis will be used to identify the viral mRNAs and determine if selected TRS mutations alter the molar ratio of the upstream and downstream subgenomic viral mRNAs or the emergence of additional transcripts originating from newly recognized body CS sites. In selected instances, cultures will be radiolabeled with <sup>3</sup>H-uridine and the labeling kinetics of viral mRNAs and replicative form RNAs will be analyzed.

**[0175]** New CS sites will be demonstrated to be stable in the SARS-CoV genome and allow for efficient gene expression and virus growth. The stability of selected mutants will be determined by serial passage in Vero E6 cells with particular focus on: 1) stability of leader/body CS sites and the evolution of group specific gene expression, and 2) mechanisms of fitness recovery. Putative compensatory mutations that restore virus replication fitness will be reintroduced into the full-length construct to prove causality between particular alterations, growth recovery and enhanced gene expression.

**[0176]** RNA Recombination with wild type SARS-CoV. Several approaches will be used to test whether "re-networked" SARS-CoV can recombine efficiently with wild type SARS-CoV. In one approach, select wild type CS sites will be introduced back into the SARS-CoV re-networked genome. For example, the M CS site will be altered back to "ACGAAC" and a determination will be made regarding whether viable recombinant viruses can be isolated. As the M glycoprotein is essential for growth, this is a strong selective screen for viability. It is anticipated that such viruses will grow poorly and require passage for selection of compensatory changes that reestablish growth. Such changes would likely evolve within the CS site, given only 1 or 2 mutations might be required to restore functionality of the body CS. In a second approach, a theoretical single

crossover recombinant virus that contains the “mutated” leader CS will be coupled with all 3' end CS sites derived from the wild type virus. Reflecting what would happen if a recombination event occurred in the replicase of a “re-networked” virus and wild type virus, it will be determined if the engineered leader CS site will communicate and drive subgenomic transcripts from a wild type 3' end genome. It is anticipated that such viruses would be heavily impaired and that subgenomic transcription might occur from non-canonical sites. Subgenomic transcription will be measured by quantitative RT-PCR using primer sets in the leader sequence and in various structural genes. Transfected cultures and supernatants will be passaged (15-20x) to allow for the emergence of revertants that will be analyzed as described herein. Reversion, if it occurs, would likely occur through the emergence of one or more mutations in the leader CS. In a third approach, as in the second approach, a similar experiment will be performed, but in this instance, the 3' end of the re-networked SARS-CoV genome will be replaced with the 3' end of MHV (S through N). In one construct, the 3'-most ~500 nucleotides of the SARS-CoV genome that might contain cis-acting sequences needed for replication will be maintained and in a second construct, the entire SARS-CoV N gene and CS site will be included. In essence, a potential double recombinant genome will be generated that might occur between another group II coronavirus and SARS-CoV. Such viruses are not anticipated to be viable because of communication problems between the SARS-CoV replicase and the leader/body CS sites and potential encapsidation sequence problems associated with an MHV N protein (construct 1). In the second construct, the N protein will be derived from SARS-CoV but it must communicate with the MHV M, E and S proteins to mediate assembly of infectious virus, which is not likely.

[0177] Other rewiring approaches to prevent recombination repair of live attenuated viruses: Other approaches to produce recombination resistant viruses include: 1) repositioning the SARS-CoV encapsidation sequence to different locations in the genome, 2) rewiring the replicase protein cleavage sites to a different recognition sequence, 3) reorganizing the SARS-CoV gene order and 4) any other methods that establish genome incompatibilities following recombination with any other circulating human strain.

#### Example 4

##### SARS Virus Gene Order Rearrangement

[0178] Because N gene rearrangements are well tolerated in the group I and II coronavirus genomes, in initial studies, the N gene position will be moved to new sites immediately downstream (SARS SNEM) and upstream of S (SARS NSEM). Second generation constructs will include rearrangement of ORF7 or 8, the N gene upstream of S (SARS 7/8NSEM6) and an E protein/M glycoprotein reposition upstream of S as well (SARS EMSN). The SARS S glycoprotein will also be repositioned to the 3' proximal location (SARS EMNS). The putative IFN antagonist genes encoded in ORF3a and ORF6 will also be repositioned, potentially altering levels of gene expression and activity. Depending upon the outcome of these experiments, more radical rearrangements may be performed in which multiple genes are repositioned to multiple sites. Care will be taken to maintain tight genetic juxtaposition and the appropriate TRS control

such that rearranged genes will use their normal CS site for expression of subgenomic mRNAs. After introducing these rearrangements into the SARS F subclone, sequence analysis will be used to identify the consensus cDNA used in the assembly of full length SARS cDNAs. Molecularly cloned viruses will be rescued as described. Any number of rearranged SARS-CoV genomes can be built using the molecular clone or any other method to alter the 3' end of the SARS-CoV genome. Gene order rearranged viruses should be significantly resistant to recombination repair.

[0179] Phenotype and Genotype Analysis of Rescued Viruses. Rescued viruses will be isolated between 24-36 hrs post transfection, plaque purified and stocks grown in Vero E6 cells as previously described. Virus passage will be minimized to prevent the emergence of mutations that restore growth fitness in culture.

[0180] Genotype Analysis. Plaque purified molecularly cloned viruses will be inoculated into Vero E6 cells and intracellular RNA isolated using Trizol reagents. Using RT-PCR and primer-pairs that flank the newly rearranged genes, the presence of the mutant gene order in rescued viruses will be demonstrated. The amplicons will be sequenced to confirm the expected gene rearrangements. As rearrangements may alter expression of flanking genes, leader-containing amplicons will be RT-PCR cloned and sequenced to determine if subgenomic RNA synthesis has originated from appropriate CS sites.

[0181] Phenotype Analysis. Virus growth will be evaluated in culture and by plaque morphology. Rescued molecularly cloned viruses will be analyzed for SARS protein production by Western Blot using antiserum from mice inoculated with VRPs encoding the various SARS virus structural proteins and group specific ORFs. By Western Blot or immunoprecipitation with <sup>35</sup>S-methionine labeled cell extracts, the expression of the structural and various group specific nonstructural proteins that are expressed following gene rearrangement will be measured. Northern Blot analysis will be used to identify the viral mRNAs and determine if selected gene rearrangements alter the molar ratio of the upstream and downstream viral mRNAs. In selected instances, cultures will be radiolabeled with <sup>3</sup>H-Uridine and the kinetics of viral mRNA and replicative forms (RF) RNA synthesis will be analyzed. These experiments will allow for the determination of 1) whether full length and subgenomic length RF RNAs are present in SARS-CoV infected cells, 2) effects of reorganization on RNA expression, and 3) reorganization effects on relative molar ratios of mRNA and RF RNA.

[0182] Stability and Recombination Repair. Compensatory Evolution. Recombinants will be inoculated into Vero E6 cells and progeny viruses passaged into fresh cultures at 24 hrs postinfection. After 15-20 passages, growth curves and plaque size will be compared to wild type virus. Then, consensus SARS E and F subclones that contain mutations of interest will be reintroduced back into the full length cDNA to identify 2<sup>nd</sup> site mutations that restore replication fitness in the rearranged viruses.

[0183] RNA Recombination Repair: Gene order rearranged viruses are expected to be highly stable and several fold more resistant to RNA recombination repair by wild type viruses. Using SARS NSEM and SARS wild type as an example, it is noted that most recombination events origi-



nating from SARS NSEM to SARS wild type will result in an N gene deletion, a lethal event which reduces the recombination frequency by one-half. In contrast, recombination events originating from SARS wild type to SARS NSEM will predominantly lead to recombinant genomes containing two copies of N, which at least in the case of TGEV reduces replication fitness by about 1 log of titer. Only recombination events occurring within the pol genes of both viruses will reconstruct wild type virus, hardly a significant problem as it would only add to the wild type virus burden already present. Consequently, it is expected that gene order rearrangements will attenuate RNA recombination repair to give wild type genomes. In the absence of wild type virus, it is highly unlikely that RNA recombination can restore the gene order of SARS NSEM. This is because several recombination events are necessary, including duplication of the N gene at the 3' end of the genome (a double recombination event), followed by deletion of N from the upstream position.

[0184] To test this hypothesis, two approaches will be used. In the first approach, cultures of cells will be coinfecting with selected gene order rearranged viruses (SARS SNEM) and SARS wild type virus containing a deletion in the S CS and N-terminal S protein sequence and a GFP indicator. This genome should replicate but result in noninfectious virus particles lacking S. Rescue of the genome can occur by recombination or by phenotypic mixing of S glycoprotein provided by SARS SNEM in trans. The basic approach is to transfect with SARS AS and then infect with SARS SNEM at a MOI of 2 and passage progeny 3x (at high MOI<SARS ΔS should passage if S provided in trans). Using quantitative RT-PCR and primer pairs spanning the N/adjacent gene (SARS SNEM would be N/ORF3a; SARS wild type would be ORF8/N), virions will be harvested from supernatants at each passage, concentrated through gradients, and titered by plaque assay and quantitative RT-PCR using primer sets that detect SARS SNEM, SARS wild type and SARS SNEMN recombinant viruses. Green plaques (3x) can also be plaque-purified and the genotype of recombinant viruses characterized. In the second approach, a SARS 2N virus (the progeny of a wild type x SNEM recombination event) is constructed, which is expected to be replication impaired as compared to wild type.

#### Example 5

##### Remodeling SARS-CoV Genome Regulatory Networks Obstructs RNA Recombination

[0185] Live virus vaccines are a crucial intervention strategy documented to improve the overall health of communities because they have induced long-term immune responses that protect against a variety of highly pathogenic human and animal pathogens over the past century. Concerns regarding reversion to virulence by mutation and recombination, coupled with the associated challenges in commercially developing these vaccines, have diminished the appeal of using live virus vaccines in human and animal populations. The dichotomy between the well known protective efficacy and costs and risks of developing live virus vaccines has been recognized as one of the grand challenges in global health by the National Foundation for Infectious Diseases, calling for the development of new methods to stabilize and prevent reversion or recombination repair of attenuating alleles in live virus vaccines.

[0186] SARS-CoV was recognized as a highly pathogenic respiratory human pathogen that emerged suddenly and subsequently spread worldwide during the first few months of 2003. By the end of the outbreak, over 8,000 people had been infected, resulting in some 800 deaths worldwide. Demonstration of zoonotic forms of the SARS-CoV, both in farm animals and in bat populations, dictate a need for continued surveillance and the development of efficacious vaccines and therapeutics targeted to prevent reemergence and spread of this important human pathogen. SARS-CoV is an excellent model system for coronavirus vaccine development as this pathogen replicates efficiently in a variety of animal models, it is highly pathogenic in its normal host, the genome is well characterized, a reverse genetic system exists to manipulate the genome (49), and other important coronaviruses have been identified that cause significant diseases in human and in animal populations.

[0187] The SARS-CoV virion contains a single-stranded positive polarity 29,700 nucleotide RNA genome bound by the nucleocapsid protein, N. The capsid is surrounded by a lipid bilayer containing at least four structural proteins, designated S, ORF3a, M and E. The SARS-CoV genome contains nine open reading frames (ORFs), the first of which encodes the viral replicase proteins required for subgenomic and genome length RNA synthesis. ORFs 2-8 are encoded in eight subgenomic mRNAs synthesized as a nested set of 3' co-terminal molecules in which the leader RNA sequences, encoded at the 5' end of the genome, are joined to body sequences at distinct transcription regulatory sequences (TRS) which contains a highly conserved consensus sequence (CS), ACGAAC (SEQ ID NO: \_\_\_\_\_) in SARS-CoV and ACTAAAC (SEQ ID NO: \_\_\_\_\_) in other group II coronaviruses, respectively. Based on studies with other coronaviruses, SARS-CoV likely uses transcription attenuation to synthesize both full length and subgenomic length negative strand RNAs containing antileader sequences, which then function as the templates for the synthesis of like-sized mRNAs (44, 47, 48). It has been well established that alterations in body TRS sequences attenuate expression of individual subgenomic mRNAs while alterations in the leader TRS globally impact the synthesis of all subgenomic mRNAs. Interspaced among the SARS-CoV structural genes are the accessory or group specific genes, ORF3a/b, ORF6, ORF7a/b, ORF8a/b and ORF9b, which are not conserved in other coronaviruses and whose function in replication and pathogenesis are generally unknown. Group specific ORFs encoded in the SARS-CoV, mouse hepatitis virus (MHV), feline infectious peritonitis virus (FIPV) and transmissible gastroenteritis virus (TGEV) genome often encode luxury functions for replication in vitro, but may attenuate virulence in vivo (22, 46, 51). The exact functions of the SARS-CoV group specific ORFs are unclear in replication and pathogenesis. Importantly, coronaviruses undergo RNA recombination events at high frequency, presumably because of the large size of the genome, the presence of transcriptionally active full length and subgenomic length templates and a transcription strategy that requires disassociation/reassociation with template strands.

[0188] In the present invention, the SARS-CoV transcription network has been globally remodeled by engineering a novel six nucleotide TRS domain that is not present in the genome of SARS-CoV or any other known coronavirus. Studies described herein demonstrate that the recombinant virus, icSARS-CRG, replicates efficiently in cell culture and



expressed normal levels of the expected complement of subgenomic mRNAs that encoded both structural and non-structural proteins. A second recombinant virus, icSARS-PRG, containing a new TRS network that regulated expression of the genome length RNA as well as the subgenomic RNAs encoding essential structural proteins, but not the group specific ORFs 3a/b, ORF6, ORF7a/b and ORF8a/b, also replicated efficiently, demonstrating that the group specific ORFs were not essential for in vitro and in vivo replication. In contrast, chimeras containing a mixture of natural and rewired TRS networks that influenced essential gene expression were lethal. This is the first example of successfully designing new regulatory circuits into the genome of a mammalian virus. These studies serve as a paradigm for designing unique networks of interacting alleles that function as lethal genetic traps following RNA recombination with wild type viruses.

[0189] Viruses and Cells. The Urbani and icSARS strains of SARS-CoV (AY278741), icSARS-CoV Luc, icSARS-CoV Luc1, icSARS-CoV Luc-2 and the icSARS-CRG and PRG recombinant viruses were propagated on VeroE6 cells in Eagle's MEM supplemented with 10% fetal calf serum, kanamycin (0.25 µg/ml) and gentamycin (0.05 µg/ml) at 37° C. in a humidified CO<sub>2</sub> incubator. For virus growth, cultures of VeroE6 cells were infected at a multiplicity of infection (MOI) of 0.1 PFU for 1 hr, the monolayer washed 2× with 2 mls of PBS and overlaid with complete MEM. Virus samples were harvested at different times post infection and titered by plaque assay in 60 mm dishes. Plaques were visualized by neutral red staining and counted at 48 hrs. All virus work was performed in a biological safety cabinet in a biosafety level 3 laboratory containing redundant exhaust fans. Personnel were double-gloved and dressed in TYVEK suits with full hoods and face shields. Powered air purifying respirators (PAPR) with high efficiency particulate air (HEPA) and organic vapor filters were used to provide positive pressure environment within the hoods.

[0190] Construction of *Renilla* Luciferase encoding SARS-CoV Recombinant Clones. Plasmid DNAs were amplified in One Shot® Top 10 chemically competent cells (Invitrogen) and purified with the QIAPREP miniprep kit (Qiagen Inc., Valencia, Calif.). All restriction enzymes were purchased from New England BioLabs (NEB, Beverly, Mass.) and used according to manufacture's instructions. DNA fragments were isolated from 1.0% agarose gels with the QIAQUICK gel extraction kit (Qiagen Inc.). All DNA was visualized using DARKREADER technology (Clare Chemical Research, Denver, Colo.) to prevent UV-induced damage that potentially could affect efficacy of assembly at later stages, such as transcription. The six subgenomic cDNA clones (A to F) that include the entire ic SARS genome when combined were isolated using known molecular techniques. ORF 7a/b is located within cDNA F, nucleotides 27273-27772. The *Renilla* luciferase gene was inserted to replace ORF 7a/b.

[0191] Mutations were introduced to the TRS site by utilizing the 'no see'm' strategy. Three primer sets were designed to amplify three cDNA fragments. A forward primer (Ppum3: 5'-GCTGTGACATTAAGGACCTGCCAAAAG-3'; SEQ ID NO: \_\_\_\_\_) extending from the PpumI site was used concurrently with a reverse primer (3MUT3: 5'-AGGTGCACCTGCAGCCATTTTAATT-TATCCGGTTTATGGATA-3'; SEQ ID NO: \_\_\_\_\_) or

2MUT3: 5'-AGGTGCACCTGCAGCCATTTTAATT-TATCCGGTTTATGGATA-3'; SEQ ID NO: \_\_\_\_\_) ending at the ORF 7a/b TRS site, which included the appropriate TRS mutations and the outside cutter restriction site for AarI (Fermentas). This resulted in Amplicon 1 (TRS2), by PCR, with three mutations (CCGGAT; SEQ ID NO: \_\_\_\_\_) in the TRS site and Amplicon 2 (TRS1), with two mutations (ACGGAT (SEQ ID NO: \_\_\_\_\_) in the TRS site. The third amplicon was created using a forward primer (3MUT5: 5'-GGTGCACCTGCAAATAAATGGCTTCCA-3'; SEQ ID NO: \_\_\_\_\_) that overlapped with the previously mentioned AarI site designed into AMP1 and AMP2. The reverse primer (PacI3: 5'-TAAAGTGAGCTCTTAATTAATTACT-GCTCG-3'; SEQ ID NO: \_\_\_\_\_) ended at the downstream PacI site. This created, by PCR techniques, Amplicon 3 (AMP3). Each amplicon was digested with the appropriate restriction enzyme cocktail: AMP1 and AMP2-PpumI/AarI and AMP3-AarI/PacI. AMP1 and AMP2 were separately ligated to AMP3 for a 1.34 kb cDNA fragment that was subsequently cloned into pTOPO PCR-XL plasmid (Invitrogen). The ic SARS wild type luciferase (icSARS wt-Luc) construct and the new mutated clones were digested with PpuMI and PacI and the 7.91 kb band of the icSARS wt-Luc and the 1.34 kb band of the mutated clones were gel purified and ligated overnight at 4° C. in the presence of 1% T4 DNA Ligase (NEB). Clones with the correct mutations were verified by DNA sequencing with an automated sequencer (UNC-CH Genome Analysis Facility).

[0192] Construction of SARS Plasmids Containing Mutated Leader and Intergenic Sequences. To create a leader with the consensus sequence CCGGAT (SEQ ID NO: \_\_\_\_\_), the SARS A plasmid was PCR amplified with primer set M13R3 (CAGGAAACAGCTATGAC; SEQ ID NO: \_\_\_\_\_) and MuL1-(AAAATCCGGTTAGAGAACA-GATCTACAAGAG; SEQ ID NO: \_\_\_\_\_) or MuL1+(CTAACCGGATTTTAAAATCTGTGTAGCTGTC; SEQ ID NO: \_\_\_\_\_) and SARS 453-(ATAGGGCTGT-TCAAGCTGGGG; SEQ ID. NO: \_\_\_\_\_). The resulting fragments were combined in an overlapping PCR reaction to create an approximately 620 bp product that was subsequently cloned and sequenced. A plasmid containing the appropriate changes was digested with restriction enzymes MluI and AvrII and this fragment was used to replace the same fragment from the SARS A plasmid. Further sequencing of this plasmid confirmed that it was identical to the SARS A plasmid except for the altered leader. To mutate the spike (S) gene TRS, the SARS E fragment was PCR amplified with either primer set SARS #37 (TGCTG-GCTCTGATAAAGGAG; SEQ ID NO: \_\_\_\_\_) and MuS-gene-(NNNCACCTGCACATATCCGGTTAGTTGT-TAACAGAATATCAC (SEQ ID NO: \_\_\_\_\_) or MuSgene+(NNNCACCTGCAACCGGATATGTT-TATTTTCTTATTATTCTTACTCTC; SEQ ID NO: \_\_\_\_\_) and #10AgeI-(CATCAAGCGAAAAGGCAT-CAG; SEQ ID NO: \_\_\_\_\_). These fragments were digested with restriction enzyme AarI, ligated and subcloned. A consensus clone with the desired changes, was digested with BsmBI and AgeI, and used to replace the corresponding fragment in the SARS E plasmid. Next, the SARS F plasmid was PCR amplified with the following sets of primers: SARS #44 (TGATCCTCTGCAACCTGAGC (SEQ ID NO: \_\_\_\_\_) and MuEgene-(NNNCACCTGCATAAATC-CGGACTCACTTTCTTGTGCTTAC; SEQ ID NO: \_\_\_\_\_); MuEgene+(NNNCACCTGCGTCCGGATT-

TATGTACTCATTTCGTTTCGG; SEQ ID NO:\_\_\_\_\_) and MuMgene-(NNNCACCTGCAATAGTTAATCCGGTTA-GACCAGAAGATCAGGAAC; SEQ ID NO:\_\_\_\_); and MuMgene+(NNNCACCTGCGGATTAACCTATTATTAT-TATTCTGTTTGG; SEQ ID NO:\_\_\_\_) and 28033-(TAC-CAACACCTAGCTATAAGC; SEQ ID NO:\_\_\_\_). The three fragments were digested with the restriction enzyme AarI, directionally ligated and subcloned. A clone, containing the new consensus sequence CCGGAT (SEQ ID NO:\_\_\_\_) for the E and M genes, was digested with SmaI and NdeI and this fragment was inserted into the SARS F plasmid that had been identically digested. The resulting plasmid was designated FmuEandM. The SARS N gene TRS was mutated by PCR amplifying the F plasmid with primers MuNgene1 (GCTGCATTTAGAGACGTACTTGT-TGTTTTAAATAACCGGATAAATTAATAA TGCTC-GATAATGG; SEQ ID NO:\_\_\_\_) and SARS 3' Ng (TTAATTAATTATGCCTGAGTTGAATCAGCAG; SEQ ID NO:\_\_\_\_). The product was digested with BsmBI and used to replace the corresponding section in plasmid FmuE-andM, this new plasmid was called FmuE/M/N. Altering the ORF 3a TRS consisted of amplifying plasmid FmuE/M/N with the following primer sets; SARS #44 and SARSX1-(CGTCTCATGTGTAATGTAATTTGACACCC; SEQ ID NO:\_\_\_\_) or SARSX1+(CGTCTCACACATAACCG-GATTTATGGATTTGTTTATGAGATTTTT TAC; SEQ ID NO:\_\_\_\_) and 28033-, and then joining the two fragments using the restriction endonuclease BsmBI. This product was used to replace the SmaI-NdeI portion of FmuE/M/N. Primer sets SARS #47 (GTGCTTGCTGTTGTCTACAG; SEQ ID NO:\_\_\_\_) and SARSX3-(CGTCTCCGTCGG-GATGTAGCCACAGTGATCTC; SEQ ID NO:\_\_\_\_), SARSX3+(CGTCTCCGACGCTTTCTT-ATTA-CAAATTAGGAG; SEQ ID NO:\_\_\_\_) and SARSX4-(CGTCTCTCATATCCGGTTTATG-GATAATCTAACTCCATAG; SEQ ID NO:\_\_\_\_), and SARSX4+(CGTCTCATATGAAAATTATTCTCTCCCT-GAC; SEQ ID NO:\_\_\_\_) and 28033- were used to generate three PCR fragments that were digested with BsmBI, ligated with T4 DNA ligase and subcloned. A clone containing only the required changes was digested with AvrII and inserted into plasmid FmuE/M/N/X1 that had also been digested with AvrII. Finally, primer set SARS #48 (GGACTTTTCAGGATTGCTATTTG; SEQ ID NO:\_\_\_\_) and SARSX5-(CGTCTCATCCGGTTAGACTTTGGTA-CAAGGTTT; SEQ ID NO:\_\_\_\_) and set SARSX5+(CGTCTCCCGGATATGAAACTTCTCATTGTTTGGAC; SEQ ID NO:\_\_\_\_) and SARS3'X5 (NNNTTAATTAAT-TAATTGTTTCGTTTATTAAACAACA; SEQ ID NO:\_\_\_\_) created PCR products that were similarly joined using BsmBI and T4 DNA ligase. This product was introduced into plasmid FmuE/M/N/X1/X3/X4 using the NdeI-BstEII restriction sites. This plasmid, named FmuE/M/N/Xorfs, was sequenced to verify all mutations.

**[0193]** Assembly of Full Length cDNAs. The SARS A through F inserts were digested, separated through 0.8% agarose gels, visualized with a DARKREADER lightbox (Claire Chemical), excised and purified using the QIAEX II DNA purification kit. The SARS A+B, C+D and E+F fragments were ligated overnight, and the products isolated (17, 49). The SARS AB+CD+EF fragments were ligated overnight at 4° C., phenol/chloroform extracted and precipitated under isopropyl alcohol. Full-length transcripts were generated in vitro as described by the manufacturer

(Ambion, MMESAGE MMACHINE) with certain modifications. To produce full length capped SARS N gene transcripts, 1 µg of plasmid DNA encoding the SARS N gene was PCR amplified using forward primer (5'-NNGGC-CTCGATGGCCATTTAGGTGACACTA-TAGATGTCT-GATAATGGACCCCAATC-3'; SEQ ID NO:\_\_\_\_) and reverse primer (5'-NNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-TATGCCTGAGTTGAATCAGCAG-3'; SEQ ID NO:\_\_\_\_) and the amplicons purified from gels. Full length, polyadenylated N gene transcripts were transcribed by SP6 RNA polymerase with a 2:1 ratio of cap analog to GTP (Ambion, Austin, Tex.), mixed with full length transcripts and electroporated into cells.

**[0194]** Transfection of Full-Length Transcripts. RNA transcripts were added to 800 µl of the Vero E6 cell suspension (8.0x10<sup>6</sup>) in an electroporation cuvette and four electrical pulses of 450 V at 50 µF were given with a GENE PULSER II electroporator (BioRad) similar to protocols previously described (17, 49). The presence of full length cDNAs and transcripts was verified by separation on agarose gels and visualization by UV light. The transfected Vero cells were seeded in a 75 cm<sup>2</sup> flask and incubated at 37° C. for 2 days. Viruses were plaque purified in Vero E6 cells and stock grown in 75 cm<sup>2</sup> flasks.

**[0195]** Northern Blot Analysis. Cultures of Vero E6 cells were inoculated with the wild type SARS-CoV Urbani strain and various recombinant viruses at a MOI of 1.0 and incubated for 1 hr at 37° C. At 12 hrs post infection, intracellular RNA was isolated using RiboPure™ reagents as directed by the manufacturer (Ambion, Austin, Tex.). The mRNA was isolated using Qiagen's Oliogtex® mRNA spin-column reagents according to the manufacturer's direction (Qiagen, Valencia, Calif.). The mRNA was treated with glyoxal and separated on agarose gels using NorthernMax®-Gly according to the manufacturer's directions (Ambion, Austin, Tx). The RNA was transferred to BRIGHTSTAR-PLUS membrane (Ambion) for 4-5 hrs and the RNA cross-linked to the membrane by UV light. The blot was prehybridized and probed with an N gene-specific oligodeoxynucleotide probe (5'-CTTGACTGCCGCTCT-GCT<sup>a</sup>T<sup>b</sup>CCCT<sup>c</sup>CT<sup>d</sup>GC<sup>b</sup>-3'; SEQ ID NO:\_\_\_\_), where biotinylated nucleotides are designated with a superscript<sup>b</sup>. Blots were hybridized overnight and washed with low and high stringency buffers as recommended by the manufacturer. Filters were incubated with streptavidin-AP, washed, and then incubated with chemiluminescence substrate CDP-STAR. The blots were overlaid with film and developed.

**[0196]** Western Blot Analysis. Twelve hours post infection, Urbani, icSARS-CoV, icSARS-CoV Luc, icSARS-CoV Luc-1, icSARS-CoV Luc2, icSARS-CRG or icSARS PRG virus infected cells were washed in 1xPBS, lysed in buffer containing 20 mM Tris-HCl, pH 7.6, 150 mM NaCl, 0.5% deoxycholine, 1% nonidet-p-40, 0.1% sodium dodecyl sulphate (SDS), and post nuclear supernatants were added to an equal volume of 5 mM EDTA/0.9% SDS, resulting in a final SDS concentration of 0.5%. Samples were then heat inactivated for 30 minutes at 90° C. in the BL3 facility prior to removal. In a BL2 facility, samples were again heat inactivated for 30 minutes at 90° C. before use. Equivalent sample volumes were loaded onto 4 to 20% Criterion gradient gels (BioRad) and transferred to PVDF membrane (BioRad). For detecting SARS-CoV antigens, lots were

probed with polyclonal mouse antisera directed against Venezuelan equine encephalitis virus replicon particles (VRPs) that expressed the SARS-CoV ORF3a (VRP-ORF3a), S (VRP-S) or N (VRP-N) proteins diluted 1:200 and developed using ECL chemiluminescence reagents (Amersham Biosciences). *Renilla* luciferase expression was verified using antibodies purchased from commercial vendors.

**[0197]** TRS Function in SARS-CoV Transcription. Previous studies using TGEV as a model have established that high fidelity complementary base-pair communication between sequences encoded within the body TRS and the leader TRS encoded in the genome is absolutely essential for efficient expression of coronavirus subgenomic RNAs. As a first step toward remodeling the SARS-CoV TRS network, the nonessential ORF7a/b domain was replaced with the *Renilla* luciferase gene under the control of the ORF7a/b TRS motif (icSARS-CoV Luc). Double (icSARS-CoV Luc1) and triple (icSARS-CoV Luc2) mutations were then engineered in that should disrupt the ORF7a/b TRS communication network in the genome, ablating efficient mRNA 7 subgenomic transcription. Although the wild type SARS-CoV TRS is ACGAAC (SEQ ID NO: \_\_\_\_\_), the double mutant (TRS1-ACGGAT (SEQ ID NO: \_\_\_\_\_)) and triple mutant (TRS2-CCGGAT (SEQ ID NO: \_\_\_\_\_)) TRS sequences were unique with the latter not being encoded elsewhere in the Urbani genome and not used as a regulatory sequence in any known coronavirus. Recombinant cDNA genomes were assembled as previously described and recombinant viruses isolated by plaque purification. RT-PCR RFLP and sequence analysis demonstrated the appropriate  $\Delta$ ORF7a/b and luciferase replacement. Recombinant viruses encoding the luciferase gene replicated as efficiently as wild type virus, achieving titers greater than  $10^7$  PFU/ml within 20 hrs post infection, consistent with previous reports that demonstrated that ORF7a/b deletion did not significantly reduce virus replication in vitro or in vivo. In Vero cells, a progressive increase in *Renilla* luciferase expression was noted over the course of infection that peaked at 4-5 logs above background in icSARS-CoV Luc infected cultures. Under identical conditions, icSARS-CoV Luc-1 and icSARS-CoV Luc-2 displayed significant 90-95% reductions in global levels of luciferase protein expression, respectively over the course of infection. Western blot analysis confirmed the significant reduction in *Renilla* luciferase expression, but not N protein expression following infection of Vero cells. Further, Northern blot analyses clearly demonstrated that the TRS-1 and TRS-2 motifs significantly ablated expression of subgenomic mRNA 7 encoding luciferase.

**[0198]** Rewiring Coronavirus TRS Transcription Networks. Having demonstrated that the remodeled TRS-2 motif significantly attenuates communication with leader TRS elements and results in significant reductions in subgenomic mRNA synthesis, all of the SARS-CoV TRS elements were changed to the TRS-2 signature (icSARS-CoV CRG). A second mutant was engineered that contained the novel TRS-2 regulatory network, effectively establishing efficient communication between the leader sequence and the four essential structural genes, S, E, M and N (icSARS-PRG). However, this second regulatory network retained the wild type TRS sites that normally regulate expression of the group specific genes ORF3a/b, ORF6, ORF7a/b and ORF8a/b. The inefficient communication between the leader TRS-2

site and wild type TRS sites located just upstream of the group specific genes should significantly attenuate expression of the group specific genes. Recombinant viruses encoding the new TRS-2 and chimeric networks were readily isolated and plaque purified. Recombinant viruses icSARS-CRG and icSARS-PRG both replicated efficiently in Vero cells, approaching titers of mid  $10^7$  PFU/ml within 20 hrs post infection, equivalent to wild type viruses. Northern blot analyses revealed appropriately sized subgenomic mRNAs in icSARS-CRG infected cells, typical of wild type SARS-CoV infection. Importantly, subgenomic RNA profiles in icSARS-PRG infected cells displayed the expected set of subgenomic mRNAs encoding the structural genes (mRNA 2, 4, 5 and 9) that were driven from networked TRS-2 sites, but reduced and/or mis-sized subgenomic mRNAs driven from the wild type TRS sites regulating expression of the group specific ORFs. For example, expression of mRNA 3 is reduced by about 50%, while mRNAs 6 and 8 are mis-sized and mRNA 7 is apparently not expressed. Western blot analyses confirmed abundant levels of expression of the structural proteins S and N in all recombinant and wild type viruses, abundant expression of ORF3a in wild type and icSARS-CRG, but little if any expression of ORF3a in icSARS-PRG infected cultures. Analysis of leader-body TRS junctions in wild type and icSARS-CRG revealed usage of the appropriate wild type or mutant TRS sites. In icSARS-PRG, leader-body TRS-2 sites drove expression of subgenomic mRNAs encoding the structural proteins demonstrating efficient communication between networked leader/body TRS sites. The wild type TRS site ACGAAC (SEQ ID NO: \_\_\_\_\_) was rarely used for initiating expression of the group specific ORF encoding subgenomic RNAs. When this was the case, the body TRS sequence was preserved in mRNA consistent with the transcription attenuation model for Nidovirus mRNA synthesis. Most often, upstream and downstream noncanonical TRS sites were typically activated; most likely because they displayed increased homology with TRS-2. Aberrant leader-body junction sites oftentimes result in mRNAs that encode deletions in a group specific ORF or potentially silence group specific ORF expression because new upstream ATG start codons and small ORFs are encoded in the mRNA that are likely recognized and translated by host translational machinery, interfering with efficient translation and expression of the group specific ORF. The data demonstrate that deletion or efficient expression of the group specific ORFs is not essential for SARS-CoV replication in vitro

**[0199]** Remodeled TRS Networks Encode Lethal Genetic Traps Following RNA Recombination with Wild type Viruses. A series of wild type and chimeric recombinant viruses were engineered as described herein. In the first example, wild type or TRS-2 networks were preserved to efficiently express mRNA 2 encoding the S glycoprotein, but encode the heterologous TRS site for driving expression of the other structural genes (icSARS-Rec1; icSARS-Rec2). In another case, only the N gene TRS site was misaligned with the leader TRS site (icSARS-Rec3).

**[0200]** Full length cDNAs were constructed for wild type icSARS-CoV, icSARS-CRG and the three chimeric recombinant viruses and full length transcripts were electroporated in Vero cells. One-fifth of the electroporated cells were overlaid onto confluent monolayers, allowed to attach for 3 hrs and overlaid with agarose for plaque assay to determine the number of infectious centers. The remaining cells were

maintained in complete medium and virus samples and RNA harvested at different times post-electroporation. Approximately  $10^3$  infectious centers were detected for icSARS-CoV and icSARS-CRG, but no infectious centers were detected from icSARS-Rec1-3 transfected cultures. Moreover, infectious virus was readily detected, which increased to greater than  $10^7$  PFU/ml after about 72 hrs in icSARS-CoV and icSARS-CRG, but no virus was detected in cSARS-Rec 1-3 transfected cultures.

[0201] Using RT-PCR, leader-containing transcripts were detected in all wild type and icSARS-CRG transfected cultures at 24 and 48 hrs. These leader-containing transcripts originated at the appropriately networked combination of leader/body TRS sites. In contrast, only low levels of subgenomic mRNA transcripts were detected in icSARS-Rec1-3 transfected cultures at 24 hrs that had mostly disappeared by 48 hrs post-electroporation. Sequence analysis revealed that most leader-containing RNAs originated from noncanonical TRS sites located upstream or more often, downstream of the appropriate start location (FIG. 1A). In many cases, the noncanonical site usage results in large lethal deletions in critical structural genes like M (FIG. 1B) that prevent the production of infectious progeny.

