



(86) Date de dépôt PCT/PCT Filing Date: 2004/07/09  
(87) Date publication PCT/PCT Publication Date: 2005/01/20  
(85) Entrée phase nationale/National Entry: 2006/01/10  
(86) N° demande PCT/PCT Application No.: IB 2004/002237  
(87) N° publication PCT/PCT Publication No.: 2005/005596  
(30) Priorités/Priorities: 2003/07/10 (CN03141517.2);  
2003/07/10 (CN03129330.1);  
2004/01/29 (CN2004100160632);  
2004/01/29 (CN2004100160628);  
2004/05/08 (CN2004100435974);  
2004/05/08 (CN200410043596X)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01),  
*A61K 39/215* (2006.01), *A61K 39/12* (2006.01)  
(71) Demandeurs/Applicants:  
CHINESE NATIONAL HUMAN GENOME CENTER AT  
SHANGHAI, CN;  
GUANGDONG CENTER FOR DISEASE CONTROL  
AND PREVENTION, CN;  
...  
(72) Inventeurs/Inventors:  
ZHAO, GUOPING, CN;  
XU, RUI, HENG, CN;  
WU, XINYAN, CN;  
TU, CHENG, CN;  
...

(54) Titre : CARACTERISATION DES STADES LES PLUS PRECOCES DU VIRUS DU SYNDROME RESPIRATOIRE  
AIGU SEVERE (SRAS), ET APPLICATIONS  
(54) Title: CHARACTERIZATION OF THE EARLIEST STAGES OF THE SEVERE ACUTE RESPIRATORY SYNDROME  
(SARS) VIRUS AND USES THEREOF

(57) **Abrégé/Abstract:**

Severe Acute Respiratory Syndrome ("SARS") is a human respiratory disease of recent origin, widespread infectivity, recurring incidence, and significant mortality. Although there is abundant evidence suggesting that the coronavirus responsible for the disease ("SARS-CoV") evolves during an outbreak, there is currently little data on the earliest strains of this coronavirus. The present invention is directed to the characterization of the genomic RNA sequences of these earliest SARS coronaviruses, to the identification of nucleotide positions within the SARS-CoV genomic RNA that are characteristic of the different evolutionary stages of the coronavirus, to kits based on these positions for use in diagnosis of the disease in patients, and for the development of vaccines to the disease based on the lowered virulence and contagiousness of these earliest strains of SARS-CoV.



- (71) **Demandeurs(suite)/Applicants(continued):** GUANGZHOU CENTER FOR DISEASE CONTROL AND PREVENTION, CN;  
CHANGCHUN UNIVERSITY OF AGRICULTURE AND ANIMAL SCIENCES, CN;  
RUIJIN HOSPITAL AFFILIATED TO SHANGHAI SECOND MEDICAL UNIVERSITY, CN;  
SHANGHAI INSTITUTES FOR BIOLOGICAL SCIENCES, CHINESE ACADEMY OF SCIENCES, CN;  
NANFANG HOSPITAL, FIRST MEDICAL UNIVERSITY OF PLA, CN;  
GUANGDONG J-TECH SCIENCE DEVELOPMENT CO. LTD., CN;  
SECOND AFFILIATED HOSPITAL OF SUN YET-SEN UNIVERSITY, CN
- (72) **Inventeurs(suite)/Inventors(continued):** SONG, HUAI-DONG, CN; LI, YIHONG, CN; HOU, JINLIN, CN; XU, JUN, CN;  
MIN, JUN, CN
- (74) **Agent:** BLAKE, CASSELS & GRAYDON LLP

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

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
20 January 2005 (20.01.2005)

PCT

(10) International Publication Number  
**WO 2005/005596 A3**

(51) International Patent Classification<sup>7</sup>: **C12Q 1/68**,  
A61K 39/12, 39/215

(21) International Application Number:  
PCT/IB2004/002237

(22) International Filing Date: 9 July 2004 (09.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

|               |                              |    |
|---------------|------------------------------|----|
| 03129330.1    | 10 July 2003 (10.07.2003)    | CN |
| 03141517.2    | 10 July 2003 (10.07.2003)    | CN |
| 2004100160628 | 29 January 2004 (29.01.2004) | CN |
| 2004100160632 | 29 January 2004 (29.01.2004) | CN |
| 200410043596X | 8 May 2004 (08.05.2004)      | CN |
| 2004100435974 | 8 May 2004 (08.05.2004)      | CN |