[0202] Animal studies. Mice and ferrets are inoculated with  $1.0 \times 10^6$  plaque forming units (PFU) of icSARS-CRG and icSARS-PRG. A variety of immune responses (e.g., innate, mucosal, humoral, cellular) are measured according to standard protocols well known in the art at different times postinfection.

#### Example 6

##### Secondary Genetic Traps

[0203] A TRS consensus sequence (which is the wild type sequence and not a mutation) is engineered into a nidovirus genome or replicon RNA upstream or downstream from the normal TRS consensus sequence (e.g., of an essential structural protein gene such as S, M and/or N) site. This is done by identifying locations that are "CS-like, e.g., that naturally have 3-5 nucleotides of a six nucleotide CS and that fall just upstream or downstream of the wild type CS. The primary nucleotide sequence is modified at the "CS-like" site to be as close to the wild type CS as possible without altering the amino acid sequence of the protein. This modified CS site functions as a site for subgenomic transcription in recombinant virus progeny after recombination with wild type virus. In a resulting recombinant genome, expression of wild type genes is driven from these engineered start sites, resulting in N-terminal truncated proteins and/or out-of-frame ORFs.

[0204] In one embodiment, the sequence of the CRG or PRG genome S, M and N genes is analyzed and six nucleotide domains with close homology to the wild type TRS sequence, ACGAAC, are identified and characterized. Close sequence motifs within about 50, 75, 100, 150, 200, 250, 300, 350, 400, 450 or 500 nucleotides (upstream or downstream) of the normal mRNA CS initiation site are identified and those that can be mutated to near exact (5/6 or 6/6) duplicates of the wild type TRS sequence, ACGAAC, are introduced within one or more of the essential gene coding sequences, simultaneously maintaining the normal protein sequence. Following recombination events in which

the wild type TRS site is joined to the body of a CRG genome, the secondary trap becomes activated as a preferred site for subgenomic mRNA synthesis. Secondary traps can be introduced in the S, M or N proteins to introduce N-terminal or C-terminal deletions or to introduce out-of-frame ATG starts as the translated product of the subgenomic mRNA. All result in poor expression and/or deleted protein products. Using the sequence of the virus genes, the rewired TRS CS sites and the wild type TRS CS sites, different secondary traps can be introduced into any nidovirus rewired genome of this invention.

[0205] In a particular example, the CS of a SARS coronavirus is ACGAAC. Locations nearby this sequence in a structural protein gene (e.g., S, E, M, N) of a rewired genome or a replicon RNA comprising a structural protein gene would be identified that have 4 or 5 of the 6 nucleotides of this CS (e.g., AGGAAG). This site would be changed using standard procedures from AGGAAG to ACGAAC. This site would then be preferentially used when a wild type leader TRS is present in a recombinant virus genome. In an example where the secondary trap is introduced into the S gene, the S protein is truncated by 37 amino acids and a new ATG start site is used to produce a smaller S glycoprotein. The truncated protein functions poorly, resulting in a lethal phenotype.

[0206] The foregoing examples are illustrative of the present invention, and are not to be construed as limiting thereof. The invention is described by the following claims, with equivalents of the claims to be included therein.

[0207] All publications, patent applications, patents and other references cited herein are incorporated by reference in their entireties for the teachings relevant to the sentence and/or paragraph in which the reference is presented.

#### REFERENCES

- [0208] 1. Tsang et al. (2003). *N Engl J Med* 348:1977-1985.
- [0209] 2. Lee et al. (2003). *N Engl J Med* 348:1986-1994.
- [0210] 3. Peiris et al. (2003) *Lancet* 361: 1767-1772.
- [0211] 4. Poutanen et al. (2003). *N Engl J. Med.* 348: 1995-2005.
- [0212] 5. Peiris et al. (2003) *Lancet* 361: 1319-1325.
- [0213] 6. Ksiazek et al. (2003) *N Engl J Med* 348: 1953-1966.
- [0214] 7. Drosten et al. (2003) *N Engl J Med* 348: 1967-1976.
- [0215] 8. Rota et al. (2003) *Science* 300:1394-1398.
- [0216] 9. Marra et al. (2003) *Science* 300: 1399-1404.
- [0217] 10. Ruan et al. (2003) *Lancet* 361: 1779-1785.
- [0218] 11. Kim et al. (1995). *Virology* 208:1-8.
- [0219] 12. Kanjanahaluethai et al. (2003) *J Virol* 77: 7376-7382.
- [0220] 13. Gorbalenya et al. (1991) *FEBS Lett* 288: 201-205.
- [0221] 14. Lu et al. (1998) *J Virol* 72: 2265-71.

- [0222] 15. von Grotthuss et al. (2003). *Cell* 113: 701-702.
- [0223] 16. Zeibuhr et al. (2000) *J Gen Virol* 81: 853-879.
- [0224] 17. Yount et al. (2002) *J. Virol.* 76: 11065-11078.
- [0225] 18. Yount et al. (2000) *J. Virol.* 74: 10600-10611.
- [0226] 19. Almazon et al. (2000) *Proc Natl Acad Sci USA* 97: 5516-5521.
- [0227] 20. Casais et al. (2001) *J Virol* 75:12359-12369.
- [0228] 21. Thiel et al. (2001). *J Gen Virol* 82: 1273-81.
- [0229] 22. Curtis et al. (2002) *J Virol* 76:1422-1434.
- [0230] 23. Pensaert and de Bouck. (1978) *Arch Virol* 58: 243-247.
- [0231] 24. Pensaert et al. (1986). *Vet Q* 8: 257-261.
- [0232] 25. Duarte et al. *Virology* 198: 486-476.
- [0233] 26. Baric et al. (1997) *J Virol* 71:1946-1955.
- [0234] 27. Baric et al. (1999) *J. Virol.* 73:638-649.
- [0235] 28. Fischer et al. (1998) *J Virol* 71: 7885-7894.
- [0236] 29. de Haan et al. (2002) *Virology* 296:177-189.
- [0237] 30. Sola et al. (2002). *J Virol* 77: 4357-4369.
- [0238] 31. Ortega et al. (2002). *J Virol* 76: 11518-11529.
- [0239] 32. Enjuanes & van der Zeijst (1995). In: *The Coronaviridae*; S G Siddell, ed, p. 331-376.
- [0240] 33. Ladman et al. (2002). *Avian Dis* 46: 938-44.
- [0241] 34. Saif (1999). *Adv Vet Med* 41: 429-46.
- [0242] 35. van Nieuwstadt et al. (1989) *Vet Res.* 125: 58-60.
- [0243] 36. Crouch et al. (2000) *Vaccine* 19:189-196.
- [0244] 37. Park et al. (1998) *Am J Vet Res* 59:1002-8.
- [0245] 38. Rottier (1999) *Vet Microbiol* 69:117-25.
- [0246] 39. Vennema et al. (1997) *J Virol* 64:1407-1409.
- [0247] 40. Wang et al. (2002) *Avian Dis*, 46: 831-838.
- [0248] 41. Baron et al. (1997) *J Gen Virol.* 80:2031-2039.
- [0249] 42. Fouchier et al. (2003) *Nature* 423: 240.
- [0250] 43. Komatsu et al. (1986) *Exp Neurol* 91: 23-29.
- [0251] 44. Baric and Yount. 2000. Subgenomic negative-strand RNA function during mouse hepatitis virus infection. *J Virol* 74:4039-46.
- [0252] 45. de Haan et al. 2002. The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host. *Virology* 296:177-89.
- [0253] 46. Haijema et al. 2004. Live, attenuated coronavirus vaccines through directed deletion of group-specific genes provide protection against feline infectious peritonitis. *J. Virol.* 78:3863-3871.
- [0254] 47. Sawicki and Sawicki. 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *J Virol* 64:1050-6.

- [0255] 48. Schaad and Baric. 1994. Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. *J Virol* 68:8169-79.

- [0256] 49. Yount et al. 2003. Reverse genetics with a full length infectious cDNA of the severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci USA* 100:12995-13000.

TABLE 1

SARS Consensus Sequences				
TRS SITE	WT SEQUENCE	CS MUTANT #1	CS MUTANT #2	
Leader CS	TAAACGAAC	TAAACGgtC	TAAcCGgtC	
S CS	TAAACGAAC	TAAACGgtC	TAAcCGgtC	
ORF3a	TAAACGAAC	TAAACGgtC	TAAcCGgtC	
E CS	AGTACGAAC	AGTACGgtC	AGTcCGgtC	
M CS	TAAACGAAC	TAAACGgtC	TAAcCGgtC	
ORF6 CS	ATCACGAAC	ATCACGgtC	ATCcCGgtC	
ORF7 CS	AAAACGAAC	AAAACGgtC	AAAcCGgtC	
ORF8 CS	TAAACGAAC	TAAACGgtC	TAAcCGgtC	
N CS	TAAACGAAC	TAAACGgtC	TAAcCGgtC	

\* the CS is an ACGAAC

[0257]

TABLE 2

SARS Consensus Sequences				
TRS SITE	WT SEQUENCE	CS MUTANT #1	CS MUTANT #2	
Leader CS	TAAACGAAC	TAAACGgAt	TAAcCGgAt	
S CS	TAAACGAAC	TAAACGgAt	TAAcCGgAt	
ORF3a	TAAACGAAC	TAAACGgAt	TAAcCGgAt	
E CS	AGTACGAAC	AGTACGgAt	AGTcCGgAt	
M CS	TAAACGAAC	TAAACGgAt	TAAcCGgAt	
ORF6 CS	ATCACGAAC	ATCACGgAt	ATCcCGgAt	
ORF7 CS	AAAACGAAC	AAAACGgAt	AAAcCGgAt	
ORF8 CS	TAAACGAAC	TAAACGgAt	TAAcCGgAt	
N CS	TAAACGAAC	TAAACGgAt	TAAcCGgAt	

\* the functional CS is an ACGAAC

[0258]

## SEQUENCE LISTING

&lt;160&gt; NUMBER OF SEQ ID NOS: 76

&lt;210&gt; SEQ ID NO 1

&lt;211&gt; LENGTH: 29748

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: SARS coronavirus

&lt;400&gt; SEQUENCE: 1

```
atattagggt ttacctacc caggaaaagc caaccaacct c gatctcttg tagatctgtt    60
ctctaaacga actttaaaat ctgtgtagct gtcgctcggc tgc atgccta gtcacctac    120
gcagtataaa caataataaa ttttactgtc gttgacaaga aacgagtaac tcgtccctct    180
tgtgcagact gcttacgggt tcgtccgtgt tgcagtcgat catcagcata cctaggtttc    240
gtccgggtgt gaccgaaag taagatggag agccttggtc ttggtgtcaa cgagaaaaca    300
cacgtccaac tcagtttgcc tgccttcag gttagagacg tgctagtgcg tggcttcggg    360
gactctgtgg aagaggccct atcggaggca cgtgaacacc tcaaaaatgg cacttgtggt    420
ctagtagagc tggaaaaag cgtactgccc cagcttgaa agccctatgt gttcattaaa    480
cgttctgatg ccttaagcac caatcacggc cacaaggtcg ttgagctggt tgcagaaatg    540
gacggcattc agtacggtcg tagcgggata aactgggag tactcgtgcc acatgtgggc    600
gaaaccccaa ttgataccg caatgttctt cttcgtaaga acgtaataa gggagccggg    660
ggtcatagct atggcatcga tctaaagtct tatgacttag gtgacgagct tggcactgat    720
cccattgaag attatgaaca aaactggaac actaagcatg gcagtggtgc actccgtgaa    780
ctcactcgtg agctcaatgg aggtgcagtc actcgctatg tcgacaacaa tttctgtggc    840
ccagatgggt accctcttga ttgcatcaaa gattttctcg cacgcgctgg caagtcaatg    900
tgcactcttt ccgaacaact tgattacatc gagtcgaaga gaggtgtcta ctgctgccgt    960
gaccatgagc atgaaattgc ctggttcaact gagcgtctctg ataagagcta cgagcaccag    1020
acacccttcg aaattaagag tgccaagaaa tttgacactt tcaaggggga atgccaaaag    1080
tttgtgtttc ctcttaactc aaaagtcaaa gtcattcaac cacgtgttga aaagaaaaag    1140
actgagggtt tcatggggcg tatacgtctc gtgtaccctg ttgcatctcc acaggagtgt    1200
aacaatatgc acttgcttac cttgatgaaa tgtaatcatt gcgatgaagt ttc atggcag    1260
acgtgcgact ttctgaaagc cacttgtgaa cattgtggca ctgaaaatgt agttattgaa    1320
ggacctacta catgtgggta cctacctact aatgctgtag tgaaaatgcc atgtcctgcc    1380
tgtcaagacc cagagattgg acctgagcat agtgttgagc attatcacia cactcaaac    1440
attgaaactc gactccgcaa gggaggtagg actagatggt ttggaggctg tgtgtttgcc    1500
tatgttggtc gctataataa gcgtgcctac tgggttcctc gtgctagtgc tgatattggc    1560
tcaggacata ctggaattac tgggtgacaat gtggagacct tgaatgagga tctccttgag    1620
atactgagtc gtgaacgtgt taacattaac attgttggcg attttcattt gaatgaagag    1680
gttgccatca ttttggcacc tttctctgct tctacaagtg cctttattga cactataaag    1740
agtcttgatt acaagtcttt caaaaccatt gttgagtcct gcgtaacta taaagtacc    1800
aagggaaagc ccgtaaaagg tgcttggaac attggacaac agagatcagt ttaaacacca    1860
```

-continued

---

ctgtgtggtt	ttccctcaca	ggctgctggt	gttatcagat	caatTTTTgc	gcgcacactt	1920
gatgcagcaa	accactcaat	tcctgatttg	caaagagcag	ctgtcaccat	acttgatggt	1980
atTTctgaac	agtcattacg	tcttTgcgac	gccatggttt	atacttcaga	cctgctcacc	2040
aacagtgtca	ttattatggc	atatgtaact	ggtggtcttg	tacaacagac	ttctcagtgg	2100
ttgtctaatac	TTTTgggcac	tactgttgaa	aaactcaggc	ctatcttTga	atggattgag	2160
gcgaaactta	gtgcaggagt	tgaatttctc	aaggatgctt	gggagattct	caaatttctc	2220
attacaggtg	TTTTtgacat	cgTcaagggT	caaatacagg	ttgcttcaga	taacatcaag	2280
gattgtgtaa	aatgcttcat	tgatgttTgt	aacaaggcac	tcgaaatgtg	cattgatcaa	2340
gtcactatcg	ctggcgcaaa	gtTgcgatca	ctcaacttag	gtgaagtctt	catcgctcaa	2400
agcaagggac	ttaccgtca	gtgtatacgt	ggcaaggagc	agctgcaact	actcatgcct	2460
cttaaggcac	caaaagaagt	aaccttctt	gaaggTgatt	cacatgacac	agtacttacc	2520
tctgaggagg	ttgttctcaa	gaacggTgaa	ctcgaagcac	tcgagacgcc	cgttgatagc	2580
ttcacaatg	gagctatcgt	tggcacacca	gtctgtgtaa	atggcctcat	gctcttagag	2640
attaaggaca	aagaacaata	ctgcgcattg	tctcctggtt	tactggctac	aaacaatgtc	2700
tttcgcttaa	aagggggTgc	accaattaaa	ggtgtaacct	ttggagaaga	tactgtttgg	2760
gaagttcaag	gttacaagaa	tgtgagaatc	acatttgagc	ttgatgaacg	tgttgacaaa	2820
gtgcttaaty	aaaagtgcTc	tgtctacact	gttgaatccg	gtaccgaagt	tactgagttt	2880
gcatgtgttg	tagcagaggc	tgttgTgaag	actttacaac	cagtttctga	tctccttacc	2940
aacatgggta	ttgatcttga	tgagtggagt	gtagctacat	tctacttatt	tgatgatgct	3000
ggtgaagaaa	acttttcatc	acgtatgtat	tgttcctttt	accctccaga	tgaggaagaa	3060
gaggacgatg	cagagtgtga	ggaagaagaa	attgatgaaa	cctgtgaaca	tgagtacggT	3120
acagagggat	attatcaagg	tctccctctg	gaatttggtg	cctcagctga	aacagttcga	3180
gTtgaggaag	aagaagagga	agactggctg	gatgatacta	ctgagcaatc	agagattgag	3240
ccagaaccag	aacctacacc	tgaagaacca	gttaatcagt	ttactggTta	tttaaaactt	3300
actgacaatg	ttgccattaa	atgtgttgac	atcgTtaagg	aggcacaag	tgctaactct	3360
atggtgattg	taaatyctgc	taacatacac	ctgaaacatg	gtggtggtgt	agcaggtgca	3420
ctcaacaagg	caaccaatgg	Tgccatgcaa	aaggagagtg	atgattacat	taagctaaat	3480
ggcctcttta	cagtaggagg	gtcttgtttg	ctttctggac	ataatcttgc	taagaagtgt	3540
ctgcatgttg	ttggacctaa	cctaaatgca	ggtgaggaca	tccagcttct	taaggcagca	3600
tatgaaaatt	tcaattcaca	ggacatctta	cttgccacat	tgttTgcagc	aggcatattt	3660
ggtgctaaac	cacttcagtc	tttacaagtg	Tgcgtgcaga	cggttcgtac	acaggtttat	3720
attgcagtca	atgacaaaagc	tctttatgag	caggtTgtca	tggattatct	tgataacctg	3780
aagcctagag	tggaaagcacc	taaacaagag	gagccaccaa	acacagaaga	ttccaaaact	3840
gaggagaaat	ctgtcgtaca	gaagcctgtc	gatgtgaagc	caaaaattaa	ggcctgcatt	3900
gatgaggTta	ccacaacact	ggaagaaact	aagtttctta	ccaataagtt	actctgtttt	3960
gctgatataca	atggtaagct	ttaccatgat	tctcagaaca	Tgcttagagg	tgaagatag	4020
tctttccttg	agaaggatgc	accttacatg	gtaggTgatg	ttatcactag	tgtgatatac	4080
actTgtgttg	taataccctc	caaaaaggct	ggtggcacta	ctgagatgct	ctcaagagct	4140

-continued

---

ttgaagaaag tgccagttga tgagtataata accacgtacc ctggacaagg atgtgctggt	4200
tatacacttg aggaagctaa gactgctctt aagaaatgca aatctgcatt ttatgtacta	4260
ccttcagaag cacctaatac taaggaagag attctaggaa ctgtatcctg gaatttgaga	4320
gaaatgcttg ctcatgctga agagacaaga aaattaatgc ctatatgcat ggatgttaga	4380
gccataatgg caaccatcca acgtaagtat aaaggaatta aaattcaaga gggcatcggt	4440
gactatgggt tccgattctt cttttatact agtaaagagc ctgtagcttc tattattacg	4500
aagctgaact ctctaaatga gccgcttgtc acaatgcca ttggttatgt gacacatggt	4560
tttaactctg aagaggctgc gcgctgtatg cgttctctta aagctcctgc cgtagtgtca	4620
gtatcatcac cagatgctgt tactacatat aatggatacc tcacttcgtc atcaaagaca	4680
tctgaggagc actttgtaga aacagtttct ttggctggct cttacagaga ttggctcctat	4740
tcaggacagc gtacagagtt aggtgtttaa tttcttaagc gtggtgacaa aattgtgtac	4800
cacactctgg agagccccgt cgagtttcat cttgacgggt aggttctttc acttgacaaa	4860
ctaaagagtc tcttatccct gcgggaggtt aagactataa aagtgttcac aactgtggac	4920
aacactaatc tccacacaca gcttgtggat atgtctatga catatggaca gcagtttgg	4980
ccaacatact tggatgggtc tgatgttaca aaaattaaac ctcatgtaaa tcatgaggg	5040
aagactttct ttgtactacc tagtgatgac aactacgta gtgaagcttt cgagtactac	5100
catactcttg atgagagttt tcttggtagg tacatgtctg ctttaaacca cacaaagaaa	5160
tggaaatttc ctcaagttgg tggtttaact tcaattaaat gggctgataa caattgttat	5220
ttgtctagtg ttttattagc acttcaacag cttgaagtca aattcaatgc accagactt	5280
caagaggctt attatagagc ccgtgctggt gatgctgcta acttttgtgc actcacttc	5340
gcttacagta ataaaactgt tggcgagctt ggtgatgtca gagaaactat gaccatctt	5400
ctacagcatg ctaatttgga atctgcaaag cgagttctta atgtggtgtg taaacattgt	5460
ggtcagaaaa ctactacctt aacgggtgta gaagctgtga tgatatggg tactctatct	5520
tatgataatc ttaagacagc tgtttccatt ccatgtgtgt gtggtcgtga tgcacacaa	5580
tatctagtag aacaagagtc tcttttgggt atgatgtctg caccacctgc tgagtataaa	5640
ttacagcaag gtacattctt atgtgogaat gagtacactg gtaactatca gtgtggtcat	5700
tacactcata taactgctaa ggagaccctc tatcgtattg acggagctca ccttacaag	5760
atgtcagagt acaaaggacc agtgactgat gttttctaca aggaacatc ttacactaca	5820
accatcaagc ctgtgtcgtg taaactcgat ggagttactt acacagagat tgaacaaaa	5880
ttggatgggt attataaaaa ggataatgct tactatacag agcagcctat agacctgta	5940
ccaactcaac cattacaaa tgcgagtttt gataatttca aactcacatg ttctaacaca	6000
aaatttgctg atgattttaa tcaaagaca ggcttcacaa agccagcttc acgagagcta	6060
tctgtcacat tcttcccaga cttgaatggc gatgtagtgg ctattgacta tagacactat	6120
tcagcgagtt tcaagaaagg tgctaatta ctgcataagc caattgtttg gcacattaac	6180
caggctacaa ccaagacaac gttcaaacca aacacttgggt gtttacgttg tctttggagt	6240
acaaagccag tagatacttc aaattcattt gaagtctctg cagtagaaga cacacaagga	6300
atggacaatc ttgcttgtag aagtcacaaa cccacctctg aagaagtagt ggaaaatcct	6360
accatacaga aggaagtcac agagtgtgac gtgaaaacta ccgaagtgtg aggcaatgct	6420



-continued

---

atacttaaac catcagatga aggtgttaaa gtaacacagg agttaggtca tgaggatcctt	6480
atggctgctt atgtggaaaa cacaagcatt accattaaga aacctaata gctttcacta	6540
gccttagggt taaaaacaat tgccactcat ggtattgctg caattaatag tgttccttgg	6600
agtaaaattt tggcttatgt caaacatttc ttaggacaag cagcaattac aacatcaaat	6660
tgcgctaaga gattagcaca acgtgtgttt aacaattata tgccttatgt gtttacatta	6720
ttgttccaat tgtgtacttt tactaaaagt accaattcta gaattagagc ttcactacct	6780
acaactattg ctaaaaatag tgtaagagt gttgctaaat tatgtttgga tgccggcatt	6840
aattatgtga agtcacccaa attttctaaa ttgttcacaa tcgctatgtg gctattgttg	6900
ttaagtattt gcttaggttc tctaactctgt gtaactgctg cttttggtgt actcctatct	6960
aattttggtg ctccctctta ttgtaatggc gttagagaat tgtatcttaa ttcgtctaac	7020
gttactacta tggatttctg tgaaggttct tttccttga gcatttgttt aagtggatta	7080
gactcccttg attcctatcc agctcttgaa accattcagg tgacgatttc atcgtacaag	7140
ctagacttga caattttag tctggccgct gagtgggttt tggcatatat gttgttcaca	7200
aaattctttt atttattag tctttcagct ataatgcagg tgttcttgg ctattttgct	7260
agtcatttca tcagcaattc ttggctcatg tggtttatca ttagtattgt acaaatggca	7320
cccgtttctg caatggttag gatgtacatc ttctttgctt ctttctacta catatggaag	7380
agctatgttc atatcatgga tggttgcacc tcttcgactt gcatgatgtg ctataagcgc	7440
aatcgtgcca cacgctgtga gtgtacaact attgttaatg gcatgaagag atctttctat	7500
gtctatgcaa atggaggccg tggtctctgc aagactcaca attggaattg tctcaattgt	7560
gacacatttt gactggttag tacattcatt agtgatgaag ttgctcgtga tttgtcactc	7620
cagtttaaaa gaccaatcaa ccctactgac cagtcacgt atattgttga tagtgttget	7680
gtgaaaaatg gcgcttca cctctacttt gacaaggctg gtcaaaagac ctatgagaga	7740
catccgctct ccattttgt caatttagac aatttgagag ctaacaacac taaaggttca	7800
ctgctatta atgtcatagt ttttgatgac aagtccaaat gcgacgagtc tgcttctaag	7860
tctgctctg tgtactacag tcagctgatg tgccaaccta ttctgttget tgaccaagtt	7920
cttgtatcag acgttgagaga tagtactgaa gtttccgtta agatggttga tgcttatgtc	7980
gacacctttt cagcaacttt tagtgttct atggaaaaac ttaaggcact tgttgctaca	8040
gctcacagcg agttagcaaa ggggttagct ttagatggtg tcctttctac attcgtgtca	8100
gctgcccgac aagggttgtg tgataccgat gttgacacaa aggatgttat tgaatgtctc	8160
aaactttcac atcactctga cttagaagt acaggtgaca gttgtaacaa tttcatgctc	8220
acctataata aggttggaaa catgaogccc agagatcttg gcgcatgtat tgactgtaat	8280
gcaaggcata tcaatgccca agtagcaaaa agtcacaatg tttcactcat ctggaatgta	8340
aaagactaca tgtctttatc tgaacagctg cgtaaacaaa ttcgtagtgc tgccaagaag	8400
aacaacatac cttttagact aacttgtgct acaactagac aggttgtcaa tgcataact	8460
actaaaatct cactcaaggg tggtaagatt gttagtactt gttttaaact tatgcttaag	8520
gccacattat tgtgcttct tgctgactg gtttgttata tcgttatgcc agtacatata	8580
ttgtcaatcc atgatggtta cacaaatgaa atcattggtt acaagccat tcaggatggt	8640
gtcactcgtg acatcatttc tactgatgat tgttttgcaa ataaacatgc tggttttgac	8700

-continued

---

gcatggttta gccagcgtgg cggttcatac aaaaatgaca aaagctgccc thtagtagct	8760
gctatcatta caagagagat tggtttcata gtgcctggct taccgggtac tgtgctgaga	8820
gcaatcaatg gtgacttctt gcattttcta cctcgtgttt ttagtgctgt tggcaacatt	8880
tgctacacac cttccaaact cattgagtat agtgattttg ctacctctgc ttgcgttctt	8940
gctgctgagt gtacaathtt taaggatgct atgggcaaac ctgtgccata ttgttatgac	9000
actaatgtgc tagaggggtc tatttcttat agtgagcttc gtccagacac tcgttatgtg	9060
cttatggatg gttccatcat acagtttcct aacacttacc tggaggggtc tgttagagta	9120
gtaacaactt ttgatgctga gtactgtaga catggtacat gcgaaaggtc agaagtaggt	9180
atgtgcttat ctaccagtgg tagatgggtt cttaataatg agcattacag agctctatca	9240
ggagttttct gtggtgttga tgcgatgaat ctcatagcta acatctttac tcctcttctg	9300
caacctgtgg gtgctttaga tgtgtctgct tcagtagtgg ctggtggtat tattgccata	9360
ttggtgactt gtgctgccta ctactttatg aaattcagac gtgtttttgg tgagtacaac	9420
catgttgttg ctgctaatac acttttgttt ttgatgtctt tcactatact ctgtctggta	9480
ccagcttaca gctttctgcc gggagtctac tcagtccttt acttgaactt gacattctat	9540
ttccaatg atgtttcatt cttggctcac cttcaatggt ttgccatggt ttctcctatt	9600
gtgctttttt ggataacagc aatctatgta ttctgtatth ctctgaagca ctgccattgg	9660
ttctttaaca actatcttag gaaaagatc atgtttaatg gagttacatt tagtaccttc	9720
gaggaggctg ctttgtgtac ctttttgctc aacaaggaaa tgtacctaaa attgcgtagc	9780
gagacactgt tgccacttac acagtataac aggtatcttg ctctatataa caagtacaag	9840
tatttcagtg gagccttaga tactaccagc tatcgtgaag cagcttctg ccacttagca	9900
aaggctctaa atgactttag caactcaggt gctgatgttc tctaccaacc accacagaca	9960
tcaatcactt ctgctgttct gcagagtggg tttaggaaaa tggcattccc gtcaggcaaa	10020
gttgaagggt gcatgttaca agtaacctgt ggaactacaa ctcttaatgg attgtggttg	10080
gatgacacag tatactgtcc aagacatgct atttgcacag cagaagacat gcttaatcct	10140
aactatgaag atctgctcat tcgcaaatcc aaccatagct ttcttgttca ggctggcaat	10200
gttcaacttc gtgttatgtg ccattctatg caaaattgct tgcttaggct taaagtgtat	10260
acttctaacc ctaagacacc caagtataaa tttgtccgta tccaacctgg tcaaacattt	10320
tcagttctag catgctacaa tggttcacca tctggtgttt atcagtgctc catgagacct	10380
aatcatacca ttaaagggtc tttccttaat ggatcatgtg gtagtgttg ttttaacatt	10440
gattatgatt gcgtgtcttt ctgctatatg catcatatgg agcttccaac aggagtacac	10500
gctggtactg acttagaagg taaattctat ggtccatttg ttgacagaca aactgcacag	10560
gctgcaggta cagacacaac cataacatta aatgttttgg catggctgta tgcgtctggt	10620
atcaatgggt atagggtgtt tcttaataga ttcaccacta ctttgaatga ctttaacctt	10680
gtggcaatga agtacaacta tgaaccttg acacaagatc atgttgacat attgggacct	10740
ctttctgctc aaacaggaat tgccgtctta gatatgtgtg ctgctttgaa agagctgctg	10800
cagaatggta tgaatggctg tactatcctt ggtagcacta ttttagaaga tgagtttaca	10860
ccatttgatg ttgttagaca atgctctggt gttaccttcc aagtaagtt caagaaaatt	10920
gttaagggca ctcatcattg gatgctttta actttcttga catcactatt gattcttgtt	10980

---

-continued

---

caaagtacac agtggctcact gtttttcttt gtttacgaga atgctttctt gccatttact 11040  
cttggтата tggcaattgc tgcattgctc atgctgcttg ttaagcataa gcacgcattc 11100  
ttgtgcttgt ttctgttacc ttctcttgca acagttgctt actttaatat ggtctacatg 11160  
cctgctagct ggggtgatgc tatcatgaca tggcttgaat tggctgacac tagcttgtct 11220  
ggttataggc ttaaggattg tgttatgtat gcttcagctt tagttttgct tattctcatg 11280  
acagctcgca ctgtttatga tgatgctgct agacgtgttt ggacactgat gaatgtcatt 11340  
acacttgctt acaaagtcta ctatggtaat gcttttagatc aagctatttc catgtgggccc 11400  
ttagttatct ctgtaacctc taactattct ggtgctgta cgactatcat gtttttagct 11460  
agagctatag tgtttgtgtg tgttgagtat taccattgtt tatttattac tggcaacacc 11520  
ttacagtgta tcatgcttgt ttattgttct ttaggctatt gttgctgctg ctactttggc 11580  
ctttctgtt tactcaaccg ttacttcagg cttactcttg gtgtttatga ctacttggc 11640  
tctacacaag aatttaggta tatgaactcc caggggcttt tgcctcctaa gagtagtatt 11700  
gatgctttca agcttaacat taagtgttg ggtattggag gtaaacctatg tatcaagggt 11760  
gctactgtac agtctaaat gtctgacgta aagtgcacat ctgtggtact gctctcgggt 11820  
cttcaacaac ttagagtaga gctatcttct aaattgtggg cacaatgtgt acaactccac 11880  
aatgatattc ttcttgcaaa agacacaact gaagctttcg agaagatggt ttctcttttg 11940  
tctgttttgc tatccatgca ggggtgctga gacattaata ggttgtgca ggaatgctc 12000  
gataaccgtg ctactcttca ggctattgct tcagaattta gttctttacc atcatatgcc 12060  
gcttatgcca ctgcccaaga ggctatgag caggctgtag ctaatggtga ttctgaagtc 12120  
gttctcaaaa agttaaagaa atctttgaat gtggctaaat ctgagtttga cctgtatgct 12180  
gccatgcaac gcaagttgga aaagatggca gatcaggcta tgaccctaat gtacaaacag 12240  
gcaagatctg aggacaagag gccaaaagta actagtgcta tgcaacaat gctcttctc 12300  
atgcttagga agcttgataa tgatgcactt aacaacatta tcaacaatgc cctgtatggt 12360  
tgtgttccac tcaacatcat accattgact acagcagcca aactcatggt tgtgttccct 12420  
gattatggtta cctacaagaa cactgtgat ggtaacacct ttacatagc atctgcactc 12480  
tgggaaatcc agcaagttgt tgatgctgat agcaagattg ttcaacttag tgaattaac 12540  
atggacaatt caccaaatct ggcttggcct cttattgtta cagctctaag agccaactca 12600  
gctgttaaac tacagaataa tgaactgagt ccagtagcac tacgacagat gctctgtgctg 12660  
gctgttacca cacaaacagc ttgtactgat gacaatgac ttgctacta taacaattcg 12720  
aaggaggtta ggtttgtgct ggcttacta tcagaccacc aagatctcaa atgggctaga 12780  
ttcctaaga gtgatgttac aggtacaatt tacacagaac tggaaaccacc ttgtaggttt 12840  
gttacagaca caccaaaagc gcctaaagt aaatactgt acttcatcaa aggcttaaac 12900  
aacctaaata gaggtatggt gctgggcagt ttagctgcta cagtacgtct tcaggctgga 12960  
aatgctacag aagtacctgc caattcaact gtgctttcct tctgtgcttt tgcagtagac 13020  
cctgctaaaag catataagga ttacctagca agtggaggac aaccaatcac caactgtgtg 13080  
aagatgttgt gtacacacac tggtagagga caggcaatta ctgtaacacc agaagctaac 13140  
atggaccaag agtcttttgg tgggtcttca tgtgtctctg attgtatagc ccacattgac 13200  
catccaaatc ctaaaggatt ctgtgacttg aaaggtaagt acgtccaaat acctaccact 13260

-continued

---

tgtgctaata	accagtgagg	ttttacactt	agaaacacag	tctgtaccgt	ctgcggaatg	13320
tgaaagggtt	atggctgtag	ttgtgaccaa	ctccgcgaac	ccttgatgca	gtctgcggat	13380
gcatcaacgt	ttttaaacgg	gtttgcggtg	taagtgcagc	ccgtcttaca	ccgtgcggca	13440
caggcactag	tactgatgtc	gtctacaggg	cttttgatat	ttacaacgaa	aaagttgctg	13500
gttttgcaaa	gttcctaaaa	actaattgct	gtcgcctcca	ggagaaggat	gaggaaggca	13560
atltattaga	ctcttacttt	gtagttaaga	ggcatactat	gtctaactac	caacatgaag	13620
agactattta	taacttggtt	aaagattgtc	cagcgggtgc	tgtccatgac	tttttcaagt	13680
ttagagtaga	tggtgacatg	gtaccacata	tatcacgtca	gcgtctaact	aaatacacia	13740
tggttgat	agtctatgct	ctacgtcatt	ttgatgagg	taattgtgat	acattaaaag	13800
aaatactcgt	cacatacaat	tgctgtgatg	atgattat	caataagaag	gattggatg	13860
acttcgtaga	gaatcctgac	atcttacgcg	tatatgctaa	cttaggtgag	cgtgtacgcc	13920
aatcattatt	aaagactgta	caattctgcg	atgctatgcg	tgatgcaggc	attgtaggcg	13980
tactgacatt	agataatcag	gatcttaaat	ggaactggta	cgatttcggg	gatttcgtac	14040
aagtagcacc	aggctgcgga	gttcctattg	tggattcata	ttactcattg	ctgatgcca	14100
tcctcacttt	gactagggca	ttggctgctg	agtcccatat	ggatgctgat	ctcgcaaac	14160
cacttattaa	gtgggatctg	ctgaaatatg	attttacgga	agagagactt	tgtctcttcg	14220
accgttat	taaatattgg	gaccagacat	accatcccaa	ttgtattaac	tgtttgatg	14280
ataggtgat	ccttcattgt	gaaaacttta	atgtgttatt	ttctactgtg	ttccacctta	14340
caagttttg	accactagta	agaaaaatat	ttgtagatgg	tgttcctttt	gttgtttcaa	14400
ctggatacca	ttttcgtgag	ttaggagtgc	tacataatca	ggatgtaaac	ttacatagct	14460
cgcgtctcag	tttcaaggaa	cttttagtgt	atgctgctga	tccagctatg	catgcagctt	14520
ctggcaattt	attgctagat	aaacgcacta	catgcttttc	agtagctgca	ctaacaaca	14580
atgttgcttt	tcaaactgtc	aaaccoggta	attttaataa	agacttttat	gactttgctg	14640
tgtctaaagg	tttctttaag	gaaggaagtt	ctggtgaact	aaaacacttc	ttctttgctc	14700
aggatggcaa	cgctgctatc	agtgattatg	actattatcg	ttataatctg	ccaacaatgt	14760
gtgatatcag	acaactccta	ttcgtagtgt	aagttgttga	taaatacttt	gattgttacg	14820
atggtggctg	tattaatgcc	aaccaagtaa	tcgttaacaa	tctggataaa	tcagctgggt	14880
tcccatttaa	taaatggggt	aaggctagac	tttattatga	ctcaatgagt	tatgaggatc	14940
aagatgcact	tttcgcgtat	actaagcgta	atgtcatccc	tactataact	caaatgaatc	15000
ttaagtatgc	cattagtgtca	aagaatagag	ctcgcaccgt	agctgggtgc	tctatctgtg	15060
gtactatgac	aaatagacag	tttcatcaga	aattattgaa	gtcaatagcc	gccactagag	15120
gagctactgt	ggtaattgga	acaagcaagt	tttacgggtg	ctggcataat	atgttaaaaa	15180
ctgtttacag	tgatgtagaa	actccacacc	ttatgggttg	ggattatoca	aaatgtgaca	15240
gagccatgcc	taacatgctt	aggataatgg	cctctcttgt	tcttgctcgc	aaacataaca	15300
cttgctgtaa	cttatcacac	cgtttctaca	ggtagctaa	cgagtgtgcg	caagtattaa	15360
gtgagatggt	catgtgtggc	ggctcactat	atgttaaacc	agggtgaaca	tcacccgggtg	15420
atgtacaac	tgcttatgct	aatagtgctt	ttaacatttg	tcaagctggt	acagccaatg	15480
taaatgcact	tctttcaact	gatggtaata	agatagctga	caagtatgct	cgcaatctac	15540

-continued

---

aacacaggct ctatgagtgt ctctatagaa atagggatgt tgcacatgaa ttcgtggatg	15600
agttttacgc ttacctgcgt aaacatttct ccatgatgat tctttctgat gatgccgttg	15660
tgtgtataa cagtaactat gcggctcaag gtttagtagc tagcattaag aactttaagg	15720
cagttcttta ttatcaaaac aatgtgttca tgtctgaggc aaaatgttgg actgagactg	15780
accttactaa aggacctcac gaattttgct cacagcatac aatgctagtt aaacaaggag	15840
atgattacgt gtacctgcct taccagatc catcaagaat attaggcgca ggctgttttg	15900
tcgatgatat tgtcaaaaca gatggtacac ttatgattga aaggttcgtg tcaactggcta	15960
ttgatgctta cccacttaca aaacatccta atcaggagta tgctgatgtc tttcaactgt	16020
atttacaata cattagaaag ttacatgatg agcttactgg ccacatgttg gacatgtatt	16080
ccgtaatgct aactaatgat aacacctcac ggtactggga acctgagttt tatgaggcta	16140
tgtacacacc acatacagtc ttgcaggctg taggtgcttg tgtattgtgc aattcacaga	16200
cttcaacttg ttgcgggtgc tgtattagga gaccattcct atgttgcaag tgctgctatg	16260
accatgtcat ttcaacatca cacaatag tagttgtctgt taatccctat gtttgcattg	16320
ccccaggttg tgatgtcact gatgtgacac aactgtatct aggaggtatg agctattatt	16380
gcaagtcaca taagcctccc attagtttct cattatgtgc taatggtcag gtttttggtt	16440
tatacaaaaa cacatgtgta ggcagtgaca atgtcactga cttcaatgag atagcaacat	16500
gtgattggac taatgctgac gattacatac ttgccaacac ttgtactgag agactcaagc	16560
ttttcgagc agaaacgctc aaagccactg aggaacatt taagctgtca tatggtattg	16620
ctactgtacg cgaagtactc tctgacagag aattgcatct ttcattggag gttggaaaac	16680
ctagaccacc attgaacaga aactatgtct ttactgggta ccgtgtaact aaaaatagta	16740
aagtacagat tggagagtac acctttgaaa aaggtgacta tggatgatgct gttgtgtaca	16800
gaggactac gacatacaag ttgaatgttg gtgattactt tgtgttgaca tctcactg	16860
taatgccact tagtgacct actctagtc cacaagagca ctatgtgaga attactggct	16920
tgtaccacac actcaacatc tcagatgagt tttctagcaa tgttgcaaat tatcaaaag	16980
tcggcatgca aaagtactct aactccaag gaccacctgg tactggtaag agtcattttg	17040
ccatcgact tgctctctat taccatctg ctcgcatagt gtatacggca tgctctcatg	17100
cagctgttga tgccctatgt gaaaaggcat taaaatattt gccatagat aatgttagta	17160
gaatcatacc tgcgctgctg cgcgtagagt gttttgataa attcaaagt aattcaacac	17220
tagaacagta tgtttctgc actgtaaatg cattgccaga aacaactgct gacattgtag	17280
tctttgatga aatctctatg gctactaatt atgacttgag tgttgcaat gctagacttc	17340
gtgcaaaaca ctacgtctat attggogac ctgctcaatt accagcccc cgcacattgc	17400
tgactaaagg cacactagaa ccagaatatt ttaattcagt gtgcagactt atgaaaacaa	17460
taggtccaga catgttcctt ggaactgtc gccgtgtcc tgctgaaatt gttgacactg	17520
tgagtgttt agtttatgac aataagctaa aagcacacaa ggataagtca gctcaatgct	17580
tcaaaatggt ctacaaagg gttattacac atgatgtttc atctgcaatc aacagacctc	17640
aaataggcgt tgtaagagaa tttcttacac gcaatcctgc ttggagaaaa gctgttttta	17700
tctcacctta taattcacag aacgctgtag cttcaaaaat cttaggattg cctacgcaga	17760
ctgttgattc atcacagggt tctgaatatg actatgtcat attcacacaa actactgaaa	17820

-continued

---

cagcacactc	ttgtaatgtc	aaccgcttca	atgtggctat	cacaagggca	aaaattggca	17880
ttttgtgcat	aatgtctgat	agagatcttt	atgacaaaact	gcaatttaca	agtctagaaa	17940
taccacgctc	caatgtggct	acattacaag	cagaaaatgt	aactggactt	ttaaggact	18000
gtagtaagat	cattactggt	cttcaccta	cacaggcacc	tacacacctc	agcgttgata	18060
taaagttcaa	gactgaagga	ttatgtgttg	acataccagg	cataccaaag	gacatgacct	18120
accgtagact	catctctatg	atgggtttca	aatgaatta	ccaagtcaat	ggttacccta	18180
atatgtttat	caccgcgaa	gaagctattc	gtcacgttcg	tgcgtggatt	ggctttgatg	18240
tagagggctg	tcatgcaact	agagatgctg	tgggtactaa	cctacctctc	cagctaggat	18300
ttctacag	tgtaactta	gtagctgtac	cgactgttca	tggtgacact	gaaaataaca	18360
cagaattcac	cagagttaat	gcaaacctc	caccagggtga	ccagtttaaa	catcttatac	18420
cactcatgta	taaaggcttg	ccctggaatg	tagtgcgtat	taagatagta	caaagtctca	18480
gtgatacact	gaaaggattg	tcagacagag	tcgtgttcgt	cctttgggcg	catggctttg	18540
agcttacatc	aatgaagtac	tttgtcaaga	ttggacctga	aagaacgtgt	tgtctgtgtg	18600
acaaacgtgc	aacttgcttt	tctacttcat	cagatactta	tgcctgctgg	aatcattctg	18660
tgggttttga	ctatgtctat	aaccattta	tgattgatgt	tcagcagtg	ggctttacgg	18720
gtaaccttca	gagtaacct	gaccaacatt	gccagggtaca	tggaaatgca	catgtggcta	18780
gttgtgatgc	tatcatgact	agatgtttag	cagtccatga	gtgctttggt	aagcgcgttg	18840
attgtctgt	tgaataccct	attataggag	atgaactgag	ggttaattct	gcttgacaga	18900
aagtacaaca	catggtgtg	aagtctgccc	tcctggctga	taagtttcca	gttcttcctg	18960
acattggaaa	tccaaaggct	atcaagtgtg	tgcctcaggc	tgaagtagaa	tggaaattct	19020
acgatgctca	gcatgtag	gacaaagctt	acaaaataga	ggagctcttc	tattcttatg	19080
ctacacatca	cgataaatc	actgatggg	tttgtttgtt	ttggaattgt	aacgttgatc	19140
gttaccagc	caatgcaatt	gtgtgtaggt	ttgacacaag	agtcttgta	aacttgaact	19200
taccaggctg	tgatgggtg	agtttgatg	tgaataagca	tgcatccac	actccagctt	19260
tcgataaaa	tgcatctact	aatttaaac	aattgccttt	ctttactat	tctgatagtc	19320
cttgtgagtc	tcatggcaaa	caagtagtgt	cggatattga	ttatgttoca	ctcaaatctg	19380
ctacgtgtat	tacacgatgc	aatttaggtg	gtgctgtttg	cagacacat	gcaaatgagt	19440
accgacagta	cttgatgca	tataaatga	tgatttctgc	tggatttagc	ctatggattt	19500
acaaacaatt	tgatacttat	aacctgtgga	atacatttac	caggttacag	agtttagaaa	19560
atgtggctta	taatgtgtt	aataaaggac	actttgatgg	acacgccggc	gaagcacctg	19620
ttocatcat	taataatgct	gtttacacaa	aggtagatgg	tattgatgtg	gagatctttg	19680
aaaataagac	aacacttctc	gttaatgttg	catttgagct	tgggctaag	cgtaacatta	19740
aaccagtgcc	agagattaag	atactcaata	atttgggtgt	tgatctcgt	gctaactactg	19800
taatctggga	ctataaaaga	gaagccocag	cacatgtatc	tacaataggt	gtctgcacaa	19860
tgactgacat	tgccaagaaa	cctactgaga	gtgctgttcc	ttcacttact	gtctgtttg	19920
atggtagagt	ggaaggacag	gtagacctt	ttagaaacgc	ccgtaatggt	gttttaataa	19980
cagaaggctc	agtcaaaggt	ctaacacctt	caaagggacc	agcacaagct	agcgtcaatg	20040
gagtcacatt	aattggagaa	tcagtaaaaa	cacagtttaa	ctactttaag	aaagtagacg	20100

-continued

---

gcattattca acagttgcct gaaacctact ttactcagag cagagactta gaggatttta	20160
agcccagatc acaaatggaa actgactttc tcgagctcgc tatggatgaa ttcatacagc	20220
gatataagct cgagggctat gccttcgaac acatcgttta tggagatttc agtcatggac	20280
aacttgccgg tcttcattta atgataggct tagccaagcg ctcacaagat tcaccactta	20340
aattagagga ttttatccct atggacagca cagtgaaaa ttacttcata acagatgcgc	20400
aaacaggttc atcaaaatgt gtgtgttctg tgattgatct tttacttgat gactttgtcg	20460
agataataaa gtcacaagat ttgtcagtga tttcaaaagt ggtcaagggt acaattgact	20520
atgctgagat ttcattcatg ctttggtgta aggacggaca tgttgaacc ttctacccaa	20580
aactacaagc aagtcaagcg tggcaaccag gtgttgcat gcctaacttg tacaagatgc	20640
aaagaatgct tcttgaaaag tgtgacctc agaattatgg tgaaaatgct gttataccaa	20700
aaggaataat gatgaatgtc gcaaagtata ctcaactgtg tcaatactta aatacactta	20760
ctttagctgt accctacaac atgagagtta ttcactttgg tgctggctct gataaaggag	20820
ttgcaccagc tacagctgtg ctcagacaat ggttgccaac tggcacacta cttgtcgatt	20880
cagatcttaa tgacttcgtc tccgacgcag attctacttt aattggagac tgtgcaacag	20940
tacatacggc taataaatgg gaccttatta tttagcgatat gtatgaccct aggaccaaac	21000
atgtgacaaa agagaatgac tctaaagaag ggtttttcac ttatctgtgt ggatttataa	21060
agcaaaaact agccctgggt ggttctatag ctgtaaagat aacagagcat tcttggatg	21120
ctgaccttta caagcttatg ggccatttct catggtggac agcttttgtt acaaatgtaa	21180
atgcatcatc atcggaagca tttttaattg gggctaacta tcttggaag ccgaagggaac	21240
aaattgatgg ctataccatg catgctaact acattttctg gaggaacaca aatcctatcc	21300
agttgtcttc ctattcactc tttgacatga gcaaatctc tcttaaatta agaggaactg	21360
ctgtaatgtc tcttaaggag aatcaaatca atgatatgat ttattctctt ctggaaaaag	21420
gtaggcttat cattagagaa aacaacagag ttgtggtttc aagtgatatt cttgttaaca	21480
actaaacgaa catgtttatt ttcttattat ttcttactct cactagtggg agtgacctg	21540
accggtgcac cacttttgat gatgttcaag ctcctaatta cactcaacat acttcatcta	21600
tgaggggggt ttactatcct gatgaaattt ttagatcaga cactctttat ttaactcag	21660
atthatttct tccattttat tctaattgta cagggtttca tactattaat catacgtttg	21720
gcaaccctgt catacccttt aaggatggtt tttattttgc tggcacagag aatcaaatg	21780
ttgtccgtgg ttgggttttt ggttctacca tgaacaacaa gtcacagtcg gtgattatta	21840
ttaacaattc tactaatggt gttatcacgag catgtaactt tgaattgtgt gacaaccctt	21900
tctttgctgt ttctaaccac atgggtacac agacacatac tatgatattc gataatgcat	21960
ttaattgcac tttcagtagc atatctgatg ccttttcgct tgatggttca gaaaagtcag	22020
gtaattttaa acacttacga gagtttgtgt ttaaaaataa agatgggttt ctctatggtt	22080
ataagggcta tcaacctata gatgtagttc gtgatctacc ttctggtttt aacactttga	22140
aacctattht taagttgcct cttggtatta acattacaaa ttttagagcc attcttacag	22200
ccttttcacc tgctcaagac atttggggca cgtcagctgc agcctattht gttggctatt	22260
taaagccaac tacatttatg ctcaagtatg atgaaaatgg tacaatcaca gatgctgttg	22320
attgttctca aaatccactt gctgaactca aatgctctgt taagagcttt gagattgaca	22380

-continued

---

aaggaattta	ccagacctct	aatttcaggg	ttgttccctc	aggagatggt	gtgagattcc	22440
ctaataattac	aaacttgtgt	ccttttgag	aggtttttaa	tgctactaaa	ttcccttctg	22500
tctatgcatg	ggagagaaaa	aaaatttcta	attgtgttgc	tgattactct	gtgctctaca	22560
actcaacatt	tttttcaacc	tttaagtgt	atggcgtttc	tgccactaag	ttgaatgatc	22620
tttgcttctc	caatgtctat	gcagattcct	ttgtagtcaa	gggagatgat	gtaagacaaa	22680
tagcgccagg	acaaactggt	gttattgctg	attataatta	taaattgcc	gatgatttca	22740
tgggttgtgt	ccttgcttgg	aatactagga	acattgatgc	tacttcaact	ggtaattata	22800
attataaata	taggtatcct	agacatggca	agcttaggcc	ccttgagaga	gacatatcta	22860
atgtgccttt	ctcccctgat	ggcaaacctt	gcaccccacc	tgctcttaat	tgttattggc	22920
cattaaatga	ttatggtttt	tacaccacta	ctggcattgg	ctaccaacct	tacagagttg	22980
tagtactttc	ttttgaactt	ttaaatgcac	cggccacggg	ttgtggacca	aaattatcca	23040
ctgaccttat	taagaaccag	tgtgtcaatt	ttaattttaa	tggactcact	ggtactgggtg	23100
tgtaactcc	ttcttcaaa	agatttcaac	catttcaaca	atttggccgt	gatgtttctg	23160
atctcactga	ttccgtctga	gatcctaaaa	catctgaaat	attagacatt	tcaccttgct	23220
cctttggggg	tgtaagtga	attacacctg	gaacaaatgc	ttcatctgaa	gttgctgttc	23280
tatatcaaga	tgtaactgc	actgatgttt	ctacagcaat	tcatgcagat	caactcacac	23340
cagcttggcg	catataattct	actggaaaca	atgtattcca	gactcaagca	ggctgtctta	23400
taggagctga	gcatgtcgac	acttcttatg	agtgcgacat	tcctattgga	gctggcattt	23460
gtgctagtta	ccatacagtt	tctttattac	gtagtactag	ccaaaaatct	attgtggctt	23520
atactatgtc	tttaggtgct	gatagttcaa	ttgcttactc	taataacacc	attgctatac	23580
ctactaactt	ttcaattagc	attactacag	aagtaatgcc	tgtttctatg	gctaaaacct	23640
ccgtagattg	taatatgtac	atctgoggag	attctactga	atgtgctaat	ttgcttctcc	23700
aatatggtag	cctttgcaca	caactaaatc	gtgcactctc	aggtattgct	gctgaacagg	23760
atcgcaacac	acgtgaagtg	ttcgctcaag	tcaaacaaat	gtacaaaacc	ccaactttga	23820
aatattttgg	tggttttaa	ttttcacaaa	tattacctga	ccctctaaag	ccaactaaga	23880
ggtcttttat	tgaggacttg	ctctttaaata	aggtgacact	cgctgatgct	ggcttcatga	23940
agcaatatgg	cgaatgccta	ggtgatatta	atgctagaga	tctcatttgt	ggcgagaagt	24000
tcaatggact	tacagtgttg	ccacctctgc	tcactgatga	tatgattgct	gcctacacgg	24060
ctgctctagt	tagtggtagt	gccactgctg	gatggacatt	tgggtgctggc	gctgctcttc	24120
aaataccttt	tgctatgcaa	atggcatata	ggttcaatgg	cattggagtt	acccaaaatg	24180
ttctctatga	gaaccaaaaa	caaatogcca	accaatttaa	caaggcgatt	agtcaaatc	24240
aagaatcact	tacaacaaca	tcaactgcat	tgggcaagct	gcaagacggt	gttaaccaga	24300
atgctcaagc	attaacaca	cttgtaaac	aacttagctc	taattttggg	gcaatttcaa	24360
gtgtgctaaa	tgatattcct	tcgctacttg	ataaagtcga	ggcggaggta	caaattgaca	24420
ggttaattac	aggcagactt	caaagccttc	aaacctatgt	aacacaacaa	ctaactcagg	24480
ctgctgaaat	cagggcttct	gctaactctg	ctgctactaa	aatgtctgag	tgtgttcttg	24540
gacaatcaaa	aagagttgac	ttttgtggaa	agggtacca	ccttatgtcc	ttcccacaag	24600
cagccccgca	tgggtgtgtc	ttcctacatg	tcacgtatgt	gccatcccag	gagaggaact	24660



-continued

---

tcaccacagc gccagcaatt tgtcatgaag gcaaagcata cttccctcgt gaagggtgtt 24720
ttgtgtttaa tggcacttct tggtttatta cacagaggaa cttcttttct ccacaaataa 24780
ttactacaga caatacattt gtctcaggaa atttgtatgt cgttattggc atcattaaca 24840
acacagttta tgatcctctg caacctgagc tcgactcatt caaagaagag ctggacaagt 24900
acttcaaaaa tcatacatca ccagatgttg atcttggcga catttcaggc attaacgctt 24960
ctgtcgtcaa cattcaaaaa gaaattgacc gcctcaatga ggtcgctaaa aatttaaatg 25020
aatcactcat tgaccttcaa gaattgggaa aatatgagca atatattaaa tggccttggg 25080
atgtttggct cggcttcatt gctggactaa ttgccatcgt catggttaca atcttgcttt 25140
gttgcatgac tagttgttgc agttgcctca agggtgcatg ctcttggtgt tcttgctgca 25200
agtttgatga ggatgactct gagccagttc tcaagggtgt caaattacat tacacataaa 25260
cgaaactatg gatttgttta tgagattttt tactcttggg tcaattactg cacagccagt 25320
aaaaattgac aatgcttctc ctgcaagtac tgttcatgct acagcaacga taccgctaca 25380
agcctcactc ctttctggat ggcttgttat tggcgttgca tttcttgctg tttttcagag 25440
cgctacaaaa ataattgcgc tcaataaaaag atggcagcta gccctttata agggcttcca 25500
gttcatttgc aatttactgc tgctatttgt taccatctat tcacatcttt tgcttgcgc 25560
tgagggtatg gaggcgcaat ttttgctact ctatgccttg atatattttc tacaatgcat 25620
caacgcatgt agaattatta tgagatgttg gctttgttgg aagtgcaaat ccaagaacct 25680
attactttat gatgccaaact actttgtttg ctggcacaca cataactatg actactgtat 25740
accatataac agtgtcacag atacaattgt cgttactgaa ggtgacggca tttcaacacc 25800
aaaactcaaa gaagactacc aaattggtgg ttattctgag gataggcact cagggtgtaa 25860
agactatgtc gttgtacatg gctatttccac cgaagtttac taccagcttg agtctacaca 25920
aattactaca gacactggta ttgaaaatgc tacattcttc atctttaaca agcttgtaa 25980
agaccaccg aatgtgcaaa tacacacaat cgacggctct tcaggagttag ctaatccagc 26040
aatggatcca atttatgatg agccgacgac gactactagc gtgcctttgt aagcacaaga 26100
aagtgagtac gaacttatgt actcattcgt ttcggaagaa acaggtaoct taatagttaa 26160
tagcgtactt cttttcttg ctttctggtt attcttgcta gtcacactag ccatccttac 26220
tgcgcttcca ttgtgtgcgt actgtgcaa tattgttaac gtgagtttag taaaaccaac 26280
ggtttacgtc tactcgcggt ttaaaaaatct gaactcttct gaaggagttc ctgatcttct 26340
ggtctaaacg aactaactat tattattatt ctgtttgaa ctttaacatt gcttatcatg 26400
gcagacaacg gtactattac cgttgaggag cttaacaac tcctggaaca atggaaccta 26460
gtaataggtt tcctattcct agcctggatt atgttactac aattgccta ttctaacgg 26520
aacaggtttt gtacataat aaagcttgtt ttcctctggc tcttggtggc agtaacactt 26580
gcttgttttg tgcttgcgc tgtctacaga attaattggg tgactggcgg gattgctgatt 26640
gcaatggctt gtattgtag ctgtatgttg cttagctact tcgttgcttc ctacaggctg 26700
tttgctcgta cccgctcaat tgggtcattc aaccagaaa caaacattct tctcaatgtg 26760
ctctccggg ggacaattgt gaccagaccg ctcatggaaa gtgaacttgt cattggtgct 26820
gtgatcattc tgggtcactt gcgaatggcc ggacaccccc tagggcctg tgacattaag 26880
gacctgcaa aagagatcac tgtggctaca tcacgaacgc tttcttatta caaattagga 26940

-continued

---

gcgtcgcagc	gtgtaggcac	tgattcaggt	tttgctgcat	acaaccgcta	ccgtattgga	27000
aactataaat	taaatacaga	ccacgccggt	agcaacgaca	atattgcttt	gctagtagacag	27060
taagtgacaa	cagatgtttc	atcctgttga	cttccaggtt	acaatagcag	agatattgat	27120
tatcattatg	aggactttca	ggattgctat	ttggaatcct	gacgttataa	taagttcaat	27180
agtgagacaa	ttatттаagc	ctctaactaa	gaagaattat	tcggagttag	atgatgaaga	27240
acctatggag	ttagattatc	cataaacga	acatgaaaat	tattctcttc	ctgacattga	27300
ttgtatttac	atccttgcgag	ctatatcact	atcaggagtg	tgtagaggt	acgactgtac	27360
tactaaaaga	accttgccca	tcaggaacat	acgagggcaa	ttcaccattt	caccctcttg	27420
ctgacaataa	atttgacta	acttgacta	gcacacactt	tgcttttgct	tgtgctgacg	27480
gtactcgaca	tacctatcag	ctgcgtgcaa	gatcagtttc	acaaaaactt	ttcatcagac	27540
aagaggaggt	tcaacaagag	ctctactcgc	cactttttct	cattgttgct	gctctagtat	27600
ttttaatact	ttgcttcacc	attaagagaa	agacagaatg	aatgagctca	ctttaattga	27660
cttctatttg	tgctttttag	cctttctgct	attccttggt	ttaataatgc	ttattatatt	27720
ttggttttca	ctcgaaatcc	aggatctaga	agaaccttgt	accaaagtct	aaacgaacat	27780
gaaacttctc	attgttttga	cttgtatttc	tctatgcagt	tgcatatgca	ctgtagtaca	27840
gcgctgtgca	tctaataaac	ctcatgtgct	tgaagatcct	tgtaaggtag	aacactaggg	27900
gtaataactta	tagcactgct	tggctttgtg	ctctagaaa	ggttttacct	tttcatagat	27960
ggcacactat	ggttcaaaca	tgcacaccta	atgttactat	caactgtcaa	gatccagctg	28020
gtggtgcgct	tatagctagg	tggttggtacc	ttcatgaagg	tcaccaaact	gctgcattta	28080
gagacgtact	tgttgtttta	aataaacgaa	caaattaaa	tgtctgataa	tgaccctcaa	28140
tcaaaccaac	gtagtgcccc	ccgcattaca	tttggtggac	ccacagattc	aactgacaat	28200
aaccagaatg	gaggacgcaa	tggggcaagg	ccaaaacagc	gccgacccca	aggtttacc	28260
aataaactg	cgcttgggt	cacagctctc	actcagcatg	gcaaggagga	acttagattc	28320
cctcgaggcc	aggcgcttc	aatcaacacc	aatagtggtc	cagatgacca	aattggctac	28380
taccgaagag	ctaccgcagc	agttcgtggt	ggtgacggca	aaatgaaaga	gctcagcccc	28440
agatgttact	tctattacct	aggaactggc	ccagaagctt	cacttcccta	cgcgctaac	28500
aaagaaggca	tcgtatgggt	tgcaactgag	ggagccttga	atacacccaa	agaccacatt	28560
ggcaccgcga	atcctaataa	caatgctgcc	accgtgctac	aacttctca	aggaacaaca	28620
ttgccaagg	gcttctacgc	agagggaagc	agagcgggca	gtcaagcctc	ttctcgtcc	28680
tcatcagta	gtcgcggtaa	ttcaagaaat	tcaactctg	gcagcagtag	gggaaattct	28740
cctgctcgaa	tggctagcgg	agggtggtgaa	actgccctcg	cgctattgct	gctagacaga	28800
ttgaaccagc	ttgagagcaa	agtttctggt	aaaggccaac	aacaacaagg	ccaaactgtc	28860
actaagaaat	ctgctgctga	ggcatctaaa	aagcctcgcc	aaaaacgtac	tgccacaaaa	28920
cagtacaacg	tactcaagc	atttgggaga	cgtggtccag	aacaaccca	aggaaatttc	28980
ggggaccaag	acctaatacag	acaaggaaact	gattacaaac	attggccgca	aattgcacaa	29040
tttgctccaa	gtgcctctgc	attcttttga	atgtcacgca	ttggcatgga	agtcacacct	29100
tcgggaacat	ggctgactta	tcatggagcc	attaatttgg	atgacaaaga	tccacaattc	29160
aaagacaacg	tcatactgct	gaacaagcac	attgacgcat	acaaaacatt	cccaccaaca	29220

-continued

---

```

gagcctaaaa aggacaaaaa gaaaagact gatgaagctc agcctttgcc gcagagacaa 29280
aagaagcagc ccaactgtgac tcttcttctc gcggtgaca tggatgattt ctccagacaa 29340
cttcaaaatt ccatgagtgg agcttctgct gattcaactc aggcataaac actcatgatg 29400
accacacaag gcagatgggc tatgtaaacy ttttcgcaat tccgtttacg atacatagtc 29460
tactcttgtg cagaatgaat tctcgtact aacacagaca agtaggttta gttaacttta 29520
atctcacata gcaatcttta atcaatgtgt aacattaggg aggactttaa agagccacca 29580
cattttcctc gagggccagc ggagtacgat cgaggggtaca gtgaataatg ctagggagag 29640
ctgcctatat ggaagagccc taatgtgtaa aattaathtt agtagtgcta tccccatgtg 29700
attttaatag ctcttagga gaatgacaaa aaaaaaaaa aaaaaaaa 29748

```

```

<210> SEQ ID NO 2
<211> LENGTH: 29727
<212> TYPE: DNA
<213> ORGANISM: SARS coronavirus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(29727)
<223> OTHER INFORMATION: SARS Urbani 278741 sequence

```

```

<400> SEQUENCE: 2

```

```

atattaggtt ttacctacc caggaaaagc caaccaacct cgatctcttg tagatctggt 60
ctctaaacga actttaaaat ctgtgtgact gtcgctcggc tgcagccta gtgcacctac 120
gcagtataaa caataataaa ttttactgtc gttgacaaga aacgagtaac tcgtccctct 180
tctgcagact gttacgggtt tcgctcgtgt tgcagtcgat catcagcata cctaggtttc 240
gtccgggtgt gaccgaaag taagatggag agccttgctt ttggtgtcaa cgagaaaaca 300
cacgtccaac tcagtttgcc tgccttcag gttagagacg tgctagtgcg tggcttcggg 360
gactctgtgg aagagggcct atcggaggca cgtgaacacc tcaaaaatgg cacttgtggt 420
ctagtagagc tggaaaaagc cgtactgccc cagcttgaac agccctatgt gttcattaaa 480
cgttctgatg ccttaagcac caatcacgac cacaaggtcg ttgagctggt tgcagaaatg 540
gacggcattc agtacggtcg tagcgggata aactgggag tactcgtgcc acatgtgggc 600
gaaaccccaa ttgcataccg caatgttctt cttcgtgaa acggaataa gggagccggg 660
ggtcatagct atggcatcga tctaaagtct tatgacttag gtgacgagct tggcactgat 720
cccattgaag attatgaaca aaactggaac actaagcatg gcagtgtgac actccgtgaa 780
ctcactcgtg agctcaatgg aggtgcagtc actcgtatg tgcacaaca tttctgtggc 840
ccagatgggt accctcttga ttgcataaaa gattttctcg cacgcccggg caagtcaatg 900
tgcactcttt ccgaacaact tgattacatc gagtcaaga gaggtgtcta ctgctgccgt 960
gaccatgagc atgaaatgc ctggttccact gagcgtctg ataagagcta cgagcaccag 1020
acacccttcg aaattaagag tgccaagaaa tttgacactt tcaaaaggga atgccaaaag 1080
tttgtgttct ctcttaactc aaaagtcaaa gtcattcaac cacgtgttga aaagaaaaag 1140
actgaggggt tcatggggcg tatacgtctt gtgtaccctg ttgcatctcc acaggagtgt 1200
aacaatatgc acttgtctac cttgatgaaa tgtaatcatt gcgatgaagt ttcattggag 1260
acgtgcgact ttctgaaagc cacttgtgaa cattgtggca ctgaaaatth agttattgaa 1320
ggacctacta catgtgggta cctacctact aatgctgtag tgaaaatgcc atgtcctgcc 1380

```

-continued

---

tgtcaagacc	cagagattgg	acctgagcat	agtggtgcag	attatcacia	ccactcaaac	1440
attgaaactc	gactccgcaa	gggaggtagg	actagatggt	ttggagctg	tgtgttgccc	1500
tatgttggct	gctataataa	gcgtgcctac	tgggttcctc	gtgctagtgc	tgatattggc	1560
tcaggccata	ctggcattac	tggtgacaat	gtggagacct	tgaatgagga	tctccttgag	1620
atactgagtc	gtgaacgtgt	taacattaac	attgttggcg	atcttcattt	gaatgaagag	1680
gttgccatca	ttttggcatc	tttctctgct	tctacaagtg	cctttattga	cactataaag	1740
agctttgatt	acaagctctt	caaaaccatt	gttgagtcct	gcgtaacta	taaagttacc	1800
aagggaagc	ccgtaaaagg	tgcttggaac	attggacaac	agagatcagt	tttaacacca	1860
ctgtgtggtt	ttccctcaca	ggctgctggt	gttatcagat	caatctttgc	gcgcacactt	1920
gatgcagcaa	accactcaat	tcctgatttg	caaagagcag	ctgtcaccat	acttgatggt	1980
atctctgaac	agtcattacg	tcttgtcgac	gccatggttt	atacttcaga	cctgctcacc	2040
aacagtgtca	ttattatggc	atatgtaact	gggtgtcttg	tacaacagac	ttctcagtgg	2100
ttgtctaata	ttttgggcat	tactgttgaa	aaactcaggc	ctatctttga	atggattgag	2160
gcgaaactta	gtgcaggagt	tgaatttctc	aaggatgctt	gggagattct	caaatttctc	2220
attacaggtg	tttttgacat	cgtaagggtt	caaatacagg	ttgcttcaga	taacatcaag	2280
gattgtgtaa	aatgcttcat	tgatgttgtt	aacaaggcac	tcgaaatgtg	cattgatcaa	2340
gtcactatcg	ctggcgcaaa	gttgcgatca	ctcaacttag	gtgaagtctt	catcgctcaa	2400
agcaagggac	tttacgctca	gtgtatacgt	ggcaaggagc	agctgcaact	actcatgcct	2460
cttaaggcac	caaaagaagt	aacctttctt	gaaggtgatt	cacatgacac	agtacttacc	2520
tctgaggagg	ttgttctcaa	gaacggtgaa	ctcgaagcac	tcgagacgcc	cgttgatagc	2580
ttcacaatg	gagctatcgt	tggcacacca	gtctgtgtaa	atggcctcat	gctcttagag	2640
attaaggaca	aagaacaata	ctgcgcattg	tctcctggtt	tactggctac	aaacaatgtc	2700
tttcgcttaa	aagggggtgc	accaatataa	gggttaacct	ttggagaaga	tactgtttgg	2760
gaagtccaag	gttacaagaa	tgtgagaatc	acatttgagc	ttgatgaacg	tgttgacaaa	2820
gtgttaaatg	aaaagtgtct	tgtctacact	gttgaatccg	gtaccgaagt	tactgagttt	2880
gcatgtgttg	tagcagaggc	tgttgtgaag	actttacaac	cagtttctga	tctccttacc	2940
aacatgggta	ttgatcttga	tgagtggagt	gtagctacat	tctacttatt	tgatgatgct	3000
gggtaagaaa	acttttctac	acgtatgtat	tgttcctttt	accctccaga	tgaggaagaa	3060
gaggacgatg	cagagtgtga	ggaagaagaa	attgatgaaa	cctgtgaaca	tgagtacggt	3120
acagaggatg	attatcaagg	tctccctctg	gaatttggtg	cctcagctga	aacagttcga	3180
gttgagggaag	aagaagagga	agactggctg	gatgatacta	ctgagcaatc	agagattgag	3240
ccagaaccag	aacctacacc	tgaagaacca	gttaatcagt	ttactgggta	tttaaaactt	3300
actgacaatg	ttgccattaa	atgtgttgac	atcgttaagg	aggcaciaag	tgctaactct	3360
atggtgattg	taaatgtctc	taacatacac	ctgaacatg	gtggtggtgt	agcaggtgca	3420
ctcaacaag	caaccaatg	tgccatgcaa	aaggagagtg	atgattacat	taagctaaat	3480
ggccctctta	cagtaggagg	gtccttgttg	ctttctggac	ataatcttgc	taagaagtgt	3540
ctgcatgttg	ttggacctaa	cctaaatgca	ggtagagaca	tccagcttct	taaggcagca	3600
tatgaaaatt	tcaattcaca	ggacatctta	cttgacccat	tgttgtcagc	aggcatatct	3660