(71) Applicants (for all designated States except US): **CHINESE NATIONAL HUMAN GENOME CENTER AT SHANGHAI** [CN/CN]; 250 Bibo Road, Zhangjiang Hi-Tech Park, Shanghai 201203 (CN). **GUANGDONG CENTER FOR DISEASE CONTROL AND PREVENTION** [CN/CN]; 176 Xingang Road West, Guangzhou, Guangdong 510030 (CN). **GUANGZHOU CENTER FOR DISEASE CONTROL AND PREVENTION** [CN/CN]; 23 Third Zhongshan Road, Guangzhou, Guangdong 510080 (CN). **CHANGCHUN UNIVERSITY OF AGRICULTURE AND ANIMAL SCIENCES** [CN/CN]; 5333 Xian Road, Changchun, Jilin 130062 (CN). **RUIJIN HOSPITAL AFFILIATED TO SHANGHAI SECOND MEDICAL UNIVERSITY** [CN/CN]; 197 Ruijin Road II, Shanghai 200025 (CN). **SHANGHAI INSTITUTES FOR BIOLOGICAL SCIENCES, CHINESE ACADEMY OF SCIENCES** [CN/CN]; 320 Yueyang Road, Shanghai 200031 (CN). **NANFANG HOSPITAL, FIRST MEDICAL UNIVERSITY OF PLA** [CN/CN]; 628 Tonghe Road, Guangzhou, Guangdong 510515 (CN). **GUANGDONG J-TECH SCIENCE DEVELOPMENT CO. LTD.** [CN/CN]; 151 Yanjiang Road West, Guangzhou, Guangdong 510020 (CN). **SECOND AFFILIATED HOSPITAL OF SUN YET-SEN UNIVERSITY** [CN/CN]; 107 Yanjiang Road West, Guangzhou, Guangdong 510020 (CN).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ZHAO, Guoping** [CN/CN]; Lane 250 Guiping Road, Shanghai 200233 (CN). **XU, Rui, Heng** [CN/CN]; 176 Xin Gang Road West, Guangzhou, Guangdong 510300 (CN). **WU, Xinyan** [CN/CN]; Guangzhou, Guangdong (CN). **TU, Cheng** [CN/CN]; 5333 Xian Road, Changchun, Jilin 130062 (CN). **SONG, Huai-Dong** [CN/CN]; 269 Zhaojiaband Road, Shanghai 200031 (CN). **LI, YiHong** [CN/CN]; 1433 Xi Zhang Road South, Shanghai 200011 (CN). **HOU, Jinlin** [CN/CN]; 1838 Guangzhou Blvd, Guangzhou, Guangdong 510515 (CN). **XU, Jun** [CN/CN]; 701 Kangda Road, Guangzhou, Guangdong 510230 (CN). **MIN, Jun** [CN/CN]; 107 Yanjiang Road West, Guangzhou, Guangdong 510120 (CN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:  
12 May 2005

[Continued on next page]

(54) Title: CHARACTERIZATION OF THE EARLIEST STAGES OF THE SEVERE ACUTE RESPIRATORY SYNDROME (SARS) VIRUS AND USES THEREOF

(57) Abstract: Severe Acute Respiratory Syndrome ("SARS") is a human respiratory disease of recent origin, widespread infectivity, recurring incidence, and significant mortality. Although there is abundant evidence suggesting that the coronavirus responsible for the disease ("SARS-CoV") evolves during an outbreak, there is currently little data on the earliest strains of this coronavirus. The present invention is directed to the characterization of the genomic RNA sequences of these earliest SARS coronaviruses, to the identification of nucleotide positions within the SARS-CoV genomic RNA that are characteristic of the different evolutionary stages of the coronavirus, to kits based on these positions for use in diagnosis of the disease in patients, and for the development of vaccines to the disease based on the lowered virulence and contagiousness of these earliest strains of SARS-CoV.



WO 2005/005596 A3

**WO 2005/005596 A3**



---

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# **DEMANDES OU BREVETS VOLUMINEUX**

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS  
COMPREND PLUS D'UN TOME.**

**CECI EST LE TOME \_\_1\_\_ DE \_\_2\_\_**

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

---

# **JUMBO APPLICATIONS / PATENTS**

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE  
THAN ONE VOLUME.**

**THIS IS VOLUME \_\_1\_\_ OF \_\_2\_\_**

NOTE: For additional volumes please contact the Canadian Patent Office.

TITLE

CHACTERIZATION OF THE EARLIEST  
STAGES OF THE SEVERE ACUTE RESPIRATORY  
SYNDROME (SARS) VIRUS AND USES THEREOF

5

BACKGROUND OF THE INVENTION

[001] Severe Acute Respiratory Syndrome ("SARS") is a human respiratory disease of recent origin, widespread infectivity, recurring incidence, and significant mortality.

10 Specifically, SARS is a recently-observed human disease, with the first cases seen in Guangdong Province, China, in November, 2002. During this 2002-2003 outbreak, the World Health Organization ("WHO") reported more than 30 countries in which the disease had occurred, with 774 of the 8096 patients who had contracted SARS eventually dying of the disease (see the WHO website at [who.int/csr/sars/country/table2004\\_04\\_21/en/](http://who.int/csr/sars/country/table2004_04_21/en/)).