-continued

---

ggtgctaaac cacttcagtc tttacaagtg tgcgtgcaga cggttcgtac acaggtttat	3720
attgcagtca atgacaaagc tctttatgag caggttgca tggattatct tgataacctg	3780
aagcctagag tggaaagcacc taaacaagag gagccaccaa acacagaaga ttccaaaact	3840
gaggagaaat ctgtcgtaca gaagcctgtc gatgtgaagc caaaaattaa ggcctgcatt	3900
gatgaggtta ccacaacact ggaagaaact aagtttctta ccaataagtt actcctgttt	3960
gctgatataca atggtaagct ttacatgat tctcagaaca tgcttagagg tgaagatag	4020
tctttccttg agaaggatgc accttacatg gtagggtgatg ttatcactag tgggtatatac	4080
acttggttg taataccctc caaaaaggct ggtggcacta ctgagatgct ctcaagagct	4140
ttgaagaaa tgccagttga tgagtatata accacgtacc ctggacaagg atgtgctggt	4200
tatacacttg aggaagctaa gactgctctt aagaaatgca aatctgcatt ttatgtacta	4260
ccttcagaag cacctaagtc taaggaagag attctaggaa ctgtatcctg gaatttgaga	4320
gaaatgcttg ctcatgctga agagacaaga aaattaatgc ctatatgcat ggatgttaga	4380
gccataatgg caaccatcca acgtaagat aaaggaatta aaattcaaga gggcatcgtt	4440
gactatggtg tccgattcct cttttatact agtaaagagc ctgtagcttc tattattacg	4500
aagctgaact ctctaaatga gccgcttgtc acaatgcaa ttggttatgt gacacatggt	4560
tttaactctg aagaggtg gcgctgtatg cgttctctta aagctcctgc cgtagtgtca	4620
gtatcatcac cagatgctgt tactacatat aatggatacc tcacttcgtc atcaaagaca	4680
tctgaggagc actttgtaga aacagtttct ttggctggct cttacagaga ttggctctat	4740
tcaggacagc gtacagagtt aggtgtgaa tttcttaagc gtggtgacaa aattgtgtac	4800
cacactctgg agagccccgt cgagtttcat cttgacggtg aggttctttc acttgacaaa	4860
ctaaagagtc tcttatccct gcgggaggtt aagactataa aagtgtcac aactgtggac	4920
aacactaatc tccacacaca gcttggtgat atgtctatga catatggaca gcagtttgg	4980
ccaacatact tggatggtgc tgatgttaca aaaattaac ctcatgtaa tcatgaggg	5040
aagactttct ttgtactacc tagtgatgac aactacgta gtgaagcttt cgagtactac	5100
catactcttg atgagagttt tcttggtagg tacatgtctg ctttaaaacca cacaaagaaa	5160
tggaatttc ctcaagttgg tggtttaact tcaattaat gggctgataa caattgttat	5220
ttgtctagtg ttttattagc acttcaacag cttgaagtca aattcaatgc accagcactt	5280
caagaggctt attatagagc ccgtgctggt gatgctgcta acttttgtgc actcactctc	5340
gcttacagta ataaaactgt tggcgagctt ggtgatgtca gagaactat gacctctctt	5400
ctacagcatg ctaatttgg aatctgcaaag cgagttctta atgtggtgtg taaacattgt	5460
ggtcagaaaa ctactacctt aacgggtgta gaagctgtga tgtatatggg tactctatct	5520
tatgataatc ttaagacag tgtttccatt ccatgtgtgt gtggtcgtga tgctacacaa	5580
tatctagtac aacaagagtc tcttttgg atgatgtctg caccacctgc tgagtataaa	5640
ttacagcaag gtacattcct atgtgogaat gactacactg gtaactatca gtgtggtcat	5700
tacactcata taactgctaa ggagaccctc tatcgtattg acggagctca ccttacaag	5760
atgtcagagt acaaaaggacc agtgactgat gttttctaca aggaacatc ttacactaca	5820
accatcaagc ctgtgtcgt aaaactcgat ggagttactt acacagagat tgaacccaaa	5880
ttgatgggtt attataaaaa ggataatgct tactatacag agcagcctat agaccttcta	5940

-continued

---

ccaactcaac cattacccaaa tgcgagtttt gataatttca aactcacatg ttctaacaca	6000
aaatttgctg atgatttaaa tcaaatgaca ggcttcacaa agccagcttc acgagagcta	6060
tctgtcacat tcttcccaga cttgaatggc gatgtagtgg ctattgacta tagacactat	6120
tcagcgagtt tcaagaaagg tgctaaatta ctgcataagc caattgtttg gcacattaac	6180
caggctacaa ccaagacaac gttcaaacca aacacttggg gtttacgttg tctttggagt	6240
acaaagccag tagatacttc aaattcattt gaagttctgg cagtagaaga cacacaagga	6300
atggacaatc ttgcttgtga aagtcaacaa cccacctctg aagaagtagt ggaaaatcct	6360
accatacaga aggaagtcat agagtgtgac gtgaaaacta ccgaagtgtg aggcaatgtc	6420
atacttaaac catcagatga aggtgttaaa gtaacacaag agttaggtca tgaggatcct	6480
atggctgctt atgtgaaaa cacaaagcatt accattaaga aacctaata gctttcacta	6540
gccttagggt taaaaacaat tgccactcat ggtattgctg caattaatag tgttccttgg	6600
agtaaaattt tggcttatgt caaaccttc ttaggacaag cagcaattac aacatcaaat	6660
tgcgctaaga gattagcaca acgtgtgttt aacaattata tgccttatgt gtttacatta	6720
ttgttccaat tgtgtacttt tactaaaagt accaattcta gaattagagc ttcactacct	6780
acaactattg ctaaaaatag tgttaagagt gttgctaaat tatgtttgga tgccggcatt	6840
aattatgtga agtcacccaa attttctaaa ttgttcacaa tcgctatgtg gctattgttg	6900
ttaagatatt gcttaggttc tctaactctg gtaactgctg cttttggtgt actcttatct	6960
aattttgggt ctcccttcta ttgtaatggc gttagagaat tgtatcttaa ttcgtctaac	7020
gttactacta tggatttctg tgaaggttct tttccttgca gcatttgttt aagtggatta	7080
gactcccctg attcttatcc agctcttgaa accattcagg tgacgatttc atcgtacaag	7140
ctagacttga caattttagg tctggccgct gagtgggttt tggcatatat gttgttcaca	7200
aaattctttt atttattagg tctttcagct ataatgcagg tgttcttttg ctattttgct	7260
agtcatttca tcagcaatc ttggctcatg tggtttatca ttagtattgt acaaatggca	7320
ccggttctg caatggttag gatgtacatc ttctttgctt ctttctacta catatggaag	7380
agctatgttc atatcatgga tgggtgcacc tottcgactt gcgatgtgtg ctataagcgc	7440
aatcgtgcca cacgcgttga gtgtacaact attgttaatg gcatgaagag atctttctat	7500
gtctatgcaa atggaggccg tggcttctgc aagactcaca attggaattg tctcaattgt	7560
gacacatttt gcaactgtag tacattcatt agtgatgaag ttgctcgtga tttgtcactc	7620
cagtttaaaa gaccaatcaa ccctactgac cagtcacgt atattgttga tagtgttgc	7680
gtgaaaaatg gcgcgcttca cctctacttt gacaaggctg gtcaaaagac ctatgagaga	7740
catccgctct cccattttgt caatttagac aatttgagag ctaacaacac taaaggttca	7800
ctgcctatta atgtcatagt ttttgatggc aagtccaaat gcgacgagtc tgcttctaag	7860
tctgcttctg tgtactacag tcagctgatg tgccaaceta tctgtttgct tgaccaagtt	7920
cttgtatcag acgttgagga tagtactgaa gtttccgtta agatgtttga tgcttatgct	7980
gacacctttt cagcaacttt tagtgttct atggaaaaac ttaaggcact tgttgctaca	8040
gctcacagcg agttagcaaa ggggtgatct ttagatggg tcctttctac attcgtgtca	8100
gctgcccgcg aaggtgttgt tgataccgat gttgacacaa aggatgttat tgaatgtctc	8160
aaactttcac atcactctga cttagaagtg acaggtgaca gttgtaacaa tttcatgctc	8220

-continued

---

acctataata aggttgaaaa catgacgccc agagatcttg gcgcatgtat tgactgtaat	8280
gcaaggcata tcaatgcccc agtagcaaaa agtcacaatg tttcactcat ctggaatgta	8340
aaagactaca tgtctttatc tgaacagctg cgtaaacaaa ttcgtagtgc tgccaagaag	8400
aacaacatac ctttttagact aacttgctgct acaactagac aggttgctca tgccataact	8460
actaaaatct cactcaaggg tggaagatt gttagtactt gttttaaact tatgcttaag	8520
gccacattat tgtgcgttct tgctgcattg gtttggtata tcgttatgcc agtacataca	8580
ttgtcaatcc atgatggta cacaaatgaa atcattgggt acaaaagccat tcaggatggt	8640
gtcactcgtg acatcatttc tactgatgat tgttttgcaa ataaacatgc tggttttgac	8700
gcatggttta gccagcgtgg tggttcatac aaaaatgaca aaagctgccc tgtagtagct	8760
gctatcatta caagagagat tggtttcata gtgcctggct taccgggtac tggctgaga	8820
gcaatcaatg tgacttctt gcattttcta cctcgtggtt ttagtgctgt tggcaacatt	8880
tgctacacac ctccaaact cattgagat agtgattttg ctacctctgc ttgcgttctt	8940
gctgctgagt gtacaattht taaggatgct atgggcaaac ctgtgccata ttgttatgac	9000
actaatttg tagagggttc tatttcttat agtgagcttc gtccagacac tcgttatgtg	9060
cttatggatg gttccatcat acagtttctt aacacttacc tggagggttc tgttagagta	9120
gtaacaactt ttgatgctga gtactgtaga catggtacat gcgaaaggtc agaagtaggt	9180
atthgcctat ctaccagtg tagatgggtt ctttaaatg agcattacag agctctatca	9240
ggagttttct gtgggttga tgcgatgaat ctcatagcta acatctttac tcctcttggtg	9300
caacctgtgg gtgctttaga tgtgtctgct tcagtagtgg ctggtggtat tattgccata	9360
ttggtgactt gtgctgccta ctactttatg aaattcagac gtgtttttgg tgagtacaac	9420
catgttggtg ctgctaatac acttttgttt ttgatgtctt tcactatact ctgtctggta	9480
ccagcttaca gctttctgcc gggagtctac tcagcttttt acttgtaact gacattctat	9540
ttaccaatg atgtttcatt cttggctcac cttcaatggt ttgccatggt ttctcctatt	9600
gtgccttttt ggataacagc aatctatgta ttctgtattht ctctgaagca ctgccattgg	9660
ttctttaaca actatcttag gaaaagatc atgtttaatg gagttacatt tagtaccttc	9720
gaggaggctg ctttgtgtac ctttttgctc aacaagaaa tgtacctaaa attgcgtagc	9780
gagacactgt tgccacttac acagtataac aggtatcttg ctctatataa caagtacaag	9840
tatttcagtg gagccttaga tactaccagc tadcgtgaag cagcttgctg ccaacttagca	9900
aaggctctaa atgactttag caactcaggt gctgatgttc tctaccaacc accacagaca	9960
tcaatcactt ctgctgttct gcagagtgg tttaggaaaa tggcattccc gtcaggcaaa	10020
gttgaagggt gcatggtaca agtaactgt ggaactacaa ctcttaatgg attgtgggtg	10080
gatgacacag tatactgtcc aagacatgct atthgcacag cagaagacat gcttaatcct	10140
aactatgaag atctgctcat tcgcaaatcc aaccatagct ttcttggtca ggtggcaat	10200
gttcaacttc gtgttattgg ccattctatg caaaattgct tgcttaggct taaagttgat	10260
acttctaacc ctaagacacc caagtataaa tttgtccgta tccaacctgg tcaaacattht	10320
tcagttctag catgctacaa tggttcacca tctgggtttt atcagtgctg catgagacct	10380
aatcatacca ttaaaggctc tttccttaat ggatcatgct gtagtggttg ttttaacatt	10440
gattatgatt gcgtgtcttht ctgctatatg catcatatgg agcttccaac aggagtacac	10500

-continued

---

gctggtactg	acttagaag	taaattctat	ggtccatttg	ttgacagaca	aactgcacag	10560
gctgcaggta	cagacacaac	cataacatta	aatgttttgg	catggctgta	tgctgctggt	10620
atcaatggty	ataggtyggt	tcttaataga	ttcaccacta	ctttgaatga	ctttaacctt	10680
gtggcaatga	agtacaacta	tgaacctttg	acacaagatc	atgttgacat	attgggacct	10740
ctttctgctc	aaacaggaat	tgccgtctta	gatatgtgtg	ctgctttgaa	agagctgctg	10800
cagaatggta	tgaatggctg	tactatcctt	ggtagcacta	ttttagaaga	tgagtttaca	10860
ccatttgatg	ttgttagaca	atgctctggt	gttaccttcc	aagtaagtt	caagaaaatt	10920
gttaagggca	ctcatcattg	gatgctttta	actttcttga	catcactatt	gattcttggt	10980
caaagtacac	agtggtcact	gtttttcttt	gtttacgaga	atgctttcct	gccatttact	11040
cttggtatta	tggaattg	tgcatgtgct	atgctgcttg	ttaagcataa	gcacgcattc	11100
ttgtgcttgt	ttctgttacc	ttctcttgca	acagttgctt	actttaatat	ggtctacatg	11160
cctgctagct	gggtgatg	tatcatgaca	tggcttgaat	tggctgacac	tagcttgctc	11220
ggttataggc	ttaaggatg	tgttatgat	gcttcagctt	tagttttgct	tattctcatg	11280
acagctcgca	ctgtttatga	tgatgctgct	agacgtgttt	ggacactgat	gaatgctcatt	11340
acacttggtt	acaaagtcta	ctatggtaat	gcttttagatc	aagctatttc	catgtgggcc	11400
ttagttat	ctgtaacctc	taactattct	ggtgtctgta	cgactatcat	gttttttagct	11460
agagctatag	tgtttgtgtg	tgttgagtat	taccattgt	tatttattac	tggcaacacc	11520
ttacagtgtg	tcatgcttgt	ttattgtttc	ttaggctatt	gttgctgctg	ctactttggc	11580
cttttctggt	tactcaaccg	ttacttcagg	cttactcttg	gtgtttatga	ctacttggtc	11640
tctacacaag	aatttaggta	tatgaactcc	caggggcttt	tgccctcctaa	gagtagtatt	11700
gatgctttca	agcttaacat	taagtgttg	ggtattggag	gtaaacatg	tatcaagggt	11760
gctactgtac	agtctaaaat	gtctgacgta	aagtgcacat	ctgtggtact	gctctcggtt	11820
cttcaacaac	ttagagtaga	gtcatcttct	aaattgtggg	cacaatgtgt	acaactccac	11880
aatgatattc	ttcttgcaaa	agacacaact	gaagctttcg	agaagatgg	ttctcttttg	11940
tctgttttgc	tatccatgca	gggtgctgta	gacattaata	ggttgtgoga	ggaaatgctc	12000
gataaccgtg	ctactcttca	ggctattgct	tcagaattta	gttctttacc	atcatatgcc	12060
gcttatgcca	ctgccaggga	ggcctatgag	caggctgtag	ctaaggtgga	ttctgaagtc	12120
gttctcaaaa	agttaaagaa	atctttgaat	gtggctaaat	ctgagtttga	ccgtgatgct	12180
gccatgcaac	gcaagttgga	aaagatggca	gatcaggcta	tgacccaaat	gtacaaacag	12240
gcaagatctg	aggacaagag	ggcaaaagta	actagtgcta	tgcaacaacat	gctcttcaact	12300
atgottagga	agcttgataa	tgatgcactt	aacaacatta	tcaacaatgc	gcgtgatggt	12360
tgtgttccac	tcaacatcat	accattgact	acagcagcca	aactcatggt	tgttgtccct	12420
gattatggta	cctacaagaa	cacttgtgat	ggtaacacct	ttacatatgc	atctgcactc	12480
tgggaaatcc	agcaagtgtg	tgatgoggat	agcaagattg	ttcaacttag	tgaattaac	12540
atggacaatt	caccaaat	ggcttggcct	cttattgtta	cagctctaag	agccaactca	12600
gctgttaa	tacagaataa	tgaactgagt	ccagtagcac	tacgacagat	gtcctgtg	12660
gctggtacca	cacaaacagc	ttgtactgat	gacaatgcac	ttgcctacta	taacaattcg	12720
aaggagggta	ggtttgtgct	ggcattacta	tcagaccacc	aagatctcaa	atgggctaga	12780



-continued

---

ttccctaaga	gtgatggtag	aggtacaatt	tacacagaac	tggaaccacc	ttgtaggttt	12840
gttacagaca	cacccaaaag	gcctaaagtg	aaatacttgt	acttcatcaa	aggcttaaac	12900
aacctaaata	gaggtagtgg	gctgggcagt	ttagctgcta	cagtacgtct	tcaggctgga	12960
aatgctacag	aagtacctgc	caattcaact	gtgctttcct	tctgtgcttt	tgcaagtagac	13020
cctgctaaag	catataagga	ttacctagca	agtggaggac	aaccaatcac	caactgtgtg	13080
aaagatgtgt	gtacacacac	tggtacagga	caggcaatta	ctgtaacacc	agaagctaac	13140
atggaccaag	agtcctttgg	tggtgcttca	tgttgtctgt	attgtagatg	ccacattgac	13200
catccaaatc	ctaaaggatt	ctgtgacttg	aaaggtaagt	acgtccaaat	acctaccact	13260
tgtgctaagt	accagtgagg	ttttacactt	agaaacacag	tctgtaccgt	ctgcggaatg	13320
tggaagggtt	atggctgtag	ttgtgaccaa	ctccggaac	ccttgatgca	gtctgcggat	13380
gcatcaacgt	ttttaaacgg	gtttgcgggt	taagtgcagc	ccgtcttaca	ccgtgcggca	13440
caggcactag	tactgatgtc	gtctacaggg	cttttgatat	ttacaacgaa	aaagttgctg	13500
gttttgcaaa	gttccataaa	actaattgct	gtcgtctcca	ggagaaggat	gaggaaggca	13560
atztataga	ctctactttt	gtagttaaga	ggcatactat	gtctaactac	caacatgaag	13620
agactattta	taacttggtt	aaagattgtc	cagcggttgc	tgtccatgac	tttttcaagt	13680
ttagagtaga	tggtgacatg	gtaccacata	tatcacgtca	gcgtctaact	aaatacacia	13740
tggtgatatt	agtctatgct	ctacgtcatt	ttgatgaggg	taattgtgat	acattaaaag	13800
aaatactcgt	cacatacaat	tgctgtgatg	atgattatth	caataagaag	gattggtatg	13860
acttcgtaga	gaatcctgac	atcttacgcy	tatatgctaa	cttaggtgag	cgtgtacgcc	13920
aatcattatt	aaagactgta	caattctgcy	atgctatgcy	tgatgcaggc	attgtaggcy	13980
tactgacatt	agataatcag	gatcttaatg	ggaactggta	cgatttcggg	gatttcgtac	14040
aaagtagcacc	aggctgcgga	gttccatttg	tggattcata	ttactcattg	ctgatgccca	14100
tcctcacttt	gactagggca	ttggctgctg	agtcccatat	ggatgctgat	ctcgcaaaac	14160
cacttattaa	gtgggatttg	ctgaaatag	atthttacgga	agagagactt	tgtctcttcg	14220
accgttattt	taaatattgg	gaccagacat	accatcccaa	ttgtattaac	tgthttggatg	14280
ataggtgat	ccttcattgt	gaaaacttta	atgtgttatt	ttctactgtg	ttccacctta	14340
caagthtttg	accactagta	agaaaaatat	ttgtagatgg	tgthcccttt	gthgtthtcaa	14400
ctggatacca	thttcgtgag	ttaggagtgc	tacataatca	ggatgtaaac	ttacatagct	14460
cgctctcag	thtcaaggaa	ctthttagtg	atgctgctga	tccagctatg	catgcagctt	14520
ctggcaattt	attgctagat	aaacgcacta	catgctthtc	agtagctgca	ctaaacaaaca	14580
atgttgctth	tcaaacgtgc	aaaccoggtta	atthttaataa	agactthttat	gactthtgctg	14640
tgtctaaagg	thtctttaag	gaagggaagt	ctgttgaact	aaaacacttc	thtctthgctc	14700
aggatggcaa	cgctgctatc	agtgattatg	actattatcg	ttataatctg	ccaacaatgt	14760
gtgatatcag	acaactccta	thcgtagtth	aagthgttga	taataactth	gattgttacg	14820
atgttgctg	tattaatgcc	aaccaagtaa	tcgttaacaa	tctggataaa	tcagctggth	14880
tcccatttaa	taaatggggt	aaggctagac	thtattatga	ctcaatgagt	tatgaggatc	14940
aaagatgcact	thtccgctat	actaagcgtta	atgtcatccc	tactataact	caaatgaatc	15000
thaaagtatgc	cattagtgtca	aagaatagag	ctcgcaccgt	agctgggtgc	tctatctgtta	15060

-continued

---

gtactatgac	aaatagacag	tttcatcaga	aattattgaa	gtcaatagcc	gccactagag	15120
gagctactgt	ggtaattgga	acaagcaagt	tttacggtgg	ctggcataat	atgttaaaaa	15180
ctgtttacag	tgatgtagaa	actccacacc	ttatgggttg	ggattatcca	aaatgtgaca	15240
gagccatgcc	taacatgctt	aggataatgg	cctctctgtg	tcttgctcgc	aaacataaca	15300
cttgctgtaa	cttatcacac	cgtttctaca	ggttagctaa	cgagtgtgcg	caagtattaa	15360
gtgagatggt	catgtgtggc	ggctcactat	atgttaaacc	aggtggaaca	tcatccggtg	15420
atgtcacaac	tgcttatgct	aatagtgctt	ttaacatttg	tcaagctggt	acagccaatg	15480
taaatgcact	tctttcaact	gatggtaata	agatagctga	caagtatgtc	cgcaatctac	15540
aacacaggct	ctatgagtgt	ctctatagaa	atagggatgt	tgatcatgaa	ttcgtggatg	15600
agttttacgc	ttacctgcgt	aaacatttct	ccatgatgat	tctttctgat	gatgccgttg	15660
tgtgctataa	cagtaactat	gcggtcaag	gtttagtagc	tagcattaag	aactttaagg	15720
cagttcttta	ttatcaaaa	aatgtgttca	tgtctgaggc	aaaatgttgg	actgagactg	15780
accttactaa	aggacctcac	gaattttgct	cacagcatac	aatgctagtt	aaacaaggag	15840
atgattacgt	gtacctgcct	taccagatc	catcaagaat	attagggcga	ggctgttttg	15900
tcgatgatat	tgtcaaaa	gatggtacac	ttatgattga	aaggttcgtg	tcactggcta	15960
ttgatgctta	cccacttaca	aaacatccta	atcaggagta	tgctgatgtc	tttcaactgt	16020
atttacaata	cattagaaag	ttacatgatg	agcttactgg	ccacatggtg	gacatgtatt	16080
ccgtaatgct	aactaatgat	aacacctcac	ggactgggga	acctgagttt	tatgaggcta	16140
tgtcacacacc	acatacagtc	ttgcaggctg	taggtgcttg	tgtattgtgc	aattcacaga	16200
cttcaacttcg	ttgcgggtgcc	tgtattagga	gaccattcct	atggtgcaag	tgctgctatg	16260
accatgtcat	ttcaacatca	cacaaattag	tgttgtctgt	taatccctat	gtttgcaatg	16320
ccccagggtg	tgatgtcact	gatgtgacac	aactgtatct	aggaggatg	agctattatt	16380
gcaagtcaaca	taagcctccc	attagtttct	cattatgtgc	taatggtcag	gtttttggtt	16440
tatacaaaaa	catatgtgta	ggcagtgaca	atgtcactga	cttcaatgcg	atagcaacat	16500
gtgattggac	taatgctggc	gattacatac	tgccaacac	ttgtactgag	agactcaagc	16560
ttttcgagc	agaaacgctc	aaagccactg	aggaaacatt	taagctgtca	tatggatttg	16620
ctactgtacg	cgaagtactc	tctgacagag	aattgcatct	ttcatgggag	gttggaaaac	16680
ctagaccacc	attgaacaga	aactatgtct	ttactggtta	ccgtgtaact	aaaaatagta	16740
aagtacagat	tgagagtagc	acctttgaaa	aagtgacta	tggatgatgct	gttgtgtaca	16800
gaggactac	gacatacaag	ttgaatgttg	gtgattactt	tgtgttgaca	tctcactctg	16860
taatgccact	tagtgcacct	actctagtgc	cacaagagca	ctatgtgaga	attactggct	16920
tgtaccacaac	actcaacatc	tcagatgagt	tttctagcaa	tgttgcaaat	tatcaaaagg	16980
tcggcatgca	aaagtactct	acactocaag	gaccacctgg	tactggtaag	agtcattttg	17040
ccatcggaact	tgctctctat	taccatctg	ctcgcatagt	gtatcggca	tgctctcatg	17100
cagctgttga	tgccctatgt	gaaaaggcat	taaaatattt	gccatagat	aaatgtagta	17160
gaatcaatacc	tgcgcggtgcg	cgcgtagagt	gttttgataa	attcaaagtg	aattcaacac	17220
tagaacagta	tgttttctgc	actgtaaatg	cattgccaga	aacaactgct	gacattgtag	17280
tctttgatga	aatctctatg	gctactaatt	atgacttgag	tgttgcfaat	gctagacttc	17340

-continued

---

gtgcaaaaca ctacgtctat attggcgatc ctgctcaatt accagccccc cgcacattgc	17400
tgactaaagc cacactagaa ccagaatatt ttaattcagt gtgcagactt atgaaaacaa	17460
taggtccaga catgttcctt ggaactgtgc gccgttgtcc tgctgaaatt gttgacactg	17520
tgagtgcctt agtttatgac aataagctaa aagcacacaa ggataagtca gctcaatgct	17580
tcaaagtgtt ctacaaggtt gttattacac atgatgtttc atctgcaatc aacagacctc	17640
aaataggcgt tgtaagagaa tttcttacac gcaatcctgc ttggagaaaa gctgttttta	17700
tctcacctta taattcacag aacgctgtag cttcaaaaat cttaggattg cctacgcaga	17760
ctgttgattc atcacagggt tctgaatag actatgtcat attcacacaa actactgaaa	17820
cagcacactc ttgtaatgtc aaccgcttca atgtggctat cacaagggca aaaattggca	17880
ttttgtgcat aatgtctgat agagatcttt atgacaaaact gcaatttaca agtctagaaa	17940
taccacgtcg caatgtggct acattacaag cagaaaatgt aactggactt tttaaggact	18000
gtagtaagat cattactggt cttcatccta cacaggcacc tacacacctc agcgttgata	18060
taaagttcaa gactgaagga ttatgtgttg acataccagg cataccaaag gacatgacct	18120
accgtagact catctctatg atgggtttca aaatgaatta ccaagtcaat ggttacccta	18180
atatgtttat caccgcgcaa gaagctattc gtcacgttcg tgctgtgatt ggctttgatg	18240
tagagggctg tcatgcaact agagatgctg tgggtactaa cctacctctc cagctaggat	18300
tttctacagg tgtaactta gtagctgtac cgactggtta tgttgacact gaaaataaca	18360
cagaattcac cagagttaat gcaaaacctc caccagggtga ccagttttaa catcttatac	18420
cactcatgta taaaggcttg ccctggaatg tagtgcgtat taagatagta caaatgctca	18480
gtgatacact gaaaggattg tcagacagag tcgtgttcgt cctttgggcg catggctttg	18540
agcttacatc aatgaagtac tttgtcaaga ttggacctga aagaacgtgt tgcctgtgtg	18600
acaaacgtgc aactgtcttt tctacttcat cagatactta tgcctgctgg aatcattctg	18660
tgggttttga ctatgtctat aaccattta tgattgatgt tcagcagtgg ggctttacgg	18720
gtaacctta gagtaacct gaccaacatt gccagggtaca tggaaatgca catgtggcta	18780
gttgtgatgc tatcatgact agatgtttag cagtccatga gtgctttggt aagcgcgttg	18840
attggtctgt tgaatacct attataggag atgaactgag ggtaattct gcttcagaa	18900
aagtacaaca catggtgtg aagtctgcat tgcttgcgta taagtttoca gttcttcatg	18960
acattggaaa tccaaaggct atcaagtgtg tgcctcaggc tgaagtagaa tggaaattct	19020
acgatgctca gccatgtagt gacaaagctt acaaaataga ggagctcttc tattcttatg	19080
ctacacatca cgataaatc actgatggtg tttgtttggt ttggaattgt aacgttgatc	19140
gttaccacgc caatgcaatt gtgtgtagggt ttgacacaag agtcttgta aactgaaact	19200
taccaggctg tgatgggtgt agttttagtg tgaataagca tgcattccac actccagctt	19260
tcgataaaag tgcatttact aatttaaagc aattgccttt cttttactat tctgatagtc	19320
cttgtgagtc tcatggcaaa caagtgtgt cggatattga ttatgttoca ctcaaatctg	19380
ctacgtgtat tacacgatgc aatttaggtg gtgctgtttg cagacacctc gcaaatgagt	19440
accgacagta cttggatgca tataaatga tgatttctgc tggatttagc ctatggattt	19500
acaaacaatt tgatacttat aacctgtgga atacatttac caggttacag agtttagaaa	19560
atgtggctta taatgttgtt aataaaggac actttgatgg acacgccggc gaagcacctg	19620

-continued

---

tttccatcat taataatgct gtttacacaa aggtagatgg tattgatgtg gagatctttg	19680
aaaataagac aacacttcct gttaatgttg catttgagct ttgggctaag cgtaacatta	19740
aaccagtgcc agagattaag atactcaata atttgggtgt tgatatcgct gctaactactg	19800
taatctggga ctacaaaaga gaagcccag cacatgtatc tacaataggt gtctgcacaa	19860
tgactgacat tgccaagaaa cctactgaga gtgcttgttc ttcacttact gtcttgtttg	19920
atggtagagt ggaaggacag gtagacctt ttagaaacgc ccgtaatggg gttttaataa	19980
cagaaggttc agtcaaaggt ctaacacctt caaagggacc agcacaagct agcgtcaatg	20040
gagtcacatt aattggagaa tcagtaaaa cacagttaa ctactttaag aaagtagacg	20100
gcattattca acagttgcct gaaacctact ttactcagag cagagactta gaggatttta	20160
agcccagatc acaaattgaa actgactttc tcgagctcgc tatggatgaa ttcatacagc	20220
gatataagct cgagggctat gccttcgaac acatcgttta tggagatttc agtcatggac	20280
aacttgcggt tcttcattta atgataggct tagccaagcg ctcacaagat tcaccactta	20340
aattagagga ttttatccct atggacagca cagtgaaaa ttacttcata acagatgcgc	20400
aaacaggttc atcaaatgt gtgtgttctg tgattgatct tttacttgat gactttgtcg	20460
agataataaa gtcacaagat ttgtcagtga tttcaaaagt ggtcaagggt acaattgact	20520
atgctgaaat ttcattcatg ctttgggtga aggatggaca tgttgaacc tctaccctaa	20580
aactacaagc aagtcaagcg tggcaaccag gtgttgcgat gcctaacttg tacaagatgc	20640
aaagaatgct tcttgaaaag tgtgacctc agaattatgg tgaatgct gttataccaa	20700
aaggaataat gatgaatgtc gcaaagtata ctcaactgtg tcaatactta aatacactta	20760
ctttagctgt accctacaac atgagagtta ttcactttgg tgctggctct gataaaggag	20820
ttgcaccagc tacagctgtg ctcagacaat ggttgccaac tggcacacta cttgtcgatt	20880
cagatcttaa tgacttctc tccgacgag attctacttt aattggagac tgtgcaacag	20940
tacatacggc taataaatgg gaccttatta ttagcgatat gtatgacctt aggaccaaac	21000
atgtgacaaa agagaatgac tctaaagaag ggtttttcac ttatctgtgt ggatttataa	21060
agcaaaaact agccctgggt ggttctatag ctgtaaagat aacagagcat tcttggatg	21120
ctgaccttta caagcttatg ggccatttct catggtggac agcttttgtt acaaatgtaa	21180
atgcatcctc atcggaagca tttttaattg gggctaacta tcttggcaag ccgaaggaa	21240
aaattgatgg ctataccatg catgctaact acattttctg gaggaacaca aatcctatcc	21300
agttgtcttc ctattcactc tttgacatga gcaaatttcc tcttaatta agaggaactg	21360
ctgtaatgct tcttaaggag aatcaaatca atgatatgat ttattctctt ctggaaaaag	21420
gtaggcttat cattaagaaa aacaacagag ttgtggtttc aagtgatatt cttgttaaca	21480
actaaacgaa catgtttatt ttcttattat ttcttactct cactagtggg agtgacctg	21540
accggtgcac cacttttgat gatgttcaag ctcctaatta cactcaacat acttcatcta	21600
tgaggggggt ttactatcct gatgaaattt ttagatcaga cactctttat ttaactcagg	21660
atttatttct tccattttat tctaattgta cagggtttca tactattaat catacgtttg	21720
gcaaccctgt catacctttt aaggatggta tttattttgc tgccacagag aatcaaatg	21780
ttgtccgtgg ttgggttttt ggttctacca tgaacaacaa gtcacagtcg gtgattatta	21840
ttaacaattc tactaatggt gttatcacgag catgtaactt tgaattgtgt gacaaccctt	21900

-continued

---

tctttgctgt	ttctaacc	atgggtac	agacacatac	tatgatattc	gataatgcat	21960
ttaattgcac	tttcgagtac	atatctgatg	ccttttcgct	tgatgtttca	gaaaagtcag	22020
gtaattttaa	acacttacga	gagtttgtgt	ttaaaaataa	agatgggttt	ctctatgttt	22080
ataagggcta	tcaacctata	gatgtagttc	gtgatctacc	ttctggtttt	aacactttga	22140
aacctatfff	taagttgcct	cttggattta	acattacaaa	ttttagagcc	attcttacag	22200
ccttttcacc	tgctcaagac	atftggggca	cgfcagctgc	agcctatfff	gftggctatt	22260
taaagccaac	tacatttatg	ctcaagtatg	atgaaaatgg	tacaatcaca	gatgctgttg	22320
attgttctca	aaatccactt	gctgaactca	aatgctctgt	taagagcttt	gagattgaca	22380
aaggaattta	ccagacctct	aatttcaggg	ttgttccctc	aggagatgft	gtgagattcc	22440
ctaataattac	aaacttgtgt	ccttttgtag	aggtttttaa	tgctactaaa	ttcccttctg	22500
tctatgcatg	ggagagaaaa	aaaatttcta	attgtgttgc	tgattactct	gtgctctaca	22560
actcaacatt	tttttcaacc	tttaagtgtc	atggcgtttc	tgccactaag	ttgaatgatc	22620
tttgcctctc	caatgtctat	gcagattcctt	ttgtagtcaa	gggagatgat	gtaagacaaa	22680
tagcgcagcag	acaaactggg	gttattgtctg	attataatta	taaattgcca	gatgatttca	22740
tgggttgtgt	ccttgccttg	aatactagga	acattgatgc	tacttcaact	ggtaattata	22800
attataaata	taggtatcctt	agacatggca	agcttaggcc	ccttgagaga	gacatatcta	22860
atgtgccttt	ctcccctgat	ggcaaacctt	gcaccccacc	tgctcttaat	tgttattggc	22920
cattaaatga	ttatggtttt	tacaccacta	ctggcattgg	ctaccaacct	tacagagttg	22980
tagtactttc	ttttgaactt	ttaaatgcac	cggccacggt	ttgtggacca	aaattatcca	23040
ctgaccttat	taagaaccag	tgtgtcaatt	ttaattttaa	tggactcact	ggtactggty	23100
tgtaactcc	ttcttcaaag	agatttcaac	catttcaaca	atftggcctg	gatgfttctg	23160
atftcactga	ttccgttctga	gatcctaaaa	catctgaaat	attagacatt	tcaccttgcct	23220
cftttggggg	tgtaagtgtta	attacacctg	gaacaaatgc	ttcatctgaa	gftgctgttc	23280
tatatcaaga	tgtaactgc	actgatgttt	ctacagcaat	tcatgcagat	caactcacac	23340
cagcttggcg	catatattct	actggaaaca	atgtattcca	gactcaagca	ggctgtctta	23400
taggagctga	gcatgtcgac	acttcttatg	agtgcgacat	tcctattgga	gctggcattt	23460
gtgctagtta	ccatacagtt	tctttattac	gtagtactag	ccaaaaatct	attgtggcctt	23520
atactatgtc	tttaggtgct	gatagttcaa	ttgcttactc	taataacacc	attgctatac	23580
ctactaactt	ttcaattagc	attactacag	aagtaatgcc	tgftttctatg	gctaaaacct	23640
ccgtagattg	taatatgtac	atctgcggag	attctactga	atgtgctaat	ttgcttctcc	23700
aatatggtag	cftttgcaca	caactaaatc	gtgcactctc	aggtattgct	gctgaacagg	23760
atcgcaacac	acgtgaagtg	ttcgctcaag	tcaacaaat	gtacaaaacc	ccaactttga	23820
aatatfttgg	tggftttaat	ttttcacaaa	tattacctga	ccctctaaag	ccaactaaga	23880
ggtctfttat	tgaggacttg	ctcttataata	aggtgacact	cgctgatgct	ggcttcatga	23940
agcaatatgg	cgaatgccta	ggtgatatta	atgctagaga	tctcatttgt	gcgcagaagt	24000
tcaatggact	tacagtgttg	ccacctctgc	tcaactgatga	tatgattgct	gcctacactg	24060
ctgctctagt	tagtggact	gccactgctg	gatggacatt	tgggtgctggc	gctgctcttc	24120
aaataccttt	tgctatgcaa	atggcatata	ggttcaatgg	cattggagtt	acccaaaatg	24180

-continued

---

ttctctatga gaaccaaaaa caaatcgcca accaatttaa caaggcgatt agtcaaatc	24240
aagaatcact tacaacaaca tcaactgcat tgggcaagct gcaagacgtt gtaaccaga	24300
atgctcaagc attaaacaca cttgttaaac aacttagctc taattttggg gcaatttcaa	24360
gtgtgctaaa tgatattcctt tcgcgacttg ataaagtcca ggcggaggta caaattgaca	24420
ggtaattac aggcagactt caaagccttc aaacctatgt aacacaacaa ctaatcaggg	24480
ctgctgaaat cagggtctct gctaattctg ctgctactaa aatgtctgag tgtgttcttg	24540
gacaatcaaa aagagtgac ttttgtggaa agggctacca ccttatgtcc tccccacaag	24600
cagccccgca tgggtgtgtc ttcctacatg tcaagtatgt gccatcccag gagaggaact	24660
tcaccacagc gccagcaatt tgtcatgaag gcaaagcata cttccctcgt gaagggtgtt	24720
ttgtgtttaa tggcacttct tggtttatta cacagaggaa cttcttttct ccacaaataa	24780
ttactacaga caatacattt gtctcaggaa attgtgatgt cgttattggc atcattaaca	24840
acacagttta tgatcctctg caacctgagc tcgactcatt caaagaagag ctggacaagt	24900
acttcaaaaa tcatacatca ccagatgttg atcttggcga catttcaggc attaacgctt	24960
ctgtcgtcaa cattcaaaaa gaaattgacc gcctcaatga ggtcgtctaaa aatttaaatg	25020
aatcactcat tgaccttcaa gaattgggaa aatatgagca atatattaaa tggccttggg	25080
atgtttggct cggtctcatt gctggactaa ttgccatcgt catggttaca atcttgcttt	25140
gttgcattgac tagttgttgc agttgcctca aggggtgatg ctcttgggtt tcttgcgca	25200
agtttgatga ggatgactct gagccagttc tcaaggtgtg caaattacat tacacataaa	25260
cgaacttatg gatttgttta tgagattttt tactcttggg tcaattactg cacagccagt	25320
aaaaattgac aatgcttctc ctgcaagtac tgttcattgct acagcaacga taccgctaca	25380
agcctcactc ctttctggat ggcttgttat tggcgttcca tttcttctg tttttcagag	25440
cgctacccaa ataattgcgc tcaataaaaag atggcagcta gccctttata agggcttcca	25500
gttcaattgc aatttactgc tgctatttgt taccatctat tcacatcttt tgcctgtcgc	25560
tgaggatgag gaggcgcaat ttttgtacct ctatgccttg atatattttc tacaatgcat	25620
caacgcatgt agaattatta tgagatgttg gctttgttgg aagtgcacaa ccaagaaccc	25680
attactttat gatgccaact actttgttgg ctggcacaca cataactatg actactgtat	25740
accatataac agtgctacag atacaattgt cgttactgaa ggtgacggca tttcaacacc	25800
aaaactcaaa gaagactacc aaattggtgg ttattctgag gataggcact cagggtttaa	25860
agactatgct gttgtacatg gctatttcac cgaagtttac taccagcttg agtctacaca	25920
aattactaca gacactggta ttgaaaatgc tacattcttc atctttaaca agcttgttaa	25980
agaccaccg aatgtgcaaa tacacacaat cgacggctct tcaggagttg ctaatccagc	26040
aatggatcca atttatgatg agccgacgac gactactagc gtgccttctg aagcacaaga	26100
aagtgatgac gaacttatgt actcattcgt ttcggaagaa acaggtagct taatagttaa	26160
tagogtactt ctttttcttg ctttcgtggg attcttgcta gtcacactag ccatccttac	26220
tcgctctcga ttgtgtcgt actgctgcaa tattgttaac gtgagtttag taaaaccaac	26280
ggtttacgct tactcgcgtg ttaaaaaatc gaactcttct gaaggagttc ctgatcttct	26340
ggtctaaacg aactaactat tattattatt ctgtttggaa cttaacatt gcttatcatg	26400
gcagacaacg gtactattac cgttgaggag cttaacaac tcctggaaca atggaaccta	26460

-continued

---

gtaataggtt	tcctattcct	agcctggatt	atgttactac	aatttgcta	ttctaactcg	26520
aacaggtttt	tgacataat	aaagcttggt	ttcctctggc	tcttgtggcc	agtaacactt	26580
gcttgttttg	tgcttgctgc	tgtctacaga	attaattggg	tgactggcgg	gattgcgatt	26640
gcaatggcct	gtattgtag	cttgatgtgg	cttagctact	tcgttgcttc	cttcaggctg	26700
tttgctcgta	cccgctcaat	gtggtcattc	aaccagaaa	caaacattct	tctcaatgtg	26760
cctctccggg	ggacaattgt	gaccagaccg	ctcatggaaa	gtgaacttgt	cattggtgct	26820
gtgatcattc	gtggtcactt	gcgaatggcc	ggacaccccc	tagggcgctg	tgacattaag	26880
gacctgcaa	aagagatcac	tgtggctaca	tcacgaacgc	tttcttatta	caaattagga	26940
gcgtcgacg	gtgtaggcac	tgattcaggt	tttgctgcat	acaaccgcta	ccgtattgga	27000
aactataaat	taatacaga	ccacgccggt	agcaacgaca	atattgcttt	gctagtacag	27060
taagtgaaa	cagatgtttc	atcttgttga	cttccaggtt	acaatagcag	agatattgat	27120
tatcattatg	aggactttca	ggattgctat	ttggaatctt	gacgttataa	taagttcaat	27180
agtgagaaa	ttatttaagc	ctctaactaa	gaagaattat	tcggagttag	atgatgaaga	27240
acctatggag	ttagattatc	cataaaacga	acatgaaaat	tattctcttc	ctgacattga	27300
ttgtatttac	atcttgcgag	ctatatcact	atcaggagtg	tgtagaggt	acgactgtac	27360
tactaaaaga	accttgccca	tcaggaacat	acgagggcaa	ttcaccattt	cacctcttg	27420
ctgacaataa	atttgcaact	acttgcaact	gcacacactt	tgcttttgct	tgtgctgacg	27480
gtactcgaca	tacctatcag	ctgcgtgcaa	gatcagtttc	acaaaactt	ttcatcagac	27540
aagaggaggt	tcaacaagag	ctctactcgc	cactttttct	cattggtgct	gctctagtat	27600
ttttaatact	ttgcttcacc	attaagagaa	agacagaatg	aatgagctca	ctttaattga	27660
cttctatttg	tgctttttag	cctttctgct	attccttggt	ttaataatgc	ttattatatt	27720
ttggttttca	ctcgaatcc	aggatctaga	agaacctgtg	accaaagtct	aaacgaacat	27780
gaaacttctc	attgttttga	cttgtatttc	tctatgcagt	tgcatatgca	ctgtagtaca	27840
gcgtgtgca	tctaataaac	ctcatgtgct	tgaagatcct	tgtaaggtag	aacactaggg	27900
gtaactacta	tagcactgct	tgcttttgct	ctctaggaaa	ggttttacct	tttcatagat	27960
ggcacactat	ggttcaaaca	tgcacaccta	atgttactat	caactgtcaa	gatccagctg	28020
gtggtgcgct	tatagctagg	tggttggtacc	ttcatgaagg	tcaccaaact	gctgcattta	28080
gagacgtact	tggtgtttta	aataaacgaa	caaattaaaa	tgtctgataa	tggaacccaa	28140
tcaaaccaac	gtagtgcccc	ccgcattaca	tttgggtggac	ccacagattc	aactgacaat	28200
aaccagaatg	gaggacgcaa	tggggcaagg	ccaaaacagc	gccgaccoca	aggtttacc	28260
aataaactg	cgcttgggtt	cacagctctc	actcagcatg	gcaaggagga	acttagattc	28320
cctcagggcc	agggcgcttc	aatcaacacc	aatagtggtc	cagatgacca	aattggctac	28380
taccgaagag	ctaccgacg	agttcgtggt	ggtgacggca	aaatgaaaga	gctcagcccc	28440
agatggtact	tctattacct	aggaactggc	ccagaagctt	cacttcccta	cggcgctaac	28500
aaagaaggca	tcgtatgggt	tgcaactgag	ggagccttga	atacacccaa	agaccacatt	28560
ggcaccgca	atcctaataa	caatgctgcc	accgtgctac	aacttcctca	aggaacaaca	28620
ttgcaaaaag	gcttctacgc	agagggaagc	agaggcgga	gtcaagcttc	ttctcgtctc	28680
tcatacagta	gtcgcggtaa	ttcaagaaat	tcaactcctg	gcagcagtag	gggaaattct	28740

-continued

---

```

cctgctcgaa tggctagcgg aggtggtgaa actgccctcg cgctattgct gctagacaga 28800
ttgaaccagc ttgagagcaa agtttctggt aaaggccaac aacaacaagg ccaaactgtc 28860
actaagaaat ctgctgctga ggcactctaaa aagcctcgcc aaaaacgtac tgccacaaaa 28920
cagtacaacg tcaactcaagc atttgggaga cgtggtccag aacaaccca aggaaatttc 28980
ggggaccaag acctaactcag acaaggaact gattacaac attggccgca aattgcacaa 29040
tttgctccaa gtgcctctgc attctttgga atgtcacgca ttggcatgga agtcacacct 29100
tcgggaacat ggctgactta tcatggagcc attaaattgg atgacaaaga tccacaattc 29160
aaagacaacg tcatactgct gaacaagcac attgacgcat acaaacatt cccaccaaca 29220
gagcctaaaa aggacaaaaa gaaaagact gatgaagctc agcctttgcc gcagagacaa 29280
aagaagcagc ccaactgtgac tcttcttctc gggctgaca tggatgattt ctccagacaa 29340
cttcaaaatt ccatgagtgg agcttctgct gattcaactc aggcataaac actcatgatg 29400
accacacaag gcagatgggc tatgtaaacy ttttcgcaat tccgtttacg atacatagtc 29460
tactcttggt cagaatgaat tctcgttaact aacagcaca agtaggttta gtaacttta 29520
atctcacata gcaatcttta atcaatggtt aacattaggg aggacttgaa agagccacca 29580
cattttcatc gaggccacgc ggagtacgat cgagggtaca gtgaataatg ctagggagag 29640
ctgcctatat ggaagagccc taatgtgtaa aattaatttt agtagtgcta tccccatgtg 29700
attttaatag cttcttagga gaatgac 29727

```

&lt;210&gt; SEQ ID NO 3

&lt;211&gt; LENGTH: 29791

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: SARS coronavirus

&lt;400&gt; SEQUENCE: 3

```

cgcccttcat taatacgact cactatagat attaggtttt tacctaccca ggaaaagcca 60
accaacctcg atctcttgta gatctgttct ctaaacgaac tttaaaatct gtgtagctgt 120
cgctcgctg catgcctagt gcacctacgc agtataaaca ataataaatt ttactgtcgt 180
tgacaagaaa cgagtaacte gtcctcttct tgcagactgc ttacggtttc gtcogtgtg 240
cagtcgatca tcagcatacc taggtttcgt cggggtgta ccgaaaggta agatggagag 300
cettgttctt ggtgtcaacg agaaaacaca cgtccaactc agtttgctg tccttcaggt 360
tagagacgtg ctagtgcgtg gcttcgggga ctctgtgtaa gaggccctat cggaggcacg 420
tgaacacctc aaaaatggca cttgtggtct agtagagctg gaaaaggcg tactgcccc 480
gettgaacag ccctatgtgt tcattaaacg ttctgatgcc ttaagcacca atcacggcca 540
caaggtcgtt gagctggttg cagaaatgga cggcattcag tacggtcgtg gcggtataac 600
actgggagta ctcgtgccac atgtgggca aacccaatt gcataccgca atgttcttct 660
tcgtaagaac ggtaataagg gagccggtg tcatagctat ggcacgac taaagtetta 720
tgacttaggt gacgagcttg gcaactgatc cattgaagat tatgaacaaa actggaacac 780
taagcatggc agtgggtcac tccgtgaact cactcgtgag ctcaatggag gtgcagtcac 840
tcgctatgtc gacaacaatt tctgtggccc agatgggtac cctcttgatt gcatcaaaga 900
ttttctcgca cgcgcgggca agtcaatgtg cactctttcc gaacaacttg attacatcga 960
gtcgaagaga ggtgtctact gctgcccgtg ccatgagcat gaaattgcct ggttactga 1020

```



-continued

---

gcgctctgat aagagctacg agcaccagac acccttcgaa attaagagtg ccaagaaatt	1080
tgacactttc aaaggggaat gcccaaagtt tgtgtttcct cttactcaa aagtcaaagt	1140
cattcaacca cgtgttgaaa agaaaagac tgagggtttc atggggcgta tacgctctgt	1200
gtaccctgtt gcatctccac aggagtgtaa caatatgcac ttgtctacct tgatgaaatg	1260
taatcattgc gatgaagttt catggcagac gtgcgacttt ctgaaagcca cttgtgaaca	1320
ttgtggcact gaaaatthag ttattgaagg acctactaca tgtgggtacc tacctactaa	1380
tgctgtagtg aaaatgccat gtcctgcctg tcaagaccca gagattggac ctgagcatag	1440
tgttgcagat taccacaacc actcaaactc tgaactcga ctccgcaagg gaggtaggac	1500
tagatgtttt ggaggctgtg tgtttgccta tgttggctgc tataataagc gtcctactg	1560
ggttcctcgt gctagtctgt atattggctc aggacatact ggaattactg gtgacaatgt	1620
ggagacctg aatgaggatc tccttgagat actgagtcgt gaacgtgta acattaacat	1680
tgttggcgat tttcatttga atgaagaggt tgccatcatt ttggcatcct tctctgcttc	1740
tacaagtgcc tttattgaca ctataagag tcttgattac aagctttca aaaccattgt	1800
tgagtcctgc ggtaactata aagttaccaa gggaaagccc gtaaaagtg cttggaacat	1860
tgacaacag agatcagttt taacaccact gtgtggtttt ccctcacagg ctgctggtgt	1920
tatcagatca atttttgcgc gcacacttga tgcagcaaac cactcaattc ctgatttgca	1980
aagagcagct gtcaccatac ttgatggat ttctgaacag tcattacgtc ttgtcgacgc	2040
catggtttat acttcagacc tgctcaccia cagtgtcatt attatggcat atgtaactgg	2100
tggtccttga caacagactt ctcagtggtt gtctaactct ttgggcacta ctgttgaaaa	2160
actcaggcct atctttgaat ggattgagc gaaacttagt gcaggagttg aattttctca	2220
ggatgcttg gagattctca aatttctcat tacagggttt tttgacatcg tcaagggtca	2280
aatacaggtt gcttcagata acatcaagga ttgtgtaaaa tgcttcattg atgttgtaa	2340
caaggcactc gaaatgtgca ttgatcaagt cactatcgct ggcgcaaagt tgcgatcact	2400
caacttaggt gaagtcttca tcgctcaaag caagggactt taccgtcagt gtatacgtgg	2460
caaggagcag ctgcaactac tcatgctct taaggcacca aaagaagtaa cctttcttga	2520
aggtgattca catgacacag tacttacctc tgaggaggtt gttctcaaga acggtgaact	2580
cgaagcactc gagaccccg ttgatagctt cacaatgga gctatcgttg gcacaccagt	2640
ctgtgtaaat ggctcatgc tcttagagat taaggacaaa gaacaatact gcgcattgtc	2700
tcctggttta ctggctacaa acaatgtcct tcgcttaaaa ggggggtgcac caattaaagg	2760
tgtaaccttt ggagaagata ctgtttgga agttcaaggt tacaagaatg tgagaatcac	2820
atgtgagctt gatgaacgtg ttgacaaagt gcttaatgaa aagtgtctctg tctacactgt	2880
tgaatccggt accgaagtta ctgagtttgc atgtgttgta gcagaggctg ttgtgaagac	2940
tttacaacca gtttctgac tccttaccaa catgggtatt gatcttgatg agtggagtgt	3000
agctacattc tacttatttg atgatgctgg tgaagaaaac ttttcatcac gtatgtattg	3060
ttccttttac cctccagatg aggaagaaga ggacgatgca gagtgtgagg aagaagaaat	3120
tgatgaaacc tgtgaacatg agtacggtac agaggatgat tatcaaggtc tcctctgga	3180
atgtgtgccc tcagctgaaa cagttogagt tgaggaagaa gaagaggaag actggctgga	3240
tgatactact gagcaatcag agattgagcc agaaccagaa cctacacctg aagaaccagt	3300

-continued

---

taatcagttt actggttatt taaaacttac tgacaatggt gccattaaat gtgttgacat	3360
cgtaaggag gcacaaagt ctaatcctat ggtgattgta aatgctgcta acatacacct	3420
gaaacatggt ggtggtgtag cagggtgact caacaaggca accaatggg ccatgcaaaa	3480
ggagagtgat gattacatta agctaagt cctccttaca gtaggagggt cttgtttgct	3540
ttctggacat aatcttgcta agaagtgtct gcatgttggt ggacctaac taaatgcagg	3600
tgaggacatc cagcttctta aggcagcata tgaaaatttc aattcacagg acatcttact	3660
tgaccattg ttgtcagcag gcatatttgg tgctaaacca cttcagtctt tacaagtgtg	3720
cgtgcagacg gttcgtacac aggtttatat tgcagtcaat gacaaagctc tttatgagca	3780
ggttgcatg gattatcttg ataacctgaa gcctagagtg gaagcaccta aacaagagga	3840
gccaccaaac acagaagatt ccaaaactga ggagaaatct gtcgtacaga agcctgtcga	3900
tgtgaagcca aaaattaagg cctgcattga tgaggttacc acaactgag aagaaactaa	3960
gtttcttacc aataagttac tcttgtttgc tgatatcaat ggtaagcttt accatgattc	4020
tcagaacatg cttagagggt aagatatgtc tttccttgag aaggatgcac cttacatggt	4080
aggtgatggt atcactagtg gtgatcacac ttgtgttgta ataccctcca aaaaggctgg	4140
tggcactact gagatgctct caagagcttt gaagaaagt caggttgatg agtatataac	4200
cacgtacctt ggacaaggat gtgctggtta tacacttgag gaagctaaga ctgctcttaa	4260
gaaatgcaaa tctgcatttt atgtactacc ttcagaagca cctaagtcta aggaagagat	4320
tctaggaact gtatcctgga atttgagaga aatgcttgct catgctgaag agacaagaaa	4380
attaatgcct atatgcatgg atgttagagc cataatggca accatccaac gtaagtataa	4440
aggaatataa attcaagagg gcatcggtga ctatggtgct cgattcttct tttatactag	4500
taaaagacct gtagcttcta ttattacgaa gctgaactct ctaaatgagc cgcttgtcac	4560
aatgccaat gggtatgtga cacatggttt taatcttgaa gaggtgctgc gctgtatgctg	4620
ttctctaaa gctcctgccg tagtgtcagt atcatcacca gatgctgta ctacatataa	4680
tgataacct acttcgtcat caaagacatc tgaggagcac tttgtagaaa cagtttcttt	4740
ggctggctct tacagagatt ggtcctattc aggacagcgt acagagttag gtgtgaatt	4800
tcttaagcgt ggtgacaaaa ttgtgtacca cactctggag agccccgtcg agtttcatct	4860
tgacggtgag gttctttcac ttgacaaact aaagagtctc ttatccctgc gggagggtta	4920
gactataaaa gtgttcacaa ctgtggacaa cactaatctc cacacacagc ttgtggatat	4980
gtctatgaca tatggacagc agtttggcc aacatacttg gatggtgctg atgttcaaaa	5040
aattaaacct catgtaaac atgagggtaa gactttcttt gtactaccta gtgatgacac	5100
actacgtagt gaagctttc agtactacca tactcttgat gagagtttc ttggtaggta	5160
catgtctgct ttaaaccaca caaagaaatg gaaatttctc caagttggtg gtttaacttc	5220
aattaaatgg gctgataaca attgttattt gtctagtgtt ttattagcac ttcaacagct	5280
tgaagtcaaa ttcaatgac cagcaactca agaggcttat tatagagccc gtgctggtga	5340
tgctgctaac ttttgtgac tcatactcgc ttacagtaat aaaactgttg gcgagcttg	5400
tgatgtcaga gaaactatga cccatcttct acagcatgct aatttggaat ctgcaaagcg	5460
agttcttaat gtggtgtgta aacattgtgg tcagaaaact actaccttaa cgggtgtaga	5520
agctgtgatg tatatgggta ctctatctta tgataatctt aagacaggtg tttccattcc	5580

-continued

---

atgtgtgtgt ggtcgtgatg ctacacaata tctagtacaa caagagtctt cttttgttat	5640
gatgtctgca ccacctgctg agtataaatt acagcaaggt acattcttat gtgcgaatga	5700
gtacactggt aactatcagt gtggtcatta cactcatata actgctaagg agaccctcta	5760
tcgtattgac ggagctcacc ttacaagat gtcagagtac aaaggaccag tgactgatgt	5820
tttctacaag gaaacatctt aactacaac catcaagcct gtgtcgtata aactcgtatg	5880
agttacttac acagagattg aacccaaatt ggatgggtat tataaaaagg ataatgctta	5940
ctatacagag cagcctatag acctgtacc aactcaacca ttaccaaatg cgagttttga	6000
taatttcaaa ctacatggt ctaacacaaa atttgctgat gatttaaactc aaatgacagg	6060
cttcacaaa cgagcttcac gagagctatc tgtcacattc tcccagact tgaatggcga	6120
tgtagtggct attgactata gacactattc agcaggttc aagaaagggtg ctaaattact	6180
gcataagcca attgtttggc acattaacca ggctacaacc aagacaacgt tcaaaccaaa	6240
cacttggtgt ttacgtgtgc tttggagtac aaagccagta gatacttcaa attcatttga	6300
agttctggca gtagaagaca cacaaggaat ggacaatctt gcttggtgaa gtcaacaacc	6360
cacctctgaa gaagtagtgg aaaatcctac catacagaag gaagtcatag agtgtgacgt	6420
gaaaactacc gaagtgtgag gcaatgtcat acttaaacca tcagatgaag gtgttaaagt	6480
aacacaggag ttaggtcatg aggatcttat ggctgcttat gtggaaaaca caagcattac	6540
cattaagaaa cctaattgagc tttcactagc cttaggttta aaaacaattg cactcatgg	6600
tattgctgca attaatagtg ttccttggag taaaattttg gcttatgtca aaccattctt	6660
aggacaagca gcaattacaa catcaaatg cgctaagaga ttagcacaac gtgtgtttaa	6720
caattatag ccttatgtgt ttacattatt gttccaattg tgtactttta ctaaagtac	6780
caattctaga attagagctt cactacctac aactattgct aaaaatagtg ttaagagtgt	6840
tgctaaatta tgtttggatg ccggcattaa ttatgtgaag tcacccaaat tttctaaatt	6900
gttcacaatc gctatgtgac tattgttgtt aagtatttgc ttaggttctc taatctgtgt	6960
aactgctgct tttggtgtac tcttatctaa ttttgggtct ccttcttatt gtaatggcgt	7020
tagagaattg tatcttaatt cgtctaactg tactactatg gatttctgtg aaggttcttt	7080
tccttgacg atttgtttaa gtggattaga ctcccttgat tcttatccag ctcttgaac	7140
cattcaggtg acgatttcat cgtacaagct agacttgaca attttaggtc tggccgctga	7200
gtgggttttg gcatatatgt tgttcacaaa attcttttat ttattaggtc tttcagctat	7260
aatgcagggtg ttccttggct attttgctag tcatttcatc agcaattctt ggctcatgtg	7320
gtttatcatt agtattgtac aaatggcacc cgtttctgca atggttagga tgtacatctt	7380
ctttgcttct ttctactaca tatggaagag ctatgttcat atcatggatg gttgcacctc	7440
ttcgacttgc atgatgtgct ataagcgcaa tcgtgccaca cgcgttgagt gtacaactat	7500
tgtaaatggc atgaagagat ctttctatgt ctatgcaaat ggaggccgtg gcttctgcaa	7560
gactcacaat tggaaattgc tcaattgtga cacattttgc actggtagta cattcattag	7620
tgatgaagtt gctcgtgatt tgtcactcca gtttaaaaga ccaatcaacc ctactgacca	7680
gtcatcgtat attgttgata gtgttgctgt gaaaaatggc gcgcttcacc tctactttga	7740
caaggctggt caaaagacct atgagagaca tccgctctcc cttttgtca atttagacaa	7800
tttgagagct aacaacacta aaggttcaact gcctattaat gtcatagttt ttgatggcaa	7860

-continued

---

gtccaaatgc gacgagtctg cttctaagtc tgcttctgtg tactacagtc agctgatgtg	7920
ccaacctatt ctgttgcttg accaagttct tgtatcagac gttggagata gtactgaagt	7980
ttccgttaag atgtttgatg cttatgtcga caccttttca gcaactttta gtgttcctat	8040
ggaaaaactt aaggcacttg ttgctacagc tcacagcgag ttagcaaagg gtgtagcttt	8100
agatggtgtc ctttctacat tcggtgcagc tgcccgacaa ggtggtgttg ataccgatgt	8160
tgacacaaag gatgttattg aatgtctcaa actttcacat cactctgact tagaagtgc	8220
aggtgacagt tgtaacaatt tcatgctcac ctataataag gttgaaaaca tgacgcccag	8280
agatcttggc gcatgtattg actgtaatgc aaggcatatc aatgccaag tagcaaaaag	8340
tcacaatgtt tcaactcatc ggaatgtaaa agactacatg tctttatctg aacagctgcg	8400
taaaaaaatt cgtagtgtg ccaagaagaa caacatacct tttagactaa cttgtgctac	8460
aactagacag gttgtcaatg tcataactac taaaatctca ctcaagggtg gtaagattgt	8520
tagtacttgt tttaaactta tgcttaagtc cacattattg tgcggtcttg ctgcattggt	8580
ttgttatatc gttatgccag tacatacatt gtcaatccat gatggttaca caaatgaaat	8640
cattggttac aaagccattc aggatggtgt cactcgtgac atcatttcta ctgatgattg	8700
ttttgcaaat aaacatgctg gttttgacgc atggtttagc cagcgtggcg gttcatacaa	8760
aaatgacaaa agctgccttg tagtagctgc tatcattaca agagagattg gtttcatagt	8820
gcctggctta ccgggtactg tgctgagagc aatcaatggt gacttcttgc attttctacc	8880
tcgtgttttt agtgcgttg gcaacatttg ctacacacct tccaaactca ttgagtatag	8940
tgattttgc acctctgctt gcggtcttgc tgctgagtgt acaattttta aggatgctat	9000
gggcaaacct gtgccatatt gttatgacac taatttgcta gagggttcta tttcttatag	9060
tgagcttctg ccagacactc gttatgtgct tatggatggt tccatcatac agtttcctaa	9120
cacttacctg gagggttctg ttagagtagt aacaactttt gatgctgagt actgtagaca	9180
tggtacatgc gaaaggtcag aagtaggtat ttgcctatct accagtggtg gatgggttct	9240
taataatgag cattacagag ctctatcagg agttttctgt ggtgttgatg cgatgaatct	9300
catagctaac atctttactc ctcttgctga acctgtgggt gctttagatg tgtctgcttc	9360
agtagtggct ggtggtatta ttgccatatt ggtgacttgt gctgcctact actttatgaa	9420
attcagacgt gtttttggtg agtacaacca tgttgttct gctaatgcac tttgttttt	9480
gatgtcttcc actatactct gtctggtacc agcttacagc tttctgcogg gagtctactc	9540
agtcttttac ttgtacttga cattctatct caccaatgat gtttcattct tggctcacct	9600
tcaatggttt gccatgtttt ctctattgt gcctttttgg ataacagcaa tctatgtatt	9660
ctgtatctct ctgaagcact gccattggtt ctttaacaac tatcttagga aaagagtcac	9720
gtttaatgga gttacattta gtaccttcga ggaggctgct ttgtgtacct ttttgctcaa	9780
caaggaaatg tacctaaaat tgcgtagcga gacactgttg ccacttacac agtataacag	9840
gtatcttctc ctatataaca agtacaagta tttcagtggg gccttagata ctaccagcta	9900
tcgtgaagca gcttgctgcc acttagcaaa ggctctaaat gactttagca actcaggtgc	9960
tgatgttctc taccaaccac cacagacatc aatcacttct gctgttctgc agagtggttt	10020
taggaaaatg gcattcccgt caggcaaatg tgaaggggtc atggtacaag taacctgtgg	10080
aactacaact cttaatggat tgtggttggg tgacacagta tactgtccaa gacatgtcat	10140

-continued

---

ttgcacagca gaagacatgc ttaatcctaa ctatgaagat ctgctcattc gcaaatccaa	10200
ccatagcttt cttgttcagg ctggcaatgt tcaacttcgt gttattggcc attctatgca	10260
aaattgtctg cttaggctta aagttgatac ttctaaccct aagacacca agtataaatt	10320
tgccgtatc caacctggtc aacattttc agttctagca tgctacaatg gttcaccatc	10380
tggtgtttat cagtgtgcca tgagacctaa tcataccatt aaaggttcct tccttaatgg	10440
atcatgtggt agtgttggtt ttaacattga ttatgattgc gtgtcttctt gctatatgca	10500
tcatatggag cttccaacag gagtacacgc tgggtactgac ttagaaggta aattctatgg	10560
tccatttggt gacagacaaa ctgcacaggc tgcaggtaca gacacaacca taacattaaa	10620
tgttttggca tggctgtatg ctgctgttat caatgggat aggtggtttc ttaatagatt	10680
caccactact ttgaatgact ttaaccttgt ggcaatgaag tacaactatg aacctttgac	10740
acaagatcat gttgacatat tgggacctct ttctgctcaa acaggaattg ccgtcttaga	10800
tatgtgtgct gctttgaaag agctgctgca gaatgggatg aatggctgta ctatccttgg	10860
tagcactatt ttagaagatg agtttacacc atttgatggt gttagacaat gctctgggtg	10920
taccttccaa ggtaagttca agaaaattgt taagggcact catcattgga tgcttttaac	10980
ttcttgaca tcaactattga ttcttgttca aagtacacag tggctactgt tttctttgt	11040
ttacgagaat gctttcttgc catttactct tgggtattatg gcaattgctg catgtgctat	11100
gctgcttgtt aagcataagc acgcattcct gtgcttgttt ctgttacctt ctcttgcaac	11160
agttgcttac ttaatatggt tctacatgcc tgctagctgg gtgatgcgta tcatgacatg	11220
gcttgaattg gctgacacta gcttgtctgg ttataggctt aaggattgtg ttatgtatgc	11280
ttcagcttta gttttgctta ttctcatgac agctcgcact gtttatgatg atgctgctag	11340
acgtgtttg acaactgatga atgtcattac acttgtttac aaagtctact atggtaatgc	11400
tttagatcaa gctatttcca tgtgggcctt agttatttct gtaacctcta actattctgg	11460
tgctgttacg actatcatgt ttttagctag agctatagtg tttgtgtgtg ttgagtatta	11520
cccattgta tttattactg gcaacacctt acagtgtatc atgcttgttt attgtttctt	11580
aggctattgt tgctgctgct actttggcct tttctgttta ctcaaccgtt acttcaggct	11640
tactcttggg gtttatgact acttggctctc tacacaagaa tttaggtata tgaactccca	11700
ggggctttt cctcctaaga gtagtattga tgctttcaag cttaacatta agttgttggg	11760
tattggaggt aaaccatgta tcaaggttgc tactgtacag tctaaaatgt ctgacgtaaa	11820
gtgcacatct gtgggtactgc tctcggttct tcaacaactt agagtagagt catcttctaa	11880
attgtgggca caatgtgtac aactccacaa tgatattcct cttgcaaaag acacaactga	11940
agctttcgag aagatggttt ctcttttgtc tgttttgcta tccatgcagg gtgctgtaga	12000
cattaatagg ttgtgcgagg aatgctcga taaccgtgct actcttcagg ctattgcttc	12060
agaatttagt tctttaccat catatgccgc ttatgccact gcccaagagg cctatgagca	12120
ggctgtagct aatgggtgatt ctgaagctgt tctcaaaaag ttaagaaat ctttgaatgt	12180
ggctaaatct gagtttgacc gtgatgctgc catgcaacgc aagttggaaa agatggcaga	12240
tcaggctatg acccaaatgt acaaacaggc aagatctgag gacaagaggg caaaagtaac	12300
tagtgctatg caacaatgc tcttactat gottaggaag cttgataatg atgacttaa	12360
caacattatc aacaatgcgc gtgatggtg tggtccactc aacatcatac cattgactac	12420

-continued

---

agcagccaaa	ctcatggttg	ttgtccctga	ttatggtacc	tacaagaaca	cttgtgatgg	12480
taacaccttt	acatatgcat	ctgcaactctg	ggaaatccag	caagttgttg	atgocgatag	12540
caagattggt	caacttagtg	aaattaacat	ggacaattca	ccaaatttgg	cttggcctct	12600
tattgttaca	gctctaagag	ccaactcagc	tgtaaacta	cagaataatg	aactgagtcc	12660
agtagcacta	cgacagatgt	cctgtgcggc	tggtaccaca	caaacagctt	gtactgatga	12720
caatgcactt	gcctactata	acaattcgaa	gggaggtagg	tttgtgctgg	cattactatc	12780
agaccaccaa	gatctcaaat	gggctagatt	ccctaagagt	gatggtacag	gtacaattta	12840
cacagaactg	gaaccacctt	gtaggtttgt	tacagacaca	ccaaaagggc	ctaaagtga	12900
atacttgtag	ttcatcaaa	gcttaacaa	cctaaataga	ggtaggtg	tgggcagttt	12960
agctgtaca	gtacgtcttc	aggctgaaa	tgctacagaa	gtacctgcca	attcaactgt	13020
gctttccttc	tgtgcttttg	cagtagacc	tgctaaagca	tataaggatt	acctagcaag	13080
tgaggacaa	ccaatcacca	actgtgtgaa	gatgttgtgt	acacacactg	gtacaggaca	13140
ggcaattact	gtaacaccag	aagctaacat	ggaccaagag	tcctttggtg	gtgcttcag	13200
ttgtctgtat	gtagatgccc	acattgacca	tccaaatcct	aaaggattct	gtgactgaa	13260
aggtaaatg	gtccaaatg	ctaccacttg	tgctaattg	ccagtgggtt	ttacacttag	13320
aaacacagtc	tgtaccgtct	gcggaatgtg	gaaaggttat	ggctgtagtt	gtgaccaact	13380
ccggaacccc	ttgatgcagt	ctgcggtatg	atcaacgttt	ttaaaccggg	ttgocgtgta	13440
agtgcagccc	gtcttacacc	gtgocggaca	ggcactagta	ctgatgctgt	ctacagggt	13500
tttgatattt	acaacgaaaa	agttgctggt	tttgcaaggt	tcctaaaaac	taattgctgt	13560
cgcttcacag	agaaggtgta	ggaaggcaat	ttattagact	cttactttgt	agttaagagg	13620
catactatgt	ctaactacca	acatgaagag	actatttata	acttggttaa	agattgtcca	13680
gocggtgctg	tcctatgact	tttcaagttt	agagttagatg	gtgacatggt	accacatata	13740
tcacgtcagc	gtctaactaa	atacacaatg	gctgatttag	tctatgctct	acgtcatttt	13800
gatgagggta	attgtgatac	attaaaagaa	atactcgtca	catacaattg	ctgtgatgat	13860
gattatttca	ataagaagga	ttggtatgac	ttcgtagaga	atcctgacat	cttacgcgta	13920
tatgctaact	taggtgagcg	gttacgcaa	tcattattaa	agactgtaca	attctgcgat	13980
gctatgcgtg	atgcaggcat	gttaggcgta	ctgacattag	ataatcagga	tcttaattggg	14040
aactggtacg	atttcggtga	tttcgtacaa	gtagcaccag	gctgocgag	tcctattgtg	14100
gattcatatt	actcattgct	gatgcccac	ctcactttga	ctagggcatt	ggctgctgag	14160
tcccataatg	atgctgatct	cgcaaaacca	cttattaagt	gggatctgct	gaaatgat	14220
tttacggaag	agagactttg	tctcttcgac	cgttatttta	aatattggga	ccagacatac	14280
catcccaatt	gtattaactg	tttgatgat	agggtatcc	ttcattgtgc	aaactttaat	14340
gtgttatttt	ctactgtggt	tccacctaca	agttttggac	cactagtaag	aaaaatattt	14400
gtagatggtg	ttccttttgt	tgtttcaact	ggataccatt	ttcgtgagtt	aggagtctga	14460
cataatcag	atgtaaacct	acatagctcg	cgtctcagtt	tcaaggaact	tttagtgtat	14520
gctgctgac	cagctatgca	tgacgcttct	ggcaatttat	tgctagataa	acgcactaca	14580
tgcttttcag	tagctgcact	aacaacaat	gttgcttttc	aaactgtcaa	accogtaat	14640
tttaataaag	acttttatga	ctttgctgtg	tctaaagggt	tctttaagga	aggaagttct	14700

-continued

---

gttgaactaa aacacttctt ctttgctcag gatggcaacg ctgctatcag tgattatgac	14760
tattatcggtt ataatctgcc aacaatgtgt gatatcagac aactcctatt cgtagttgaa	14820
gttggtgata aatactttga ttggttacgat ggtggctgta ttaatgcaa ccaagtaac	14880
gttaacaate tggataaatc agctggtttc ccatttaata aatggggtaa ggctagactt	14940
tattatgact caatgagtta tgaggatcaa gatgcacttt tcgctgatac taagcgtaat	15000
gtcatcccta ctataactca aatgaatctt aagtatgcca ttagtgcaaa gaatagagct	15060
cgcaccgtag ctggtgtctc tatctgtagt actatgacaa atagacagtt tcatcagaaa	15120
ttattgaagt caatagccgc cactagagga gctactgtgg taattggaac aagcaagttt	15180
tacgggtggct ggcataatat gttaaaaact gtttacagtg atgtagaac tccacacctt	15240
atgggttggg attatccaaa atgtgacaga gccatgccta acatgcttag gataatggcc	15300
tctcttgttc ttgctcgcaa acataacact tgctgtaact tatcacaccg tttctacagg	15360
ttagctaacg agtgtgcgca agtattaagt gagatggtca tgtgtggcgg ctactatat	15420
gttaaaccag gtggaacatc atccgggtgat gctacaactg cttatgctaa tagtgtctt	15480
aacatttgtc aagctgttac agccaatgta aatgcacttc tttcaactga tggtaataag	15540
atagctgaca agtatgtccg caatctacaa cacaggctct atgagtgtct ctatagaaat	15600
agggatgttg atcatgaatt cgtggatgag ttttacgctt acctgcgtaa acatttctcc	15660
atgatgattc tttctgatga tgccggttggt tgctataaca gtaactatgc ggctcaaggt	15720
ttagtagcta gcattaagaa ctttaaggca gttctttatt atcaaaaca tgtgttcatg	15780
tctgaggcaa aatgttgac tgagactgac cttactaaag gacctcacga attttgctca	15840
cagcatacaa tgctagttaa acaaggagat gattacgtgt acctgcctta cccagatcca	15900
tcaagaatat taggcgcagc ctgttttctc gatgatattg tcaaaacaga tggtagactt	15960
atgattgaaa ggttcgtgtc actggctatt gatgcttacc cacttacaaa acatccta	16020
caggagtatg ctgatgtctt tcaactgtat ttacaataca ttagaaagtt acatgatgag	16080
cttactggcc acatgttga catgtattcc gtaatgctaa ctaatgataa cacctcacgg	16140
tactgggaac ctgagtttta tgaggctatg tacacaccac atacagtctt gcaggctgta	16200
ggtgcttgty tattgtgcaa ttcacagact tcaacttcgtt gcggtgcctg tattaggaga	16260
ccattcctat gttgcaagtg ctgctatgac catgtcattt caacatcaca caaattagtg	16320
ttgtctgtta atccctatgt ttgcaatgcc ccaggttgtg atgtcactga tgtgacacaa	16380
ctgtatctag gaggtatgag ctattattgc aagtcacata agcctcccat tagttttcca	16440
ttatgtgcta atggtcaggt ttttggttta tacaaaaaca catgtgtagg cagtgacaat	16500
gtcactgact tcaatgcgat agcaacatgt gattggacta atgctggcga ttacatactt	16560
gccaacactt gtactgagag actcaagctt ttcgcagcag aaacgctcaa agccactgag	16620
gaaacattta agctgtcata tggattgtct actgtacgag aagtactctc tgacagagaa	16680
ttgcatcttt catgggaggt tggaaaacct agaccaccat tgaacagaaa ctatgtcttt	16740
actggttacc gtgtaactaa aaatagtaaa gtacagattg gagagtacac ctttgaaaaa	16800
ggtgactatg gtgatgctgt tgtgtacaga ggtactacga catacaagtt gaatgttgg	16860
gattactttg tgttgacatc tcacactgta atgccactta gtgcacctac tctagtacca	16920
caagagcact atgtgagaat tactggcttg tacccaacac tcaacatctc agatgagttt	16980

---

-continued

---

tctagcaatg ttgcaaatta tcaaaaggtc ggcattgcaa agtactctac actccaagga 17040  
ccacctggta ctggtaagag tcattttgcc atcggacttg ctctctatta cccatctgct 17100  
cgcatagtg atacggcatg ctctcatgca gctgttgatg ccctatgtga aaaggcatta 17160  
aaatatttgc ccatagataa atgtagtaga atcatacctg cgcgtgctcg cgtagagtggt 17220  
tttgataaat tcaaagtga ttcaacacta gaacagtagt ttttctgac tgtaaatgca 17280  
ttgccagaaa caactgctga cattgtagtc tttgatgaaa tctctatggc tactaattat 17340  
gacttgagtg ttgtcaatgc tagacttcgt gcaaaacact acgtctatat tggcgatcct 17400  
gctcaattac cagcccccg cacattgctg actaaaggca cactagaacc agaataat 17460  
aattcagtg gcagacttat gaaaacaata ggtccagaca tgttccttgg aacttgtcgc 17520  
cgttgcctg ctgaaattgt tgacactgag agtgctttag tttatgaca taagctaaaa 17580  
gcacacaagg ataagtcagc tcaatgcttc aaaatgttct acaaagggtg tattacacat 17640  
gatgtttcat ctgcaatcaa cagacctcaa ataggcgttg taagagaatt tcttacacgc 17700  
aatcctgctt ggagaaaagc tgtttttatc tcaccttata attcacagaa cgctgtagct 17760  
tcaaaaatct taggattgcc tacgcagact gttgattcat cacagggttc tgaatatgac 17820  
tatgtcatat tcacacaaac tactgaaaca gcacactctt gtaatgcaa ccgcttcaat 17880  
gtggctatca caaggcaaa aattggcatt ttgtgcataa tgtctgatag agatctttat 17940  
gacaaactgc aatttacaag tctagaaata ccacgtcgca atgtggctac attacaagca 18000  
gaaaatgtaa ctggactttt taaggactgt agtaagatca ttactggtct tcatcctaca 18060  
caggcaccta cacacctcag cgttgatata aagttcaaga ctgaaggatt atgtgttgac 18120  
ataccaggca taccaaagga catgacctac cgtagactca tctctatgat gggtttcaaa 18180  
atgaattacc aagtcaatgg ttacccta atgtttatca cccgcgaaga agtattcgt 18240  
cacgttcgtg ctgggattgg ctttgatgta gagggctgtc atgcaactag agatgctgtg 18300  
ggtactaacc tactctcca gctaggattt tctacagggt ttaacttagt agctgtaccg 18360  
actggttatg ttgacactga aaataacaca gaattcacca gagttaatgc aaaacctcca 18420  
ccaggtgacc agtttaaca tcttatacca ctcatgtata aaggcttggc ctggaatgta 18480  
gtgctgatta agatagtaca aatgctcagt gatacactga aaggattgtc agacagagtc 18540  
gtgttcgtcc tttgggcgca tggctttgag cttacatcaa tgaagtactt tgtcaagatt 18600  
ggacctgaaa gaacgtgttg tctgtgtgac aaacgtgcaa cttgcttttc tacttcatca 18660  
gatacttatg cctgctgaa tcattctgtg ggttttgact atgtctataa cccatttatg 18720  
attgatgttc agcagtggg ctttacgggt aacctcaga gtaacctga ccaacattgc 18780  
caggtagatg gaaatgcaca tgtggctagt tgtgatgcta tcatgactag atgttttagca 18840  
gtccatgagt gctttgttaa gcgcgttgat tggctctgtg aataccctat tataggagat 18900  
gaaactgagg ttaattctgc ttgcagaaaa gtacaacaca tggttgtgaa gtctgcctc 18960  
ctggctgata agtttccagt tcttcatgac attggaaatc caaaggctat caagtgtgtg 19020  
cctcaggctg aagtagaatg gaagtctac gatgctcagc catgtagtga caaagcttac 19080  
aaaaatagag agctcttcta ttcttatgct acacatcacg ataaattcac tgatgggtgt 19140  
tgtttgtttt ggaattgtaa cgttgatcgt taccagcca atgcaattgt gtgtaggttt 19200  
gacacaagag tcttgcaaa cttgaactta ccaggctgtg atgggtgtag tttgtatgtg 19260



---

-continued

---

aataagcatg cattccacac tccagcttct gataaaagtg catttactaa tttaaagcaa 19320  
ttgcctttct ttactattc tgatagtcct tgtgagctc atggcaaaca agtagtgctg 19380  
gatattgatt atgttccact caaatctgct acgtgtatta cacgatgcaa ttaggtggt 19440  
gctgtttgca gacaccatgc aatgagtagc cgacagtact tggatgcata taatatgatg 19500  
atctctgctg gatttagcct atggatttac aaacaatttg atacttataa cctgtggaat 19560  
acatttacca ggttacagag tttagaaaat gtggcttata atgttgtaa taaaggacac 19620  
ttgatggac acgccggcga agcacctgtt tccatcatta ataagctgt ttacacaaag 19680  
gtagatggta ttgatgtgga gatctttgaa aataagacaa cacttcctgt taatgttgca 19740  
ttgagcttt gggctaagcg taacattaaa ccagtgcag agattaagat actcaataat 19800  
ttgggtgtg atatcgctgc taatactgta atctgggact ataaaagaga agccccagca 19860  
catgtatcta caataggtgt ctgcacaatg actgacattg ccaagaaacc tactgagagt 19920  
gcttgttctt cacttactgt cttgtttgat gtagagtg aaggacaggt agacctttt 19980  
agaaacgcc gtaatgtgt ttaataaca gaaggttcag tcaaaggtct aacaccttca 20040  
aagggaccag cacaagctag cgtcaatgga gtcacattaa ttggagaatc agtaaaaaca 20100  
cagtttaact actttaagaa agtagacgac attattcaac agttgcctga aacctactt 20160  
actcagagca gagacttaga ggattttaag cccagatcac aatggaaac tgactttctc 20220  
gagctcgcta tggatgaatt catacagcga tataagctcg agggctatgc cttogaacac 20280  
atcgtttatg gagatttcag tcatggacaa cttggcggtc ttcatttaat gataggctta 20340  
gccaaagcgt cacaagatc accacttaaa ttagaggatt ttatccctat ggacagcaca 20400  
gtgaaaatt acttcataac agatgcgcaa acaggttcat caaatgtgt gtgttctgtg 20460  
attgatctt tactttagta ctttgcgag ataataaagt cacaagattt gtcagtgatt 20520  
tcaaagtg tcaaggttac aattgactat gctgagattt cattcatgct ttggtgtaag 20580  
gacggacatg ttgaaacctt ctacccaaaa ctacaagcaa gtcaagcgtg gcaaccaggt 20640  
gttgcatgc ctaactgtga caagatgcaa agaagcttc ttgaaaagtg tgacctcag 20700  
aattatggtg aaaatgctgt tataccaaaa ggaataatga tgaatgtgc aaagtatact 20760  
caactgtgct aatacttaaa tacacttact ttagctgtac cctacaacat gagagttatt 20820  
cactttggtg ctggctctga taaaggagtt gcaccaggt cagctgtgct cagacaatg 20880  
ttgccaactg gcacactact tgtcgattca gatcttaatg acttctctc cgaocgagat 20940  
tctactttaa ttggagactg tgcaacagta catacggcta ataaatggga ccttattatt 21000  
agcgatatgt atgacctag gaccaaacat gtgacaaaag agaagactc taaagaagg 21060  
ttttcactt atctgtgtg atttataaag caaaaactag ccctgggtg ttctatagct 21120  
gtaaagataa cagagcatc ttggaatgct gaccttaca agcttatggg cactttctca 21180  
tgggtgacag cttttgttac aatgtaaat gcatcatcat cggaaagcatt ttaattggg 21240  
gctaactatc ttggcaagcc gaaggaacaa attgatggct ataccatgca tgctaactac 21300  
atcttctgga ggaacacaaa tcctatccag ttgtcttctt attcactctt tgacatgagc 21360  
aaatttctc ttaaattaag aggaactgct gtaatgtctc ttaaggagaa tcaaatcaat 21420  
gatatgattt attctcttct ggaaaaaggt aggttatca ttagagaaaa caacagagtt 21480  
gtggttcaa gtgatattct tgttaacaac taaacgaaca tgtttatctt cttattattt 21540

-continued

---

cttactctca	ctagtggtag	tgaccttgac	cggtgcacca	cttttgatga	tgttcaagct	21600
cctaattaca	ctcaacatac	ttcatctatg	agggggggtt	actatcctga	tgaattttt	21660
agatcagaca	ctctttat	aactcaggat	ttatttcttc	cattttatc	taatgttaca	21720
gggtttcata	ctattaatca	tacgtttggc	aacctgtca	tacctttta	ggatggtatt	21780
tattttgctg	coacagagaa	atcaaatggt	gtccgtggt	gggttttgg	ttctaccatg	21840
aacaacaagt	cacagtcggt	gattattatt	aacaattcta	ctaagtgtgt	tatagcagca	21900
tgtactttg	aattgtgtga	caacccttc	tttctgttt	ctaaacctat	gggtacacag	21960
acacatacta	tgatattcga	taatgcattt	aattgcactt	tcgagtacat	atctgatgcc	22020
ttttcgcttg	atgtttcaga	aaagtcaggt	aattttaaac	acttacgaga	gtttgtgttt	22080
aaaaataaag	atgggtttct	ctatgtttat	aagggctatc	aacctataga	tgtagtctgt	22140
gatctacctt	ctggttttaa	cactttgaaa	cctattttta	agttgcctct	tggtattaac	22200
attacaaatt	ttagagccat	tcttacagcc	ttttcacctg	ctcaagacat	ttggggcacg	22260
tcagctgcag	cctattttgt	tggtatttta	aagccaacta	catttatgct	caagtatgat	22320
gaaaatggta	caatcacaga	tgctgttgat	tgttctcaa	atccactgc	tgaactcaa	22380
tgctctgtta	agagctttga	gattgacaaa	ggaatttacc	agacctctaa	tttcaggggt	22440
gttccctcag	gagatgttgt	gagattccct	aatattacaa	acttgtgtcc	tttggagag	22500
gttttaaatg	ctactaaatt	cccttctgtc	tatgcatggg	agagaaaaaa	aatttcta	22560
tggtttgctg	attactctgt	gctctacaac	tcaacatttt	tttcaacctt	taagtgtctat	22620
ggcgtttctg	ccactaagtt	gaatgatctt	tgcttctcca	atgtctatgc	agattctttt	22680
gtagtcaagg	gagatgatgt	aagacaaata	gcgccaggac	aaactggtgt	tattgctgat	22740
tataattata	aattgccaga	tgatttcatg	ggttgtgtcc	ttgcttgga	tactaggaac	22800
attgatgcta	cttcaactgg	taattataat	tataaatata	ggtatcttag	acatggcaag	22860
cttaggccct	ttgagagaga	catatcta	gtgcccttct	cccctgatgg	caaaccttgc	22920
acccacctg	ctcttaattg	ttattggcca	ttaaatgatt	atggttttta	caccactact	22980
ggcattggct	accaacctta	cagagttgta	gtactttctt	ttgaactttt	aaatgcaccg	23040
gccacggttt	gtggaccaa	attatocact	gaccttatta	agaaccagtg	tgtcaatttt	23100
aattttaaatg	gactcactgg	tactgggtgtg	ttaactcctt	cttcaaagag	atttcaacca	23160
tttaacaat	ttggccgtga	tgtttctgat	ttcactgatt	ccgttcgaga	tcctaaaaca	23220
tctgaaatat	tagacatttc	accttctctt	tttgggggtg	taagtgtaat	tacacctgga	23280
acaaatgctt	catctgaagt	tgctgttcta	tatcaagatg	ttaactgcac	tgatgtttct	23340
acagcaattc	atgcagatca	actcacacca	gottggcgca	tatattctac	tgaaacaat	23400
gtattccaga	ctcaagcagg	ctgtcttata	ggagctgagc	atgtcgacac	ttcttatgag	23460
tgcgacattc	ctattggagc	tggtcattgt	gctagtacc	atcacgtttc	tttattacgt	23520
agtagtagcc	aaaaatctat	tgtggcttat	actatgtctt	taggtgctga	tagttcaatt	23580
gcttactcta	ataacaccat	tgctatacct	actaactttt	caattagcat	tactacagaa	23640
gtaatgcctg	tttctatggc	taaaacctcc	gtagattgta	atatgtacat	ctgcggagat	23700
tctactgaat	gtgctaattt	gcttctccaa	tatggtagct	tttgcacaca	actaaatcgt	23760
gcactctcag	gtattgctgc	tgaacaggat	cgcaacacac	gtgaagtgtt	cgctcaagtc	23820

-continued

---

aaacaaatgt	acaaaacccc	aactttgaaa	tatthttggtg	gtthttaatth	ttcacaaata	23880
ttacctgacc	ctctaaagcc	aactaagagg	tctthttattg	aggacttgct	ctthtaataag	23940
gtgacctcgc	ctgatgctgg	cttcatgaag	caatatggcg	aatgcctagg	tgatattaat	24000
gctagagatc	tcatttgctc	gcagaagttc	aatggactta	cagtgttgcc	acctctgctc	24060
actgatgata	tgattgctgc	ctacacggct	gctctagtta	gtggtactgc	cactgctgga	24120
tgacatttg	gtgctggcgc	tgctcttcaa	atacctthttg	ctatgcaaat	ggcatatagg	24180
ttcaatggca	ttggagttac	ccaaaatgth	ctctatgaga	acccaaaaaca	aatgcaccaac	24240
caatthtaaca	aggcgattag	tcaaattcaa	gaatcactta	caacaacatc	aactgcattg	24300
ggcaagctgc	aagacgttgt	taaccagaat	gctcaagcat	taaacacact	tgthtaacaa	24360
cttagctcta	atthttgctc	aatthcaagt	gtgctaaatg	atacctthtc	gctgacttgat	24420
aaagtcgagg	cggagggtaca	aattgacagg	ttaattacag	gcagacttca	aagccttcaa	24480
acctatgtaa	cacaacaact	aatcagggct	gctgaaatca	gggcttctgc	taatcttgct	24540
gctactaaaa	tgcttgatg	tgthcttgga	caatcaaaaa	gagthgactt	thgtggaaa	24600
ggctaccacc	ttatgctctt	cccacaagca	gccccgcatg	gtgthgtctt	cctacatgctc	24660
acgtatgctc	catcccagga	gaggaacttc	accacagcgc	cagcaatthg	tcatgaaggc	24720
aaagcatact	tcctctgctg	agggtgthtt	gtgthtaatg	gcacttcttg	gthttattaca	24780
cagaggaact	tctthttctc	acaaataatt	actacagaca	atacattthg	ctcaggaat	24840
thgtatgctc	ttattggcat	cattaacaac	acagthttatg	atcctctgca	acctgagctc	24900
gactcattca	aagaagagct	ggacaagtac	ttcaaaaatc	atacatcacc	agatgthgat	24960
cttgccgaca	thtcaggcat	taacgcttct	gtcgtcaaca	thcaaaaaga	aattgaccgc	25020
ctcaatgagg	tcgctaaaaa	thtaaatgaa	tcactcattg	accttcaaga	atthgggaaaa	25080
tatgagcaat	atattaaatg	gccttggtat	gthttggctc	gcttcttgct	tgactaat	25140
gccatcgtca	tggttacaat	cttgctthgt	tgcatgacta	gthgttgca	thgctcaag	25200
ggthgatgct	cttgthgttc	thgctgcaag	ththgatgagg	atgactctga	gccagthctc	25260
aaggthgtca	aattacatta	cacataaacg	aactthtgga	ththgtthtg	agatththth	25320
ctctthgatc	aattactgca	cagccagtaa	aaatthgaca	tgcttctct	gcaagtactg	25380
thcatgctac	agcaacgata	ccgctacaag	cctcactccc	thtcggatgg	ctthgttattg	25440
gcgthgcatt	tctthgctgt	thtcagagcg	ctacccaaat	aatthgcctc	aataaaagat	25500
ggcagctagc	ctthttataag	ggcttccagt	tcattthgca	ththactgctg	ctattthgtta	25560
ccatctattc	acatctthttg	ctthgtcgtg	caggtatgga	ggcgaatth	thgtacctct	25620
atgctthgat	atattthtcta	caatgcatca	acgcatgtag	aattattatg	agatgthggc	25680
ththgttgaa	gtgcaaatcc	aagaacctat	tactthttatga	tgccaaactac	ththgtthgtc	25740
ggcacacaca	taactatgac	tactgtatac	catataacag	thgtcacagat	acaattgctc	25800
ththctgaag	tgacggcatt	tcaacaccaa	aactcaaaga	agactaccaa	atthgtgtht	25860
atthctgagga	taggcactca	gthgtthaaag	actatgctctg	thgtacatggc	tattthcaccg	25920
aagththtacta	ccagctthgag	tctacacaaa	ththctacaga	cactgthtatt	gaaaatgctca	25980
catthctctcat	ctthtaacaag	ctthgtthaaag	accaccgaa	thgtgcaata	cacacaatcg	26040
acgthctctc	aggagthgtct	aatccagcaa	thgthtccaat	ththgtatgag	ccgacgacga	26100

-continued

---

ctactagcgt gcctttgtaa gcacaagaaa gtgagtacga acttatgtac tcattcgttt	26160
cggaagaaac aggtacgtta atagttaata gcgtacttct ttttcttgct ttcgtgggat	26220
tcttgctagt cacactagcc atccttactg cgcttcgatt gtgtgcgtac tgctgcaata	26280
ttgttaacgt gaggtttagta aaaccaacgg tttacgtcta ctccgctggt aaaaatctga	26340
actcttctga aggagttcct gatcttctgg tctaaccgaa ctaactatta ttattattct	26400
gtttggaact ttaacattgc ttatcatggc agacaacggc actattaccg ttgaggagct	26460
taaacaaactc ctggaacaat ggaacctagt aataggttcc ctattcctag cctggattat	26520
gttactacaa ttgacctatt ctaatcgaa caggtttttg tacataataa agcttgtttt	26580
cctctggctc ttgtggccag taacacttgc ttgttttggt cttgctgctg tctacagaat	26640
taattgggtg actggcggga ttgctgattgc aatggcttgt attgtaggct tgatgtggct	26700
tagctacttc gttgcttctc tcaggctggt tgctcgtacc cgctcaatgt ggtcattcaa	26760
cccagaaaca aacattcttc tcaatgtgcc tctccggggg acaattgtga ccagaccgct	26820
catgaaagt gaactgtca ttggtgctgt gatcattcgt ggtcacttgc gaatggccgg	26880
acacccccta gggcgtgtg acattaagga cctgcaaaa gagatcactg tggctacatc	26940
acgaacgctt tcttattaca aataggagc gtcgcagcgt gtaggcactg attcaggttt	27000
tgctgcatac aaccgctacc gtattgaaa ctataaatta aatacagacc acgccggtag	27060
caacgacaat attgctttgc tagtacagta agtgacaaca gatgttccat cttgttgact	27120
tccaggttac aatagcagag atattgatta tcattatgag gactttcagg attgctattt	27180
ggaatcttga cgttataata agttcaatag tgagacaatt atttaagcct ctaactaaga	27240
agaattattc ggagttagat gatgaagaac ctatggagtt agattatcca taaaacgaac	27300
atgaaaatta ttctcttctc gacattgatt gtatttccat cttgcgagct atatcactat	27360
caggagtgtg ttagaggtag gactgtacta ctaaaagaac cttgccatc aggaacatac	27420
gagggaatt caccatttca cctcttgct gacaataaat ttgcaactaac ttgcaactagc	27480
acacactttg cttttgcttg tgctgacggt actcgacata cctatcagct gctgcaaga	27540
tcagtttcc caaaactttt catcagacaa gaggaggttc aacaagagct ctactcgcca	27600
ctttttctca ttgttgctgc tctagtattt ttaatacttt gcttcacat taagagaaag	27660
acagaatgaa tgagctcact ttaattgact tctattttg ctttttagcc tttctgctat	27720
tccttgtttt aataatgctt attatatttt ggttttcaact cgaatccag gatctagaag	27780
aaccttgtag caaagtctaa acgaacatga aacttctcat tgttttgact tgtatttctc	27840
tatgcagttg catatgcact gtagtacagc gctgtgcatc taataaacct catgtgcttg	27900
aagatccttg taaggtacaa cactaggggt aatacttata gcaactgcttg gctttgtgct	27960
ctagaaagg ttttacctt tcatagatgg cacactatgg ttcaaacatg cacaccta	28020
gttactatca actgtcaaga tccagctggt ggtgcgctta tagctagggt ttggtacctt	28080
catgaaggtc accaaactgc tgcatttaga gacgtacttg ttgttttaa taaacgaca	28140
aattaaatg tctgataatg gaccccaatc aaaccaacgt agtgccccc gcattacatt	28200
tggtggacc acagattcaa ctgacaataa ccagaatgga ggacgcaatg gggcaaggcc	28260
aaaacagcgc cgacccaag gtttaoccaa taactactg ctttggttca cagctctcac	28320
tcagcatggc aaggaggaac ttagattccc togaggccag ggcgttccaa tcaacaccaa	28380

-continued

---

```

tagtgggtcca gatgacccaaa ttggtacta ccgaagagct acccgacgag ttcgtgggtgg 28440
tgacggcaaa atgaaagagc tcagccccag atggtacttc tattacctag gaactggccc 28500
agaagcttca cttccctacg gcgctaacaa agaaggcatc gtatgggttg caactgaggg 28560
agccttgaat acacccaaa accacattgg caccgcgaat cctaataaca atgctgccac 28620
cgtgctacaa cttcctcaag gaacaacatt gccaaaaggc ttctacgcag agggaagcag 28680
aggcggcagt caagcctctt ctgcctcctc atcacgtagt cgcggtaatt caagaaattc 28740
aactcctggc agcagtaggg gaaattctcc tgctcgaatg gctagcggag gtggtgaaac 28800
tgccctcgcg ctattgtgct tagacagatt gaaccagctt gagagcaaag tttctggtaa 28860
aggccaacaa caacaaggcc aaactgtcac taagaaatct gctgctgagg catctaaaaa 28920
gcctcgccaa aaacgtactg ccacaaaaca gtacaacgtc actcaagcat ttgggagacg 28980
tgggccagaa caaacccaag gaaatttcgg ggaccaagac ctaatcagac aaggaaactga 29040
ttacaaacat tggccgcaaa ttgcacaatt tgctccaagt gcctctgcat tctttggaat 29100
gtcacgcatt ggcattggaag tcacaccttc gggaaacatg ctgacttacc atggagccat 29160
taaattggat gacaaagatc cacaaattca agacaacgtc atactgtgta acaagcacat 29220
tgacgcatac aaaacattcc caccaacaga gcctaaaaag gacaaaaaga aaaagactga 29280
tgaagctcag cttttgccgc agagacaaaa gaagcagccc actgtgactc ttcttcctgc 29340
ggctgacatg gatgatttct ccagacaact tcaaaattcc atgagtggag cttctgctga 29400
ttcaactcag gcataaacac tcatgatgac cacacaaggc agatgggcta tgtaaacggt 29460
ttcgcaattc cgtttacgat acatagtcta ctcttggtgca gaatgaattc tcgtaactaa 29520
acagcacaag taggtttagt taactttaat ctcacatagc aatctttaat caatgtgtaa 29580
cattagggag gacttgaaag agccaccaca ttttcatcga ggccacgcgg agtacgatcg 29640
agggtagcgt gaataatgct agggagagct gcctatatgg aagagcccta atgtgtaaaa 29700
ttaattttag tagtgctatc cccatgtgat ttaaatagct tcttaggaga atgacaaaaa 29760
aaaaaaaaa aaaaaagggc gaattctgca g 29791

```

&lt;210&gt; SEQ ID NO 4

&lt;211&gt; LENGTH: 5450

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: SARS F X1 mutant

&lt;400&gt; SEQUENCE: 4

```

gtgtgctgga attcgccctt cgagtgcccg eggccgcatt gctgcctaca cggctgctct 60
agttagtggg actgccactg ctggatggac atttggtgct ggcgctgctc ttcaaatacc 120
ttttgctatg caaatggcat ataggttcaa tggcattgga gttacccaaa atgttctcta 180
tgagaaccaa aaacaatcg ccaaccaatt taacaaggcg attagtcaaa ttcaagaatc 240
acttacaaca acatcaactg cattgggcaa gctgcaagac gttgttaacc agaatgctca 300
agcattaaac acacttggtta aacaacttag ctctaatttt ggtgcaattt caagtgtgct 360
aaatgatatc ctttcgcgac ttgataaagt cgaggcggag gtacaaaattg acaggttaat 420
tacaggcaga cttcaaagcc ttcaaacctc tgtaacacaa caactaatca gggctgctga 480
aatcagggct tctgctaatc ttgctgctac taaaatgtct gagtgtgttc ttggacaatc 540

```

-continued

---

aaaaagagtt gacttttgtg gaaagggcta ccaccttatg tccttcccac aagcagcccc	600
gcatggtggt gtcttctac atgtcacgta tgtgccatcc caggagagga acttcaccac	660
agcgccagca atttgtcatg aaggcaaac atacttcctt cgtgaagggtg tttttgtggt	720
taatggcact tcttggttta ttacacagag gaacttcttt tctccacaaa taattactac	780
agacaatata tttgtctcag gaaattgtga tgtcgttatt ggcatcatta acaacacagt	840
ttatgatcct ctgcaacctg agctcgactc attcaaagaa gagctggaca agtacttcaa	900
aaatcatata tcaccagatg ttgatcttgg cgacatttca ggcataacg cttctgtcgt	960
caacattcaa aaagaaattg accgcctcaa tgaggtcgct aaaaatttaa atgaatcact	1020
cattgacctt caagaattgg gaaaatatga gcaatatatt aaatggcctt ggtatgtttg	1080
gctcggcttc attgctggac taattgccat cgtcatggtt acaatcttgc tttgttgcat	1140
gactagtgtg tgcagttgcc tcaaggggtc atgctcttgt ggttcttgcg gcaagtttga	1200
tgaggatgac tctgagccag ttctcaaggg tgtcaaatta cattacacat aaacgaactt	1260
agcaggtggg cgcgccatcg atttaattaa ggaagtgcaa atccaagaac ccattacttt	1320
atgatgcaa ctactttggt tgctggcaca cacataacta tgactactgt ataccatata	1380
acagtgtcac agatacaatt gtcgttactg aaggtgacgg catttcaaca ccaaaactca	1440
aagaagacta ccaaattggt ggttattctg aggataggca ctcaggtggt aaagactatg	1500
tcgttgta tggctatttc accgaagttt actaccagct tgagtctaca caaattacta	1560
cagacactgg tattgaaat gctacattct tcacttttaa caagcttgggt aaagaccac	1620
cgaatgtgca aatacacaca atcgacggct cttcaggagt tgctaatacca gcaatggatc	1680
caatttatga tgagccgacg acgactacta gcgtgccttt gtaagcaca gaaagtgagt	1740
acgaacttat gtactcattc gtttcggaag aaacaggtac gttaatagtt aatagcgtac	1800
ttctttttct gctttctggt gtattcttgc tagtcacact agccatcctt actgcgcttc	1860
gattgtgtgc gtactgctgc aatattgta acgtgagttt agtaaaacca acggtttacg	1920
tctactcgcg tgttaaaaat ctgaactcct ctgaaggagt tcctgatcct ctggtctaaa	1980
cgaactaact attattatta ttctgtttgg aactttaaca ttgcttatca tggcagacaa	2040
cggtactatt accggtgag agcttaaaaca actcctggaa caatggaacc tagtaatagg	2100
tttcttattc ctgacctgga ttatgttact acaatttgc ttttctaac ggaacaggtt	2160
tttgacata ataaagcttg ttttcctctg gctcttggg ccagtaacac ttgcttgttt	2220
tgtgcttgcg gctgtctaca gaattaattg ggtgactggc gggattgcca ttgcaatggc	2280
ttgtattgta ggcttgatgt ggcttagcta cttcgttgcg tccttcaggc tgtttgctcg	2340
taccgctca atgtggtcat tcaaccaga aacaacatt cttctcaatg tgcctctccg	2400
ggggacaatt gtgaccagac cgctcatgga aagtgaactt gtcattggtg ctgtgatcat	2460
tcgtggtcac ttgcgaatgg ccggacaccc cctagggcgc tgtgacatta aggacctgcc	2520
aaaagagatc actgtggcta catcaogaac gctttcttat tacaatttag gagcgtcgca	2580
gcgtgtaggc actgattcag gttttgctgc atacaaccgc taccgtattg gaaactataa	2640
attaaatata gaccacgccg gtagcaacga caatattgct ttgctagtac agtaagtgac	2700
aacagatggt tcactctggt gacttccag ttacaatagc agagatattg attatcatta	2760
tgaggacttt caggattgct atttggaatc ttgacgttat aataagtca atagtgagac	2820

-continued

---

aattatttaa gcctctaact aagaagaatt attcggagtt agatgatgaa gaacctatgg	2880
agttagatta tccataaaac gaacatgaaa attattctct tcctgacatt gattgtatgt	2940
acatcttgcg agctatatca ctatcaggag tgtgttagag gtacgactgt actactaaaa	3000
gaaccttgcc catcaggaac atacgagggc aattcacat ttcacctct tgctgacaat	3060
aaatttgac taacttgac tagcacacac tttgctttg cttgtgctga cggtaactga	3120
catacctatc agctcgtgc aagatcagtt tcaccaaacc tttcatcag acaaggagg	3180
gttcaacaag agctctact gccacttttt ctcatgttg ctgctctagt attttaata	3240
cttgcttca cattaagag aaagacagaa tgaatgagct cactttaatt gacttctatt	3300
tgtgctttt agcctttctg ctattccttg ttttaataat gcttattata ttttggttt	3360
cactgaaat ccaggatcta gaagaacct gtaccaaagt ctaaacgaac atgaaacttc	3420
tcattgtttt gacttgatct tctctatgca gttgcatatg cactgtagta cagcgtgtg	3480
catctaata acctcatgtg cttgaagatc cttgtaagg acaacactag gggtaatact	3540
tatagcactg cttggctttg tgctctagga aaggttttac ctttcatag atggcacact	3600
atggttcaaa catgcacacc taatgttact atcaactgct aagatccagc tgggtgtg	3660
cttatagcta ggtgttgga cttcatgaa ggtcaccaa ctgctgcatt tagagacgta	3720
cttgtgttt taataaacg acaaaattaa aatgtctgat aatggacccc aatcaacca	3780
acgtagtgcc ccccgatta ctttgggtg acccacagat tcaactgaca ataaccagaa	3840
tggaggacgc aatggggcaa ggcaaaaaca gcgccgacc caaggtttac ccaataatac	3900
tgctcttg ttcacagctc tcaactcagca tggcaaggag gaacttagat tccctcgagg	3960
ccaggcgctt ccaatcaaca ccaatagtg tccagatgac caaattggct actaccgaag	4020
agctaccga cgagttcgtg gtggtgacgg caaatgaaa gagctcagcc ccagatggta	4080
cttctattac cttagaactg gcccagaagc ttcacttccc tacggccta acaagaagg	4140
catcgtatgg gttgcaactg agggagcctt gaatacacc aaagaccaca ttggcaccg	4200
caatcctaata caaatgctg ccaccgtgct acaacttct caaggaacaa cattgcaaa	4260
aggcttctac gcagaggaa gcagaggcg cagtcaagcc tcttctcgt cctcatcag	4320
tagtcgctg aattcaagaa attcaactcc tggcagcagt aggggaaatt ctctgctg	4380
aatggctagc ggagggtg aaactgccct cgcgctattg ctgctagaca gattgaacca	4440
gcttgagagc aaagtttctg gtaaaggcca acaacaaca ggccaaactg tcaactaaga	4500
atctgctgct gaggcattc aaaaagcctg ccaaaaacgt actgccaca aacagtaca	4560
cgtaactca gcatgtgga gacgtggtcc agaacaacc caaggaaatt tcggggacca	4620
agaactaat agacaagaa ctgattacaa acattggcgc caaattgac aatttgctcc	4680
aagtgcctc gattctttg gaatgtcac cattggcatg gaagtccac cttcggaac	4740
atggctgact tatcatggag cattaatg gtagacaaa gatccacaat tcaagacaa	4800
cgtaactg ctgaacaagc acattgacgc atacaaaaca tcccaccaa cagagcctaa	4860
aaaggacaaa aagaaaaaga ctgatgaagc tcagccttg ccgagagac aaaagaagca	4920
gccactgtg actcttctc ctgctgctga catggatgat ttctccagac aactcaaaa	4980
ttcatgagt ggagcttctg ctgattcaac tcaggcataa aactcatga tgaccacaca	5040
aggcagatgg gctatgtaaa cgttttcgca attccgttta cgatacatag tctactctg	5100

-continued

---

tcgagaatga attctcgtaa ctaaacagca caagtaggtt tagttaactt taatctcaca	5160
tagcaatcct taatcaatgt gtaacattag ggaggacttg aaagagccac cacattttca	5220
tcgaggccac gcggagtacg atcgagggta cagtgaataa tgctagggag agctgcctat	5280
atggaagagc cctaagtgtg aaaattaatt ttagtagtgc tatccccatg tgattttaat	5340
agcttcttag gagaatgaca aaaaaaaaa aaaaaaaaa agggcgaatt ctgcagatat	5400
ccatcacact ggcggccgct cgagcatgca tctagagggc ccaattcgcc	5450

&lt;210&gt; SEQ ID NO 5

&lt;211&gt; LENGTH: 5000

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: SARS F X1X2 mutant

&lt;400&gt; SEQUENCE: 5

gtgtgctgga attcgcctt cgagtgccg cggccgcatt gctgcctaca cggctgctct	60
agttagtgtt actgccactg ctggatggac atttgggtgt ggcgctgctc ttcaaatacc	120
ttttgctatg caaatggcat ataggttcaa tggcattgga gttaccctaaa atgttctcta	180
tgagaaccaa aaacaaatcg ccaaccaatt taacaaggcg attagtcaaa ttcaagaatc	240
acttacaaca acatcaactg cattgggcaa gctgcaagac gttgttaacc agaatgctca	300
agcattaaac acacttggtt aacaacttag ctctaatttt ggtgcaattt caagtgtgct	360
aaatgatatc ctttcgcgac ttgataaagt cgaggcggag gtacaaattg acaggttaat	420
tacaggcaga cttcaaagcc ttcaaaccta tgtaacacaa caactaatca gggctgctga	480
aatcagggct tctgctaate ttgctgtctac taaaatgtct gagtgtgttc ttggacaatc	540
aaaaagagtt gacttttgtg gaaagggcta ccaccttatg tccttccac aagcagcccc	600
gcatggtgtt gtcttcttac atgtcacgta tgtgccatcc caggagagga acttcaccac	660
agcgcagca atttgtcatg aaggcaaagc atacttccct cgtgaagggtg tttttgtgtt	720
taatggcact tcttggttta ttacacagag gaacttcttt tctccacaaa taattactac	780
agacaatata tttgtctcag gaaattgtga tgtcgttatt ggcattatta acaacacagt	840
ttatgatcct ctgcaacctg agctcgactc attcaaagaa gagctggaca agtacttcaa	900
aaatcatata tcaccagatg ttgatcttgg cgacatttca ggcattaacg cttctgtcgt	960
caacattcaa aaagaaattg accgcctcaa tgaggctcgt aaaaatttaa atgaatcact	1020
cattgacctt caagaattgg gaaaatatga gcaatatatt aaatggcctt ggtatgtttg	1080
gctcggcttc attgctggac taattgccat cgtcatggtt acaatcttgc tttgttgcat	1140
gactagtgtt tgcagttgcc tcaaggtgac atgctcttgt ggttcttctg gcaagtttga	1200
tgaggatgac tctgagccag ttctcaaggg tgtcaaatta cattacacat aaacgaactt	1260
atgtactcat tcgtttcgga agaaacaggt acgttaatag ttaatagcgt acttcttttt	1320
cttgctttcg tggatttctt gctagtcaca ctagccatcc ttactgcgct tcgattgtgt	1380
gcgtactgct gcaatattgt taacgtgagt ttagtaaaac caacggttta cgtctactcg	1440
cgtgttaaaa atctgaaactt ttctgaagga gttcctgac ttctggtcta aacgaactaa	1500
ctattattat tattctgttt ggaactttaa cattgcttat catggcagac aacggtacta	1560
ttaccgttga ggagcttaaa caactcctgg aacaatgaa cctagtaata ggtttcctat	1620



-continued

---

tcctagcctg gattatgta ctacaatttg cctattctaa tcggaacagg tttttgtaca	1680
taataaagct tgttttcctc tggctcctgt ggccagtaac acttgcttgt tttgtgcttg	1740
ctgctgtcta cagaattaat tgggtgactg gcgggattgc gattgcaatg gcttgtattg	1800
taggcttgat gtggcttagc tacttctgtg cttccttcag gctgtttgct cgtaccgcgt	1860
caatgtggtc attcaacca gaaacaaaca ttcttctcaa tgtgcctctc cgggggacaa	1920
ttgtgaccag accgctcatg gaaagtgaac ttgtcattgg tgctgtgac attcgtggtc	1980
acttgcaat ggccggacac cccctagggc gctgtgacat taaggacctg ccaaaagaga	2040
tcactgtggc tacatcacga acgctttctt attacaaatt aggagcgtcg cagcgtgtag	2100
gcactgattc aggttttgct gcatacaacc gctaccgtat tggaaactat aaattaaata	2160
cagaccacgc cggtagcaac gacaatattg ctttgctagt acagtaagtg acaacagatg	2220
ttcatccttg ttgacttcca ggttacaata gcagagatat tgattatcat tatgaggact	2280
ttcaggattg ctatttggaa tcttgacggt ataataagtt caatagttag acaattattt	2340
aagcctctaa ctaagaagaa ttattcggag ttagatgatg aagaacctat ggagtttagat	2400
tatccataaa acgaacatga aaattattct cttcctgaca ttgattgtat ttacatcttg	2460
cgagctatat cactatcagg agtgtgtag aggtacgact gtactactaa aagaacctg	2520
cccctcagga acatacaggg gcaattcacc atttcaccct cttgctgaca ataaatttgc	2580
actaacttgc actagcacac actttgcttt tgcttgtgct gacggtaact gacataccta	2640
tcagctgcgt gcaagatcag tttcaccaaa acttttcatc agacaagagg aggttcaaca	2700
agagctctac tcgccacttt ttctcattgt tgctgctcta gtatttttaa tactttgctt	2760
caccattaag agaaagacag aatgaatgag ctcactttaa ttgacttcta tttgtgcttt	2820
ttagcctttc tgctattcct tgttttaata atgcttatta tattttgggt ttcactcgaa	2880
atccaggatc tagaagaacc ttgtacaaa gtctaaacga acatgaaact tctcattggt	2940
ttgacttgta tttctctatg cagttgcata tgcactgtag tacagcgtg tgcatcta	3000
aaacctcatg tgcttgaaaga tccttgtaag gtacaacact aggggtaata cttatagcac	3060
tgcttggtt tgtgctctag gaaaggtttt accttttcat agatggcaca ctatggttca	3120
aacatgcaca cctaattgta ctatcaactg tcaagatcca gctgggtgtg cgcttatagc	3180
taggtgttg taccttcatg aaggtcacca aactgctgca tttagagacg tacttgttgt	3240
tttaataaaa cgaacaaatt aaaatgtctg ataatggacc ccaatcaaac caacgtagt	3300
ccccccgcat tacatttgggt ggaccacag attcaactga caataaccag aatggaggac	3360
gcaatggggc aaggccaaaa cagcgcggac cccaaggttt acccaataat actgcgtctt	3420
ggttcacagc tctcactcag catggcaagg aggaacttag attccctoga ggccagggcg	3480
ttccaatcaa caccaatagt ggtccagatg accaaattgg ctactaccga agagctacc	3540
gacgagttcg tgggtgtgac ggcaaaatga aagagctcag cccagatgg tacttctatt	3600
acctaggaac tggcccagaa gcttcacttc cctacggcgc taacaaagaa ggcacgtat	3660
gggttgcaac tgaggaggcc ttgaatacac ccaaagacca cattggcacc cgcaatccta	3720
ataacaatgc tgccaccgtg ctacaacttc ctcaaggaa aacattgcca aaaggcttct	3780
acgcagaggg aagcagaggg gccagtcaag cctcttctcg ctcctcatca cgtagtgcg	3840
gtaattcaag aaattcaact cctggcagca gtaggggaaa ttctcctgct cgaatggcta	3900

-continued

---

```

gcggagggtgg tgaaactgcc ctgcgctat tgctgctaga cagattgaac cagcttgaga 3960
gcaaagtttc tggtaaaggc caacaacaac aaggccaac tgtcactaag aaatctgctg 4020
ctgaggcatc taaaaagcct cgccaaaaac gtactgccac aaaacagtac aacgtcactc 4080
aagcatttgg gagacgtggt ccagaacaaa cccaaggaaa tttcggggac caagacctaa 4140
tcagacaagg aactgattac aaacattggc cgcaaattgc acaatttgct ccaagtgcct 4200
ctgcattcct tggaatgtca cgcattggca tggaaagcac accttcggga acatggctga 4260
cttatcatgg agccattaaa ttggatgaca aagatccaca attcaaagac aacgtcatac 4320
tgctgaacaa gcacattgac gcatacaaaa cattcccacc aacagagcct aaaaggaca 4380
aaaaaiaaaa gactgatgaa gctcagcctt tgccgcagag acaaaagaag cagcccactg 4440
tgactcttct tcctgcggct gacatggatg atttctccag acaacttcaa aattccatga 4500
gtggagcttc tgctgattca actcaggcat aaacactcat gatgaccaca caaggcagat 4560
gggctatgta aacgtttttc caattccggt tacgatacat agtctactct tgtgcagaat 4620
gaattctcgt aactaaacag cacaagtagg tttagttaac tttaatctca catagcaatc 4680
tttaatcaat gtgtaacatt agggaggact tgaaagagcc accacatctt catcgaggcc 4740
acgcggagta cgatcggagg tacagtgaat aatgctaggg agagctgcct atatggaaga 4800
gccctaattg ttaaaattaa ttttagtagt gctatcccca tgtgatttta atagcttctt 4860
aggagaatga caaaaaaaaa aaaaaaaaaa aaaggcgcaa ttctgcagat atccatcaca 4920
ctggcgccg ctcgagcatg catctagagg gcccaattcg ccctatagtg agtcgtatta 4980
caattcactg gccgtcgttt 5000

```

```

<210> SEQ ID NO 6
<211> LENGTH: 5600
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: SARS F X3 mutant

```

```

<400> SEQUENCE: 6
gaattcgccc ttcgagtgcc cgcggccgca ttgctgccta cacggctgct ctagttagtg 60
gtactgccac tgctggatgg acatttgggt ctggcgctgc tcttcaaata ccttttgcta 120
tgcaaatggc atataggttc aatggcattg gagttacca aaatgttctc tatgagaacc 180
aaaaacaaat cgccaaccaa tttacaagc cgattagtca aattcaagaa tcacttacia 240
caacatcaac tgcatgggac aagctgcaag acgttggtta ccagaatgct caagcattaa 300
acacacttgt taaacaactt agctctaatt ttggtgcaat ttcaagtgtg ctaaatgata 360
tcctttcgcg acttgataaa gtcgagccgg aggtacaaat tgacaggtta attacaggca 420
gacttcaaag ctttcaaacc tatgtaacac aacaactaat cagggtctgct gaaatcaggg 480
cttctgctaa tcttgctgct actaaaatgt ctgagtggtg tcttggacaa tcaaaaagag 540
ttgacttttg tggaaagggc taccacttta tgccttccc acaagcagcc ccgcatggtg 600
ttgtcttctc acatgtcacg tatgtgccat cccaggagag gaacttcacc acagcgccag 660
caatttgta tgaaggcaaa gcatacttcc ctcgtgaagg tgtttttgtg tttaatggca 720
cttcttggtt tattacacag aggaacttct tttctccaca aataattact acagacaata 780
cattgtctc aggaaattgt gatgtogtta ttggcatcat taacaacaca gtttatgatc 840

```

-continued

---

ctctgcaacc	tgagctcgac	tcattcaaag	aagagctgga	caagtacttc	aaaaatcata	900
catcaccaga	tgttgatcct	ggcgacattt	caggcattaa	cgcttctgtc	gtcaacattc	960
aaaaagaaat	tgaccgcctc	aatgaggtcg	ctaaaaat	aaatgaatca	ctcattgacc	1020
ttcaagaatt	gggaaaatat	gagcaatata	ttaaatggcc	ttggtatggt	tggtcggct	1080
tcattgctgg	actaattgcc	atcgatcagg	ttacaatcct	gctttgttgc	atgactagtt	1140
gttgagttg	cctcaagggt	gcatgctcct	gtggttcttg	ctgcaagttt	gatgaggatg	1200
actctgagcc	agttctcaag	ggtgtcaaat	tacattacac	ataaacgaac	ttatggattt	1260
gtttatgaga	ttttttactc	ttggatcaat	tactgcacag	ccagtaaaaa	ttgacaatgc	1320
ttctcctgca	agtactgttc	atgctacagc	aacgataccg	ctacaagcct	cactcccttt	1380
cggtaggctt	gtttatggcg	ttgcatcttc	tgctgttttt	cagagcgcta	ccaaaataat	1440
tgcgctcaat	aaaagatgac	agctagccct	ttataagggc	ttccagtcca	tttgcaattt	1500
actgctgcta	ttgtttacca	tctattcaca	tcttttgctt	gtcgctgcag	gtatggaggc	1560
gcaatttttg	tacctctatg	ccttgatata	ttttctacaa	tgcatcaacg	catgtagaat	1620
tattatgaga	tgttgctctt	gttggaaagt	caaatccaag	aaccatttac	tttatgatgc	1680
caactacttt	gtttgctggc	acacacataa	ctatgactac	tgtataccat	ataacagtgt	1740
cacagataca	attgtcgtta	ctgaaggtga	cggcatttca	acacaaaac	tcaaagaaga	1800
ctaccaaatt	ggtggttatt	ctgaggatag	gcactcaggt	gttaaagact	atgtcgttgt	1860
acatggctat	ttcacccaag	ttactacca	gcttgagtct	acacaaatta	ctacagacac	1920
tggtattgaa	aatgctacat	tcttcatcct	taacaagctt	gttaaagacc	caccgaatgt	1980
gcaaatacac	acaatcgacg	gctcttcagg	agttgcta	ccagcaatgg	atccaattta	2040
tgatgagccg	acgacgacta	ctagcgtgcc	tttgtaagca	caagaaagtg	agtacgaact	2100
tatgtactca	ttcgtttcgg	aagaaacagg	tacgttaata	gttaatagcg	tacttctttt	2160
tcttgctttc	gtggatattc	tgctagtcac	actagccatc	cttactgcgc	ttcgattgtg	2220
tgcgactgct	tgcaatattg	ttaacgtgag	tttagtaaaa	ccaacggttt	acgtctactc	2280
gcgtgttaaa	aatctgaact	cttctgaagg	agttcctgat	cttctggtct	aaacgaacta	2340
actattatta	ttattctggt	tggaacttta	acattgctta	tcatggcaga	caacggctact	2400
attaccgttg	aggagcttaa	acaactcctg	gaacaatgga	acctagtaat	aggtttctta	2460
ttcctagcct	ggattatggt	actacaat	gcctattcta	atcggaacag	gtttttgtac	2520
ataataaagc	ttgttttctc	ctggctcttg	tggccagtaa	cacttgcttg	ttttgtgctt	2580
gtgctgtct	acagaattaa	ttgggtgact	ggcgggattg	cgattgcaat	ggcttgtatt	2640
gtaggcttga	tgtggcttag	ctacttcggt	gottccttca	ggctgtttgc	tcgtacccgc	2700
tcaatgtggt	cattcaaccc	agaaacaaac	attcttctca	atgtgcctct	ccgggggaca	2760
attgtgacca	gaccgctcat	ggaaagtga	cttgtcattg	gtgctgtgat	cattcgtggt	2820
cacttgcgaa	tggccggaca	ccccctaggg	cgctgtgaca	ttaaggacct	gcaaaaagag	2880
atcactgtgg	ctacatcacg	aacgctttct	tattacaaat	taggagcgtc	gcagcgtgta	2940
ggcactgatt	cagggttttg	tgcatacaac	cgctaccgta	ttggaaacta	taaattaaat	3000
acagaccacg	ccggtagcaa	cgacaatatt	gotttgctag	tacagtaagt	gacaacagaa	3060
cgaacatgaa	aattattctc	ttcctgacat	tgattgtatt	tacatcttgc	gagctatata	3120

-continued

---

actatcagga	gtgtgttaga	ggtacgactg	tactactaaa	agaaccttgc	ccatcaggaa	3180
catacgaggg	caattcacca	tttaccctc	ttgctgacaa	taaatttgca	ctaacttgca	3240
ctagcacaca	ctttgctttt	gcttgtgctg	acggtactcg	acatacctat	cagctgcgtg	3300
caagatcagt	ttcaccaaaa	cttttcatca	gacaagagga	ggttcaacaa	gagctctact	3360
cgccaacttt	tctcattggt	gctgctctag	tatttttaat	actttgcttc	accattaaga	3420
gaaagacaga	atgaatgagc	tcaactttaa	tgacttctat	ttgtgctttt	tagcctttct	3480
gctattcctt	gttttaataa	tgcttattat	attttggttt	tcaactgaaa	tccaggatct	3540
agaagaacct	tgtaccaaag	tctaaacgaa	catgaaactt	ctcattgttt	tgacttgtat	3600
ttctctatgc	agttgcatat	gcaactgtagt	acagcgctgt	gcataccta	aacctcatgt	3660
gcttgaagat	ccttgaag	tacaacacta	gggtaatac	ttatagcact	gcttggcttt	3720
gtgctctag	aaaggtttta	ccttttcata	gatggcacac	tatggttcaa	acatgcacac	3780
ctaattgtac	tatcaactgt	caagatccag	ctggtggtgc	gcttatagct	aggtgttgg	3840
accttcatga	aggtcaccaa	actgctgcat	ttagagacgt	acttgttgtt	ttaaataaac	3900
gaacaaatta	aatgtctga	taatggacc	caatcaaacc	aacgtagtgc	ccccgcatt	3960
acatttgggt	gacccacaga	ttcaactgac	aataaccaga	atggaggacg	caatggggca	4020
aggccaaaac	agcggcagc	ccaaggttta	cccaataata	ctgcgtcttg	gttcacagct	4080
ctcactcagc	atggcaagga	ggaacttaga	ttccctcgag	gccagggcgt	tccaatcaac	4140
accaatagtg	gtccagatga	ccaaattggc	tactaccgaa	gagctaccg	acgagttcgt	4200
ggtggtgacg	gcaaatgaa	agagctcagc	cccagatggt	acttctatta	cctaggaact	4260
ggcccagaag	ctcacttcc	ctacggcgt	aacaagaag	gcatcgtatg	ggttgcaact	4320
gagggagcct	tgaatacacc	caaagaccac	attggcacc	gcaatcctaa	taacaatgct	4380
gccaccgtgc	tacaacttcc	tcaaggaaca	acattgcca	aaggcttcta	cgagagggga	4440
agcagagggc	gcagcaagc	ctctctcgc	tcctcatcac	gtagtcgcg	taattcaaga	4500
aattcaactc	ctggcagcag	taggggaaat	tctcctgctc	gaatggctag	cggaggtggt	4560
gaaactgcc	tcgctgctat	gctgctagac	agattgaacc	agcttgagag	caaagtttct	4620
ggtaaaggcc	aacaacaaca	aggccaaact	gtcactaaga	aactctgctc	tgaggcatct	4680
aaaaagcctc	gcaaaaaacg	tactgccaca	aaacagtaca	acgtcactca	agcatttggg	4740
agcgtggtc	cagaacaaac	ccaaggaaat	ttcggggacc	aagacctaat	cagacaagga	4800
actgattaca	aacattggcc	gcaaattgca	caatttgctc	caagtgcctc	tgacttcttt	4860
ggaatgtcac	gcattggcat	ggaagtca	ccttcgggaa	catggctgac	ttatcatgga	4920
gccattaaat	tgatgacaa	agatccocaa	ttcaaagaca	acgtcactca	gctgaacaag	4980
cacattgacg	catacaaaac	attcccacca	acagagccta	aaaaggacaa	aaagaaaaag	5040
actgatgaag	ctcagccttt	gccgcagaga	caaaagaagc	agcccactgt	gactcttctt	5100
cctcgggctg	acatggatga	ttctccaga	caacttcaaa	attccatgag	tgagacttct	5160
gctgattcaa	ctcaggcata	aacactcatg	atgaccacac	aaggcagatg	ggctatgtaa	5220
acgttttgc	aattccgttt	acgatacata	gtctactctt	gtgcagaatg	aattctcgta	5280
actaaacagc	acaagtaggt	ttagttaact	ttaatctcac	atagcaatct	ttaatcaatg	5340
tgtaacatta	gggaggactt	gaaagagcca	ccacattttc	atcgaggcca	cgcgaggtac	5400

-continued

---

gatcagaggt acagtgaata atgctagggg gagctgccta tatggaagag ccctaattgtg	5460
taaaattaat tttagtagtg ctatccocat gtgattttaa tagcttctta ggagaatgac	5520
aaaaaaaaa aaaaaaaaaa aagggcgaat tctgcagata cccatcacac tggcggccgc	5580
tcgagcatgc atctagaggg	5600

<210> SEQ ID NO 7  
 <211> LENGTH: 5450  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial  
 <220> FEATURE:  
 <223> OTHER INFORMATION: SARS F X4 mutant

<400> SEQUENCE: 7

agtggtctgg aattcgcocct tcgagtgcc cgggccgcat tgctgcctac acggctgctc	60
tagttagtgg tactgccact gctggatgga catttggtgc tggcgtgct cttcaaatac	120
cttttctat gcaaatggca tataggttca atggcattgg agttaccxaa aatgttctct	180
atgagaacca aaaacaaatc gccaaccaat ttaacaaggc gattagtcaa attcaagaat	240
cacttacaac aacatcaact gcattgggca agctgcaaga cgttggttaac cagaatgctc	300
aagcattaaa cacactgtgt aaacaactta gctctaattt tggtgcaatt tcaagtgtgc	360
taaatgatat ctttctcgcga cttgataaag tcgagggcggg ggtacaaatt gacaggttaa	420
ttacaggcag acttcaaagc cttcaaacct atgtaacaca acaactaatc agggctgctg	480
aaatcagggc ttctgctaatt cttgtgctca ctaaaatgct tgagtgtgtt cttggacaat	540
caaaaagagt tgacttttgt ggaaagggct accaccttat gtccttccca caagcagccc	600
cgcatggtgt tgtcttccta catgtcacgt atgtgccatc ccaggagagg aacttcacca	660
cagcgcagc aatttgtcat gaaggcaaag catacttccc tcgtgaaggt gtttttgtgt	720
ttaatggcac ttcttggttt attacacaga ggaacttctt ttctccaxaa ataattacta	780
cagacaatac atttgtctca ggaaattgtg atgtcggtat tggcatcatt aacaacacag	840
tttatgatcc tctgcaacct gagctcgact cattcaaaga agagctggac aagtacttca	900
aaaatcatac atcaccagat gttgatcttg gogacattc aggcattaac gcttctgtcg	960
tcaacattca aaaagaaatt gaccgcctca atgaggtcgc taaaaattta aatgaatcac	1020
tcattgacct tcaagaattg ggaaaatag agcaatatat taaatggcct tggatgttt	1080
ggctcggcct cattgctgga ctaattgcc aogtcatggt tacaactctg ctttgttgca	1140
tgactagtty ttgcagttgc ctcaaaggty catgctctty tggttcttygc tgcaagttyg	1200
atgaggatga ctctgagcca gttctcaagg gtgtcaaatt acattacaca taaacgaact	1260
tatggattty tttatgagat ttttactct tggatcaatt actgcacagc cagtaaaat	1320
tgacaatgct tctcctgcaa gtactgttca tgctacagca acgataccgc tacaagcctc	1380
actoccttcc ggatggctty ttattggcgt tgcatttctt gctgttttcc agagcgtac	1440
caaaataaatt gcgctcaata aaagatggca gctagccctt tataagggct tccagttcat	1500
ttgcaattta ctgctgctat ttgttaccat ctattcacat cttttgctty tcgctgcag	1560
tatggaggcg caatttttgy acctctatgc cttgatatat tttctacaat gcatcaacgc	1620
atgtagaatt attatgagat gttggcttyg ttggaagtgc aaatccaaga acccattact	1680
ttatgatgcc aactacttyt tttgtggyca cacacataac tatgactact gtataccata	1740

-continued

taacagtgtc acagatacaa ttgtcgttac tgaaggtgac ggcatttcaa caccaaaact	1800
caaagaagac taccaaattg gtggttattc tgaggatagg cactcaggtg ttaaagacta	1860
tgctcgttga catggtattt tcaccgaagt ttactaccag cttgagtcta cacaaattac	1920
tacagacact ggtattgaaa atgctacatt cttcatcttt aacaagcttg ttaaagacct	1980
accgaatgtg caaatacaca caatcgacgg ctcttcagga gttgctaata cagcaatgga	2040
tccaatttat gatgagccga cgacgactac tagcgtgcct ttgtaagcac aagaaagtga	2100
gtacgaactt atgtactcat tcgcttcgga agaaacaggt acgttaatag ttaatagcgt	2160
acttcttttt cttgcttttcg tggattcttt gctagtcaca ctgacctcc ttactgcgct	2220
tcgattgtgt gcgtactgct gcaatattgt taacgtgagt ttagtaaaac caacggttta	2280
cgctactcgc cgtgttaaaa atctgaactc ttctgaagga gttcctgatc ttctggtcta	2340
aacgaactaa ctattattat tattctgttt ggaactttaa cattgcttat catggcagac	2400
aacggtaacta ttaccgctga ggagcttaaa caactcctgg aacaatgga cctagtaata	2460
ggtttcctat tcctagcctg gattatgtta ctacaatttg cctattctaa tcggaacagg	2520
ttttgtaca taataaagct tgttttcctc tggctcctgt ggccagtaac acttgcttgt	2580
tttgctctg ctgctgtcta cagaattaat tgggtgactg gcgggattgc gattgcaatg	2640
gctgtattg taggcttgat gtggcttagc tacttctgtt cttcctcag gctgcttct	2700
cgtaaccgct caatgtggtc attcaaccga gaaacaaaca ttcttctcaa tgtgcctctc	2760
cgggggacaa ttgtgaccag accgctcatg gaaagtgaac ttgtcattgg tgctgtgatc	2820
attcgtggtc acttgccaat ggccggacac cccctagggc gctgtgacat taaggacctg	2880
ccaaaagaga tcaactgtgc tacatcacga acgctttctt attacaaatt aggagcgtcg	2940
cagcgtgtag gcaactgattc aggttttgct gcatacaacc gctaccgtat tggaaactat	3000
aaattaaata cagaccacgc cggtagcaac gacaatattg ctttgcctagt acagtaagtg	3060
acaacagatg ttcatcttg ttgacttcca ggttacaata gcagagatat tgattatcat	3120
tatgaggact ttcaggattg ctatttggaa tcttgacgtt ataataagtt caatagtgag	3180
acaattattt aagcctctaa ctaagaagaa ttattcggag ttatgatgat aagaacctat	3240
ggagttagat tatccataaa acgaacatgg caggtgatcg attaattaag agctcacttt	3300
aattgacttc tttttgtgct ttttagcctt tctgctattc cttgttttaa taatgcttat	3360
tatatttttg ttttcaactc aaatccagga tctagaagaa ccttgtaacca aagtctaaac	3420
gaacatgaaa cttctcattg ttttgacttg ttttctcta tgcagttgca tatgcaactg	3480
agtacagcgc tgtgcatcta ataaacctca tgtgcttgaa gatccttgta aggtacaaca	3540
ctaggggtaa tacttatagc actgcttggc tttgtgctct aggaagggtt taaccttttc	3600
atagatggca cactatggtt caaacatgca cacctaattg tactatcaac tgtcaagatc	3660
cagctgggtg tgcgcttata gctaggtggt ggtaccttca tgaaggtoac caaactgctg	3720
catttagaga cgtacttggt gttttaaata aacgaacaaa ttaaagtgc tgataatgga	3780
ccccaatcaa accaacgtag tgccccccgc attacatttg gtggaccac agattcaact	3840
gacaataacc agaattggag acgcaatggg gcaaggccaa aacagcgcgc accccaaggt	3900
ttaccaata atactgcgct ttggttcaca gctctcactc agcatggcaa ggaggaactt	3960
agattccctc gaggccaggc cgttccaatc aacaccaata gtggtccaga tgaccaaat	4020

-continued

---

```

ggctactacc gaagagctac ccgacgagtt cgtgggtggtg acggcaaaat gaaagagctc 4080
agccccagat ggtacttcta ttacctagga actggcccag aagcttcaact tccctacggc 4140
gctaacaaaag aaggcatcgt atgggttgca actgagggag ccttgaatac acccaaagac 4200
cacattggca cccgcaatcc taataacaat gctgccaccg tgctacaact tcctcaagga 4260
acaacattgc caaaaggctt ctacgcagag ggaagcagag gcggcagtca agcctcttct 4320
cgctcctcat cacgtagtcg cggtaattca agaaattcaa ctctggcag cagtagggga 4380
aattctcctg ctccaatggc tagcggagggt ggtgaaactg ccctcgcgct attgctgcta 4440
gacagattga accagcttga gagcaaagtt tctggtaag gccacaaca acaaggccaa 4500
actgtcacta agaaatctgc tgctgaggca tctaaaaagc ctgcacaaa acgtactgcc 4560
acaaaacagt acaacgtcac tcaagcattt gggagacgtg gtccagaaca aaccaagga 4620
aatttcgggg accaagacct aatcagacaa ggaactgatt acaaacattg gccgcaaatt 4680
gcacaatttg ctccaagtgc ctctgcattc tttggaatgt cacgcattgg catggaagtc 4740
acaccttcgg gaacatggct gacttatcat ggagccatta aattggatga caaagatcca 4800
caattcaaaag acaacgtcat actgctgaac aagcacattg acgcatacaa aacattcca 4860
ccaacagagc ctaaaaagga caaaaagaaa aagactgatg aagctcagcc ttgcccag 4920
agacaaaaga agcagcccac tgtgactctt cttcctgcgg ctgacatgga tgatttctcc 4980
agacaacttc aaaattccat gagtggagct tctgctgatt caactcaggc ataaacactc 5040
atgatgacca cacaaggcag atgggctatg taaacgtttt cgcaattccg tttacgatac 5100
atagtctact cttgtgcaga atgaattctc gtaactaac agcacaagta ggtttagtta 5160
actttaatct cacatagcaa tctttaatca atgtgtaaca ttagggagga cttgaaagag 5220
ccaccacatt ttcacgcagg ccacgcggag tacgatcgag ggtacagtga ataatgctag 5280
ggagagctgc ctatatggaa gagcccta atgtgtaaaatt aattttagta gtgctatccc 5340
catgtgattt taatagcttc ttaggagaat gacaaaaaaa aaaaaaaaaa aaaaaaaggg 5400
cgaattctgc agatatccat cacactggcg gccgctcgag catgcatcta 5450

```

```

<210> SEQ ID NO 8
<211> LENGTH: 6150
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: SARS F X4/gfp substitution
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (648)..(648)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (872)..(872)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 8

```

```

agtgcccgcg gccgcatcgc tgcctacacg gctgctctag ttagtggtac tgccactgct 60
ggatggacat ttggtgctgg cgctgctctt caaatacctt ttgctatgca aatggcatat 120
aggttcaatg gcattggagt tacccaaaat gttctctatg agaaccaaaa acaaatcgcc 180
aaccaattta acaaggcgat tagtcaaatt caagaatcac ttacaacaac atcaactgca 240
ttgggaagc tgcaagacgt tgtaaccag aatgctcaag cattaaacac acttgtaaa 300

```

-continued

---

caacttagct ctaattttgg tgcaatttca agtgtgctaa atgatatcct ttcgcgactt	360
gataaagtcg agggcgaggt acaaattgac aggttaatta caggcagact tcaaagcctt	420
caaacctatg taacacaaca actaatcagg gctgctgaaa tcagggcttc tgctaactctt	480
gctgctacta aaatgtctga gtgtgttctt ggacaatcaa aaagagttga cttttgtgga	540
aagggtacc accttatgtc cttcccacaa gcagccccgc atgggtgtgt cttcctacat	600
gtcacgtatg tgccatccca ggagaggaac ttcaccacag cgccagcnat ttgtcatgaa	660
ggcaaagcat acttccctcg tgaaggtgtt tttgtgttta atggcacttc ttggtttatt	720
acacagagga acttcttttc tccacaaata attactacag acaatacatt tgtctcagga	780
aattgtgatg tcgttattgg catcattaac aacacagttt atgatcctct gcaacctgag	840
ctcgactcat tcaaagaaga gctggacaag tncctcaaaa atcatacatc accagatggt	900
gatcttggcg acatttcagg cattaacgct tctgtcgtca acattcaaaa agaaattgac	960
cgctcaatg aggtcgctaa aaatttaaat gaatcactca ttgacctca agaattggga	1020
aaatatgagc aatatattaa atggccttgg tatgtttggc tcggcttcat tgctggacta	1080
attgccatcg tcatggttac aatcttgctt tgttgcatga ctagtgttg cagttgcctc	1140
aagggtgcat gctcttggg tctctgtgc aagtttgatg aggatgactc tgagccagtt	1200
ctcaagggtg tcaaattaca ttacacataa acgaacttat ggatttgttt atgagatttt	1260
ttactcttgg atcaattact gcacagccag taaaaattga caatgcttct cctgcaagta	1320
ctgttcatgc tacagcaacg ataccgctac aagcctcact ccctttcgga tggcttgta	1380
ttggcggtgc atttcttgcg gtttttcaga gcgctaccaa aataattgcg ctcaataaaa	1440
gatggcagct agccctttat aagggtctcc agttcatttg caatttactg ctgctatttg	1500
ttaccatcta ttcacatcct ttgcttgcg ctgcaggat ggaggcgcaa tttttgtacc	1560
tctatgcctt gatataatct ctacaatgca tcaacgcatg tagaattatt atgagatggt	1620
ggctttgttg gaagtgcaaa tccaagaacc cttacttta tgatgccaac tactttgttt	1680
gctggcacac acataactat gactactgta taccatataa cagtgtcaca gatacaattg	1740
tcgttactga aggtgacggc atttcaacac caaaactcaa agaagactac caaattggtg	1800
gttattctga ggataggcac tcagggtgta aagactatgt cgttgatcat ggctatttca	1860
ccgaagtta ctaccagctt gagtctacac aaattactac agacactggg attgaaaatg	1920
ctacattcct catctttaac aagcttgta aagaccacc gaatgtgcaa atacacacaa	1980
tcgacggctc ttcaggagtt gctaataccag caatggatcc aatttatgat gagccgacga	2040
cgactactag cgtgccttgg taagcacaag aaagtgagta cgaacttatg tactcattcg	2100
tttoggaga aacaggtagc ttaatagtta atagcgtact tctttttctt gctttcgtgg	2160
tattcttgcg agtcacacta gccatcctta ctgcgcttgc attgtgtgcg tactgctgca	2220
atattgttaa cgtgagtta gtaaaaccaa cggtttacgt ctactcgcgt gttaaaaatc	2280
tgaactcttc tgaaggagtt cctgatcttc tggctaaac gaactaacta ttattattat	2340
tctgtttgga actttaacat tgcttatcat ggcagacaac ggtactatta ccgttgagga	2400
gcttaacaa ctcttggaac aatggaacct agtaatagg ttcctattcc tagcctggat	2460
tatgttacta caatttcctt attctaactg gaacaggttt ttgtacataa taaagcttgt	2520
ttcctctg cgcttggcg cagtaacact tgcttgtttt gtgcttgcg ctgtctacag	2580



-continued

---

aattaattgg gtgactggcg ggattgcatg tgcaatggct tgtattgtag gcttgatgtg	2640
gcttagctac ttcgttgctt ccttcaggct gtttgctcgt acccgctcaa tgtggtcatt	2700
caaccagaa acaaaccatc ttctcaatgt gcctctccgg gggacaattg tgaccagacc	2760
gctcatggaa agtgaacttg tcattggtgc tgtgatcatt cgtggtcact tgcgaatggc	2820
cggacacccc ctaggggcgt gtgacattaa ggacctgcca aaagagatca ctgtggctac	2880
atcacgaacg ctttcttatt acaaattagg agcgtcgcag cgtgtaggca ctgattcagg	2940
ttttgctgca tacaaccgct accgtattgg aaactataaa ttaatacag accacgccgg	3000
tagcaacgac aatattgctt tgctagtaca gtaagtgaca acagatgttt catcttgttg	3060
acttccagggt tacaatagca gagatattga ttatcattat gaggactttc aggattgcta	3120
tttggaatct tgacgttata ataagttcaa tagtgagaca attatttaag cctctaacta	3180
agaagaatta ttcggagtta gatgatgaag aacctatgga gttagattat ccataaacg	3240
aacaaattaa aatggtgagc aagggcgagg agctgttcac cggggtggtg cccatcctgg	3300
tcgagctgga cggcgacgta aacggccaca agttcagcgt gtcggcgag ggcgagggcg	3360
atgccaccta cggcaagctg accctgaagt tcactctgac caccggcaag ctgccctgtc	3420
cctggccacc cctcgtgacc accctgacct acggcgtgca gtgcttcagc cgctaccccg	3480
accacatgaa gcagcagcagc ttcttcaagt cggccatgcc cgaaggctac gtcaggagc	3540
gcaccatctt cttcaaggac gacggcaact acaagaccgg cgccgaggtg aagttcgagg	3600
gcgacaccct ggtgaaccgc atcgagctga agggcatcga cttcaaggag gacggcaaca	3660
tcctggggca caagctggag tacaactaca acagccacaa cgtctatatac atggccgaca	3720
agcagaagaa cggcatcaag gtgaacttca agatccgcca caacatcgag gacggcagcg	3780
tgagctcgc cgaccactac cagcagaaca ccccatcgg cgacggcccc gtgctgctgc	3840
ccgacaacca ctacctgagc acccagtcgg ccctgagcaa agacccaac gagaagcgcg	3900
atcacatggt cctgctggag ttcgtgaccg ccgcccggat cactctcggc atggacgagc	3960
tgtacaagta attaatgaag agctcacttt aattgacttc tatttgtgct ttttagcctt	4020
ctctgctattc cttgttttaa taatgcttat tataatttgg ttttactcagc aaatccagga	4080
tctagaagaa cttgtacca aagtctaacc gaacatgaaa cttctcattg ttttgacttg	4140
tatttctcta tgcagttgca tatgactgt agtacagcgc tgtgcatcta ataaacctca	4200
tgtgcttgaa gatccttgta aggtacaaca ctaggggtaa tacttatagc actgcttggc	4260
tttgctctct aggaaaggtt ttaccttttc atagatggca cactatggtt caaacatgca	4320
cacctaattgt tactatcaac tgtcaagatc cagctggtgg tgcgcttata gctagggtgt	4380
ggtaccttca tgaaggtcac caaactgctg catttagaga cgtacttgtt gttttaata	4440
aacgaacaaa ttaaaatgtc tgataatgga cccaatcaa accaacgtag tgccccccgc	4500
attacatttg gtggaccac agattcaact gacaataacc agaatggagg acgcaatggg	4560
gcaaggccaa aacagcgcgg accccaaggt ttaccaata atactgctc ttggttcaca	4620
gctctcactc agcatggcaa ggaggaactt agattccctc gaggccaggg cgttccaatc	4680
aacaccaata gtggtccaga tgaccaaatt ggctactacc gaagagctac ccgacgagtt	4740
cgtggtggtg acggcaaat gaaagagctc agccccagat ggtacttcta ttacctagga	4800
actggcccag aagcttccact tccctacggc gctaacaag aaggcatcgt atgggttgca	4860

-continued

---

```

actgagggag ccttgaatac acccaaagac cacattggca cccgcaatcc taataacaat 4920
gctgccaccg tgctacaact tcctcaagga acaacattgc caaaaggctt ctacgcagag 4980
ggaagcagag gcggcagtca agcctcttct cgctcctcat cacgtagtcg cggtaattca 5040
agaaattcaa ctcttgccag cagtagggga aattctcctg ctcgaatggc tagcggaggt 5100
ggtgaaactg ccctcgcgct attgtgcta gacagattga accagcttga gagcaaagt 5160
tctggtaaaag gccaacaaca acaaggccaa actgtcacta agaaatctgc tgctgaggca 5220
tctaaaaagc ctgcgcaaaa acgtactgcc acaaacagc acaacgtcac tcaagcattt 5280
gggagacgtg gtccagaaca aaccaagga aatttcgggg accaagacct aatcagacaa 5340
ggaactgatt acaaacattg gccgcaaatt gcacaatttg ctccaagtgc ctctgcattc 5400
tttggaatgt cacgcattgg catggaagtc acaccttcgg gaacatggct gacttatcat 5460
ggagccatta aattggatga caaagatcca caattcaaag acaacgtcat actgctgaac 5520
aagcacattg acgcatacaa aacattccca ccaacagagc ctaaaaagga caaaaagaaa 5580
aagactgatg aagctcagcc ttgcccagag agacaaaaga agcagcccac tgtgactctt 5640
cttctgcgg ctgacatgga tgatttctcc agacaacttc aaaattccat gaggggagct 5700
tctgctgatt caactcagc ataaacactc atgatgacca cacaaggcag atgggctatg 5760
taaacgtttt cgcaattccg tttacgatac atagtctact cttgtgcaga atgaattctc 5820
gtaactaaac agcacaagta ggtttagta actttaatct cacatagcaa tctttaatca 5880
atgtgtaaca ttagggagga cttgaaagag ccaccacatt ttcacgagg ccacgaggag 5940
tacgatcgag ggtacagtga ataagtctag ggagagctgc ctatatggaa gagccctaat 6000
gtgtaaaatt aattttagta gtgctatccc catgtgattt taatagcttc ttaggagaat 6060
gacaaaaaaa aaaaaaaaaa aaaaaaggg cgaattctgc agatatccat cacactggcg 6120
gccgctcgag catgcatcta gagggcccaa 6150

```

```

<210> SEQ ID NO 9
<211> LENGTH: 6370
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: SARS F X4/luciferase substitution

```

```

<400> SEQUENCE: 9

```

```

tcgagtgcc gcggccgcat tgctgctac aoggctgctc tagttagtgg tactgccact 60
gctggatgga catttggtgc tggcgtgct cttcaaatac cttttgctat gcaaatggca 120
tataggttca atggcattgg agttacccaa aatgttctct atgagaacca aaaacaaatc 180
gccaaccaat ttaacaaggc gattagtcaa attcaagaat cacttacaac aacatcaact 240
gcattgggca agctgcaaga cgttgtaac cagaatgctc aagcattaaa cacacttggt 300
aaacaactta gctctaattt tgggtcaatt tcaagtgtgc taaatgatat cctttcggca 360
cttgataaag tcgagggcga ggtacaaatt gacaggttaa ttacaggcag acttcaaagc 420
cttcaaacct atgtaacaca acaactaatc agggctgctg aaatcagggc ttctgctaata 480
cttgctgcta ctaaaatgtc tgagtgtgtt cttggacaat caaaaagagt tgacttttgt 540
ggaagggtct accaccttat gtccttccca caagcagccc cgcaggtgtg tgtcttcta 600
catgtcacgt atgtgccatc ccaggagag aacttcacca cagcgcagc aatttgcatt 660

```

-continued

---

gaagcacaag catacttccc tcgtgaaggt gtttttgtgt ttaatggcac ttcttggttt	720
attacacaga ggaacttctt ttctccacaa ataattacta cagacaatac atttgtctca	780
ggaaattgtg atgtcgttat tggcatcatt aacaacacag tttatgatcc tctgcaacct	840
gagctcgact cattcaaaga agagctggac aagtacttca aaaatcatac atcaccagat	900
gttgatcttg gcgacatttc aggcattaac gcttctgtcg tcaacattca aaaagaaatt	960
gaccgcctca atgaggtcgc taaaaattta aatgaatcac tcattgacct tcaagaattg	1020
ggaaaatgat agcaatata taaatggcct tggatgttt ggctcggctt cattgctgga	1080
ctaattgcca tcgtcatggt tacaatcttg ctttgttgca tgactagtgt ttgcagttgc	1140
ctcaagggtg catgctcttg tggttcttgc tgcaagtttg atgaggatga ctctgagcca	1200
gttctcaagg gtgtcaaatt acattacaca taaacgaact tatggatttg tttatgagat	1260
tttttactct tggatcaatt actgcacagc cagtaaaaat tgacaatgct tctcctgcaa	1320
gtactgttca tgctacagca acgataccgc tacaagcctc actcccttcc ggatggcttg	1380
ttattggcgt tgcatttctt gctgttttcc agagcgtac caaataaatt gcgctcaata	1440
aaagatggca gctagccctt tataaggcct tccagttcat ttgcaattta ctgctgctat	1500
ttgttaccat ctattcacat cttttgcttg tcgctgcagg tatggaggcg caatttttgt	1560
acctctatgc cttgatata tttctacaat gcacacacgc atgtagaatt attatgagat	1620
gttggttttg ttggaagtgc aaatccaaga acccattact ttatgatgcc aactactttg	1680
tttgctggca cacacataac tatgactact gtataccata taacagtgtc acagatacaa	1740
ttgtcgttac tgaaggtgac ggcatttcaa caccaaaact caaagaagac taccaaattg	1800
gtggttattc tgaggatagg cactcaggtg ttaaagacta tgcggttga catggctatt	1860
tcaccgaagt ttactaccag cttgagtcta cacaaattac tacagacact ggtattgaaa	1920
atgtacatt cttcatcttt aacaagcttg ttaaagacc accgaatgtg caaatacaca	1980
caatcgacgg ctcttcagga gttgctaacc cagcaatgga tccaatttat gatgagccga	2040
cgacgactac tagcgtgcct ttgtaagcac aagaaagtga gtacgaactt atgtactcat	2100
tcgtttcggg aaaaacaggt acgttaatag ttaatagcgt acttcttttt cttgctttcg	2160
tggtattctt gctagtcaca ctagccatcc ttactgcgct tcgattgtgt gcgactgct	2220
gcaatattgt taacgtgagt ttagtaaaac caacggttta cgtctactcg cgtgttaaaa	2280
atctgaactc ttctgaagga gttcctgac ttctggtcta aacgaactaa ctattattat	2340
tattctgttt ggaacttta cattgcttat catggcagac aacggtaacta ttaccgttga	2400
ggagcttaa caactcctgg aacaatggaa cctagtaata ggtttcctat tcctagcctg	2460
gattatgtta ctacaatttg cctattctaa toggaacagg tttttgtaca taataaagct	2520
tgttttcctc tggctcttgt gccagtaac acttgcttgt tttgtgcttg ctgctgtcta	2580
cagaattaat tgggtgactg gcgggattgc gattgcaatg gcttgattg taggcttgat	2640
gtggcttagc tacttcgttg cttccttcag gctgtttgct cgtaccgct caatgtggtc	2700
attcaaccga gaaacaaaca ttctctcaa tgtgcctctc cgggggacaa ttgtgaccag	2760
accgctcatg gaaagtgaac ttgtcattgg tgctgtgac attcgtgggc acttgcgaaat	2820
ggcggacac cccttagggc gctgtgacat taaggacctg caaaagaga tcactgtggc	2880
tacatcacga acgctttctt attacaaatt aggagcgtcg cagcgtgtag gcaactgattc	2940

-continued

---

aggttttgct gcatacaacc gctaccgtat tggaaactat aaattaaata cagaccacgc	3000
cggtagcaac gacaatattg ctttgctagt acagtaagtg acaacagatg tttcatcttg	3060
ttgacttcca ggttacaata gcagagatat tgattatcat tatgaggact ttcaggattg	3120
ctatttggaa tcttgacgtt ataataagtt caatagttag acaattatth aagcctctaa	3180
ctaagaagaa ttattcggag ttagatgatg aagaacctat ggagttagat tatccataaa	3240
acgaacaaat taaaatggct tccaaggtgt acgaccccgga gcaacgcaa cgcgatgatca	3300
ctgggcctca gtggtgggct cgctgcaagc aaatgaacgt gctggactcc ttcatacaact	3360
actatgattc cgagaagcac gccgagaacg ccgtgatttt tctgcatggg aacgctgcct	3420
ccagctacct gtggaggcac gtcgtgcctc acatcgagcc cgtggctaga tgcatacatc	3480
ctgatctgat cggaatgggt aagtcgggca agagcgggaa tggctcatat cgcctcctgg	3540
atcactacaa gtacctcacc gcttggttcg agctgctgaa ccttccaaag aaaatcatct	3600
ttgtgggcca cgactggggg gcttgctctg cctttcacta ctctacgag caccaagaca	3660
agatcaaggc catcgtccat gctgagagtg tcgtggacgt gatcagatcc tgggacgagt	3720
ggcctgacat cgaggaggat atcgccttga tcaagagcga agagggcgag aaaatggtgc	3780
ttgagaataa cttcttcgtc gagaccatgc tccaagcaa gatcatcggg aaactggagc	3840
ctgaggagtt cgctgcctac ctggagccat tcaaggagaa gggcgagggt agacggccta	3900
ccctctcctg gcctcgcgag atccctctcg ttaagggagg caagcccagc gtcgtccaga	3960
ttgtccgcaa ctacaacgcc taccttcggg ccagcgacga tctgcctaag atgttcatcg	4020
agtcgaccc tgggttcttt tccaacgcta ttgtcgaggg agctaagaag ttccttaaca	4080
ccgagttcgt gaagggaag gccctccact tcagccagga ggacgctcca gatgaaatgg	4140
gtaagtacat caagagcttc gtggagcgcg tgctgaagaa cgagcagtaa ttaattaaga	4200
gctcacttta attgacttct atttgtgctt tttagccttt ctgctattcc ttgttttaat	4260
aatgcttatt atattttggt tttcactcga aatccaggat ctagaagaac cttgtaccaa	4320
agtctaaacg aacatgaaac ttctcattgt tttgacttgt atttctctat gcagttgcat	4380
atgcactgta gtacagcgct gtgcatctaa taaacctcat gtgcttgaag atccttghaa	4440
ggtacaacac taggggtaat acttatagca ctgcttggtt ttgtgctcta ggaaggttt	4500
taccttttca tagatggcac actatggttc aaacatgac acctaattgt actatcaact	4560
gtcaagatcc agctggtggt gcgcttatag ctagggttg gtaccttcat gaaggtcacc	4620
aaactgctgc atttagagac gtacttgttg ttttaataa acgaacaaat taaaatgtct	4680
gataatggac cccaatacaa ccaacgtagt gcccccgca ttacatttg tggaccaca	4740
gattcaactg acaataacca gaatggagga cgcaatggg caaggccaaa acagcgccga	4800
ccccaaggtt tacccaataa tactgctctt tggttcacag ctctcactca gcatggcaag	4860
gaggaactta gattccctcg aggccagggc gttccaatca acaccaatag tggccagat	4920
gaccaaattg gctactaccg aagagctacc cgacgagttc gtggtggtga cggcaaatg	4980
aaagagctca gcccagatg gtacttctat tacctaggaa ctggcccaga agcttcaact	5040
ccctacggcg ctaacaaaga aggcacgta tgggttgcaa ctgagggagc cttgaataca	5100
cccaaagacc acattggcac ccgcaatcct aataacaatg ctgccacgt gctacaactt	5160
cctcaaggaa caacattgcc aaaaggcttc tacgcagagg gaagcagagg cggcagtcaa	5220

-continued

---

```

gcctcttctc gctcctcctc acgtagtcgc ggtaattcaa gaaattcaac tcttggcagc 5280
agtaggggaa atttctctgc tcgaatggct agcggagggtg gtgaaactgc cctcgcgcta 5340
ttgtgcttag acagattgaa ccagcttgag agcaaagttt ctggtaaagg ccaacaacaa 5400
caaggccaaa ctgtcactaa gaaatctgct gctgaggcat ctaaaaagcc tcgccaaaaa 5460
cgtactgcca caaacagta caacgtcact caagcatttg ggagacgtgg tccagaacaa 5520
acccaaggaa atttcgggga ccaagaccta atcagacaag gaactgatta caaacattgg 5580
ccgcaaatgg cacaaattgc tccaagtgcc tctgcattct ttggaatgac acgcattggc 5640
atggaagtca caccttcggg aacatggctg acttatcatg gagccattaa attggatgac 5700
aaagatccac aattcaaga caacgtcata ctgctgaaca agcacattga cgcatacaaa 5760
acattcccac caacagagcc taaaaggac aaaaagaaaa agactgatga agctcagcct 5820
ttgccgcaga gacaaaagaa gcagcccact gtgactcttc ttcctgcggc tgacatggat 5880
gatttctcca gacaacttca aaattccatg agtggagctt ctgctgattc aactcaggca 5940
taaacactca tgatgaccac acaaggcaga tgggctatgt aaacgttttc gcaattccgt 6000
ttacgatata tagtctactc ttgtgcagaa tgaattctcg taactaaaca gcacaagtag 6060
gtttagttaa ctttaatctc acatagcaat ctttaatcaa tgtgtaacat tagggaggac 6120
ttgaaagagc caccacattt tcatcgaggc cacgcggagt acgatcgagg gtacagtga 6180
taatgctagg gagagctgcc tatatggaag agccctaag tgtaaaatta attttagtag 6240
tgctatcccc atgtgatattt aatagcttct taggagaatg acaaaaaaaaa aaaaaaaaaa 6300
aaaaaagggc gaattctgca gatatccatc aactggcggc ccgctcgagc atgcatctag 6360
agggccaat 6370

```

```

<210> SEQ ID NO 10
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: SARS coronavirus

```

```

<400> SEQUENCE: 10

```

```

gccataatgg c 11

```

```

<210> SEQ ID NO 11
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: SARS coronavirus

```

```

<400> SEQUENCE: 11

```

```

gccagcgtgg c 11

```

```

<210> SEQ ID NO 12
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: SARS coronavirus

```

```

<400> SEQUENCE: 12

```

```

gcccaagagg c 11

```

```

<210> SEQ ID NO 13
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: SARS coronavirus

```

---

-continued

---

<400> SEQUENCE: 13  
gccctcctgg c 11

<210> SEQ ID NO 14  
<211> LENGTH: 11  
<212> TYPE: DNA  
<213> ORGANISM: SARS coronavirus

<400> SEQUENCE: 14  
gcctacacgg c 11

<210> SEQ ID NO 15  
<211> LENGTH: 15  
<212> TYPE: DNA  
<213> ORGANISM: SARS coronavirus

<400> SEQUENCE: 15  
ttggcgttg aggtg 15

<210> SEQ ID NO 16  
<211> LENGTH: 15  
<212> TYPE: DNA  
<213> ORGANISM: SARS coronavirus

<400> SEQUENCE: 16  
cacctgcggtt attg 15

<210> SEQ ID NO 17  
<211> LENGTH: 15  
<212> TYPE: DNA  
<213> ORGANISM: SARS coronavirus

<400> SEQUENCE: 17  
atgaaaatgc aggtg 15

<210> SEQ ID NO 18  
<211> LENGTH: 15  
<212> TYPE: DNA  
<213> ORGANISM: SARS coronavirus

<400> SEQUENCE: 18  
cacctgcgaa catga 15

<210> SEQ ID NO 19  
<211> LENGTH: 15  
<212> TYPE: DNA  
<213> ORGANISM: SARS coronavirus

<400> SEQUENCE: 19  
gacgagttgc aggtg 15

<210> SEQ ID NO 20  
<211> LENGTH: 15  
<212> TYPE: DNA  
<213> ORGANISM: SARS coronavirus

<400> SEQUENCE: 20  
cacctgcacc cgacg 15

<210> SEQ ID NO 21

---

-continued

---

<211> LENGTH: 15  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: AarI restriction site  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (8)..(15)  
<223> OTHER INFORMATION: n is a, c, g, or t  
  
<400> SEQUENCE: 21  
  
cacctgcnnn nnnnn 15  
  
<210> SEQ ID NO 22  
<211> LENGTH: 45  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 22  
  
tactaatagc actcactata gatattaggt ttttacctac ccagg 45  
  
<210> SEQ ID NO 23  
<211> LENGTH: 20  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 23  
  
acaccatagt caacgatgcc 20  
  
<210> SEQ ID NO 24  
<211> LENGTH: 23  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 24  
  
gcctatatgc atggatgta gat 23  
  
<210> SEQ ID NO 25  
<211> LENGTH: 23  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 25  
  
tgaaccgcca cgctggctaa acc 23  
  
<210> SEQ ID NO 26  
<211> LENGTH: 21  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 26  
  
agccagcgtg gcggttcata c 21  
  
<210> SEQ ID NO 27

---

-continued

---

<211> LENGTH: 24  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 27  
  
aggcctcttg ggcagtggca taag 24

<210> SEQ ID NO 28  
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 28  
  
actgccaag atgcctatga gc 22

<210> SEQ ID NO 29  
<211> LENGTH: 26  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 29  
  
cagccaggag ggcagacttc acaacc 26

<210> SEQ ID NO 30  
<211> LENGTH: 29  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 30  
  
gtctgccctc ctggctgata agttccag 29

<210> SEQ ID NO 31  
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 31  
  
gagcagccgt gtaggcagca at 22

<210> SEQ ID NO 32  
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 32  
  
attgctgcct acacggctgc tc 22

<210> SEQ ID NO 33  
<211> LENGTH: 39  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:



-continued

---

<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 33

tttttttttt tttttttttt tgtcattctc ctaagaagc 39

<210> SEQ ID NO 34  
<211> LENGTH: 56  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(2)  
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 34

nnggcctcga tggccattta ggtgacacta tagatgtctg ataatggacc ccaatc 56

<210> SEQ ID NO 35  
<211> LENGTH: 50  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(3)  
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 35

nnnttttttt tttttttttt ttttttttta tgcctgagtt gaatcagcag 50

<210> SEQ ID NO 36  
<211> LENGTH: 39  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 36

tttttttttt tttttttttt tgtcattctc ctaagaagc 39

<210> SEQ ID NO 37  
<211> LENGTH: 20  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 37

aaagccaacc aacctcgatc 20

<210> SEQ ID NO 38  
<211> LENGTH: 11  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: BglI recognition sequence  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (4)..(8)  
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 38

---

-continued

---

gccnnnnngg c 11

<210> SEQ ID NO 39  
<211> LENGTH: 15  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: AarI recognition sequence  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (8)..(15)  
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 39

cacctgcnnn nnnnn 15

<210> SEQ ID NO 40  
<211> LENGTH: 13  
<212> TYPE: DNA  
<213> ORGANISM: SARS coronavirus

<400> SEQUENCE: 40

cagtgtcaca gat 13

<210> SEQ ID NO 41  
<211> LENGTH: 13  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Mutant ORF3b premature stop stop sequence

<400> SEQUENCE: 41

cagtgtgaca gat 13

<210> SEQ ID NO 42  
<211> LENGTH: 27  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 42

gctgtgacat taaggacctg ccaaaag 27

<210> SEQ ID NO 43  
<211> LENGTH: 42  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 43

agggtcacct gcagccattt taatttatcc ggtttatgga ta 42

<210> SEQ ID NO 44  
<211> LENGTH: 42  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 44

agggtcacct gcagccattt taatttatcc gttttatgga ta 42

---

-continued

---

<210> SEQ ID NO 45  
<211> LENGTH: 27  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 45  
  
ggtgacacctg caaataaatg gcttcca 27

<210> SEQ ID NO 46  
<211> LENGTH: 30  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 46  
  
taaagtgagc tcttaattaa ttactgctcg 30

<210> SEQ ID NO 47  
<211> LENGTH: 17  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 47  
  
caggaaacag ctatgac 17

<210> SEQ ID NO 48  
<211> LENGTH: 31  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 48  
  
aaaatccggt tagagaacag atctacaaga g 31

<210> SEQ ID NO 49  
<211> LENGTH: 31  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 49  
  
ctaaccggat tttaaatct gtgtagctgt c 31

<210> SEQ ID NO 50  
<211> LENGTH: 21  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 50  
  
ataggctgt tcaagctggg g 21

<210> SEQ ID NO 51  
<211> LENGTH: 20  
<212> TYPE: DNA

---

-continued

---

<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 51  
tgctggctct gataaaggag 20

<210> SEQ ID NO 52  
<211> LENGTH: 43  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(3)  
<223> OTHER INFORMATION: n is a, c, g, or t  
  
<400> SEQUENCE: 52  
nnncacctgc acatatccgg ttagttgta acaagaatat cac 43

<210> SEQ ID NO 53  
<211> LENGTH: 48  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(3)  
<223> OTHER INFORMATION: n is a, c, g, or t  
  
<400> SEQUENCE: 53  
nnncacctgc aaccgatat gtttatttc ttattatttc ttactctc 48

<210> SEQ ID NO 54  
<211> LENGTH: 21  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 54  
catcaagcga aaaggcatca g 21

<210> SEQ ID NO 55  
<211> LENGTH: 20  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 55  
tgatcctctg caacctgagc 20

<210> SEQ ID NO 56  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(3)  
<223> OTHER INFORMATION: n is a, c, g, or t

---

-continued

---

<400> SEQUENCE: 56

nnncacctgc ataaatccgg actcactttc ttgtgcttac 40

<210> SEQ ID NO 57  
<211> LENGTH: 39  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(3)  
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 57

nnncacctgc gtccggattt atgtactcat tcgtttcgg 39

<210> SEQ ID NO 58  
<211> LENGTH: 45  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(3)  
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 58

nnncacctgc aatagttaat ccggttagac cagaagatca ggaac 45

<210> SEQ ID NO 59  
<211> LENGTH: 39  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(3)  
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 59

nnncacctgc ggattaacta ttattattat tctgtttgg 39

<210> SEQ ID NO 60  
<211> LENGTH: 21  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 60

taccaacacc tagctataag c 21

<210> SEQ ID NO 61  
<211> LENGTH: 63  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 61

gctgcattta gagacgtact tgttgtttta aataaccgga taaattaataa tgtctgataa 60

---

-continued

---

tgg 63

<210> SEQ ID NO 62  
<211> LENGTH: 31  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 62

ttaattaatt atgcctgagt tgaatcagca g 31

<210> SEQ ID NO 63  
<211> LENGTH: 29  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 63

cgtctcatgt gtaatgtaat ttgacacc 29

<210> SEQ ID NO 64  
<211> LENGTH: 48  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 64

cgtctcacac ataaccggat ttatggattt gtttatgaga ttttttac 48

<210> SEQ ID NO 65  
<211> LENGTH: 20  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 65

gtgcttgctg ttgtctacag 20

<210> SEQ ID NO 66  
<211> LENGTH: 33  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 66

cgtctccgtc cgggatgtag ccacagtgat ctc 33

<210> SEQ ID NO 67  
<211> LENGTH: 34  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer

<400> SEQUENCE: 67

cgtctccgga cgctttctta ttacaaatta ggag 34

<210> SEQ ID NO 68

---

-continued

---

<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 68  
  
cgtctctcat atccggttta tggataatct aactccatag 40  
  
<210> SEQ ID NO 69  
<211> LENGTH: 31  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 69  
  
cgtctcatat gaaaattatt ctcttctga c 31  
  
<210> SEQ ID NO 70  
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 70  
  
ggactttcag gattgctatt tg 22  
  
<210> SEQ ID NO 71  
<211> LENGTH: 33  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 71  
  
cgtctcatcc ggtagactt tggacaagg ttc 33  
  
<210> SEQ ID NO 72  
<211> LENGTH: 35  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
  
<400> SEQUENCE: 72  
  
cgtctcccgg atatgaaact tctcattggt ttgac 35  
  
<210> SEQ ID NO 73  
<211> LENGTH: 39  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic oligonucleotide primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: (1)..(3)  
<223> OTHER INFORMATION: n is a, c, g, or t  
  
<400> SEQUENCE: 73  
  
nnnttaatta attaatttgt tcgtttattt aaaacaaca 39  
  
<210> SEQ ID NO 74

-continued

---

```

<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(2)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 74
nnggcctcga tggccattta ggtgacacta tagatgtctg ataatggacc ccaatc          56

<210> SEQ ID NO 75
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(3)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 75
nnnttttttt tttttttttt ttttttttta tgcctgagtt gaatcagcag          50

<210> SEQ ID NO 76
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide probe
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (19)..(20)
<223> OTHER INFORMATION: biotinylation
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (24)..(24)
<223> OTHER INFORMATION: biotinylation
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (26)..(26)
<223> OTHER INFORMATION: biotinylation
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (28)..(28)
<223> OTHER INFORMATION: biotinylation

<400> SEQUENCE: 76
cttgactgcc gcctctgctt ccctctgc          28

```

---

What is claimed is:

1. An isolated nucleic acid comprising a nucleotide sequence encoding a nidovirus genome or replicon RNA, wherein the genome or replicon RNA comprises one or more of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural genes and further comprising a wild type CS in a TRS for ORFs 3a/b, ORF4, ORF5, ORF6, ORF7a/b, ORF8a/b and ORF9a/b.

2. An isolated nucleic acid comprising a nucleotide sequence encoding a nidovirus genome or replicon RNA, wherein the genome or replicon RNA comprises one or more

of the same mutations in a consensus sequence (CS) present in a transcription regulatory sequence (TRS) of a leader sequence and in the TRS located upstream of each of the structural genes and further comprising one or more of the same mutations in the CS of the TRS located upstream of open reading frame (ORF) 3a/3b, ORF 4, ORF 5, ORF6, ORF 7a/7b, ORF 8a/b and ORF 9a/b.

3. The nucleic acid of claim 1, wherein the nidovirus is a severe acute respiratory syndrome (SARS) coronavirus having the CS of ACGAAC.

4. The nucleic acid of claim 2, wherein the nidovirus is a severe acute respiratory syndrome (SARS) coronavirus having the CS of ACGAAC.



5. The nucleic acid of claim 3, wherein the mutation is selected from the group consisting of ACGGAC, ACGGAT, ACGGAT, CCGGAC, CCGAAT, CCGGAT, CCGGCG, CGCAAC, CCCGAT, AGCGAT, CGCGAT, CCCGTT, CGCGTT and TGCGGT.

6. The nucleic acid of claim 4, wherein the mutation is selected from the group consisting of ACGGAC, ACGGAT, ACGGAT, CCGGAC, CCGAAT, CCGGAT, CCGGCG, CGCAAC, CCCGAT, AGCGAT, CGCGAT, CCCGTT, CGCGTT and TGCGGT.

7. The nucleic acid of claim 1, wherein the Nidovirus is a group I coronavirus having the CS of CUAAAC and wherein the mutation is selected from the group consisting of GUA AAC, GCAAAC, CGAAAG, GCTAAAG, GCTTAG and GCTTGG.

8. The nucleic acid of claim 1, wherein the Nidovirus is a group II coronavirus having the CS of TCTAAAC and wherein the mutation is selected from CCTAAC, CGAAAC, CGTAAAG, CCGAAGG, CGTCCGC, CGGATTG and GGCCTG.

9. The nucleic acid of claim 1, wherein the Nidovirus is a group III coronavirus having the CS of CUUAACAA and wherein the mutation is selected from the group consisting of CUUAAGAA, GUUAAGAA, GUUGAGAA, GUUTCAG, CAAGGCAA, TCCAAGAT, GUUCCTC, GCCTAGCG and GCCTGGCT.

10. The nucleic acid of claim 1, wherein the Nidovirus is a torovirus having a CS of UUUAGA and wherein the mutation is selected from the group consisting of GUUAGA, GUUGGA, GUUGCA, GCUCCA, GCCACT and GCCTCT.

11. The nucleic acid of claim 1, wherein the Nidovirus is an arterivirus having a CS of UUAACC and wherein the mutation is selected from the group consisting of CUAACC, CCAACC, CCAAGC, CCAGGC, CCAGGT and GGT-TAG.

12. A nidovirus particle comprising the nucleic acid of claim 1.

13. A nidovirus particle comprising the nucleic acid of claim 2.

14. A composition comprising a population of the nidovirus particles of claim 12 and a pharmaceutically acceptable carrier.

15. A composition comprising a population of the nidovirus particles of claim 13 and a pharmaceutically acceptable carrier.

16. A method of eliciting an immune response in a subject, comprising delivering to the subject an effective amount of the nucleic acid of claim 1.

17. A method of eliciting an immune response in a subject, comprising delivering to the subject an effective amount of the nucleic acid of claim 2.

18. A method of eliciting an immune response in a subject, comprising administering to the subject an effective amount of the Nidovirus particle of claim 12.

19. A method of eliciting an immune response in a subject, comprising administering to the subject an effective amount of the Nidovirus particle of claim 13.

\* \* \* \* \*