15 Moreover, a second outbreak of SARS in four patients in the city of Guangzhou, Guangdong Province, China, between December, 2003, and January, 2004, demonstrated that the disease is recurrent, and therefore continues to be of serious impact to worldwide human health (see the WHO website at [who.int/csr/don/2004\\_01\\_27/en/](http://who.int/csr/don/2004_01_27/en/) and [who.int/csr/don/2004\\_01\\_31/en/](http://who.int/csr/don/2004_01_31/en/)).

[002] Subsequent to the initial SARS outbreak, an intensive collaborative research

20 effort by the international scientific community identified the etiological agent causing the disease to be a novel coronavirus, the SARS coronavirus ("SARS-CoV" or, synonymously, "SCoV" or "SARS virus") (Ksiazek *et al.*, *N. Engl. J. Med.* 348:1947 (2003); Peiris *et al.*, *Lancet* 361:1319 (2003); Drosten *et al.*, *N. Engl. J. Med.* 348:1967 (2003)). This identification of the causative agent of SARS as a coronavirus is consistent with the known

25 role of these viruses in animal and human respiratory diseases; as many as one third of all human mild upper respiratory tract illnesses, for example, are caused by human coronaviruses. Interestingly, however, although SARS-CoV is clearly a member of this diverse group of positive-stranded RNA viruses, based on RNA sequence comparisons it appears that SARS-CoV does not fall within any of the coronavirus evolutionary groups

30 previously characterized, i.e., is not closely related to any previously known coronavirus (Rota *et al.*, *Science* 300:1394 (2003); Marra *et al.*, *Science* 300:1399 (2003)).

[003] Although the agent responsible for SARS has been identified, successful prevention and treatment of the disease requires an additional understanding of the origin of SARS-CoV in humans, as well as knowledge of how the virus mutates during an outbreak of SARS. With regard to origin, as discussed above, SARS has only recently been observed in humans, suggesting the prior existence of SARS-CoV, or a close relative of SARS-CoV, in a separate, non-human source with which humans have had recent contact. Thus an important component in the control and prevention of future SARS outbreaks will be an understanding of this origin, including: knowledge of how the SARS coronavirus crosses this species barrier, i.e., the characteristics of the virus at the point when it first infects humans; and, an understanding of the non-human source of the coronavirus.

[004] In this latter regard, a variety of data strongly implicate Himalayan palm civets (*Paguma larvata*; “palm civets” or, synonymously, “civets”) as the non-human source of SARS-CoV (although probably not the ultimate repository of the disease itself; see Example 2 below). First, the early cases of SARS in both the 2002-2003 and 2003-2004 outbreaks were associated with patient exposure to these exotic food animals, suggesting that they are the vectors for transmission to humans of SARS-CoV, or a close relative of SARS-CoV. And, second, it has been shown that palm civets indeed harbor a SARS-CoV-like coronavirus (synonymously “SCoV-like coronavirus”) highly related to SARS-CoV (99.8% RNA sequence homology), further suggesting the origin of the latter human form of the coronavirus from transmission of the former palm civet form (Guan *et al.*, *Science* 302:276 (2003)). Despite this knowledge of the likely non-human source of the SARS coronavirus, however, the exact form of the virus at or immediately after transmission has not yet been elucidated.

[005] With regard to the mutation of SARS-CoV during a SARS outbreak, a number of studies have demonstrated a variety of mutational changes in the SARS-CoV RNA sequences of various patients from the 2002-2003 outbreak (Ruan *et al.*, *Lancet* 361:1779 (2003); Lan-Dian *et al.*, *Acta Pharmacol. Sin.* 24:741 (2003)). Such changes are hardly surprising, in light of the recent introduction of SARS into humans from palm civets or some other non-human source discussed above, the strong selection pressures on the virus resulting after such a change in host, and the inherently high rate of genetic mutation in the coronaviruses resulting from their use of RNA as their genetic material. Such mutations in the underlying RNA genetic material are expected to result in new SARS coronaviral strains

better adapted for growth in the human host, for ability to evade the human immune system, or with other novel properties which impact human health, for example the human-human hyper-infectivity of particular strains of the SARS coronavirus in what are termed “superspreader” events. Therefore, understanding the changes that occur in the SARS coronavirus during the course of an outbreak is critical to controlling and ultimately preventing the disease.

[006] In light of the preceding discussion, it is clear that, although changes in the SARS coronavirus occurring at all stages of a SARS outbreak are important to an understanding of how to combat SARS, it is particularly important to understand the evolution of the SARS coronavirus in the earliest stages of its infection of humans, i.e., those stages at or immediately following the point at which the coronavirus crosses the species barrier. Such understanding of the earliest strains of the SARS coronavirus can be expected to lead to a variety of insights into prevention and treatment of the disease, including, for example: the development of molecular markers for identifying different evolutionary stages of the SARS-CoV (i.e., different stages occurring during a SARS outbreak), thereby allowing for the prediction of the severity of the disease in an infected patient, as well as likelihood of infectivity to others; the development of procedures based on the properties of the early SARS-CoV strains obtained, e.g., the use of the RNA genetic material of early SARS strains to obtain SARS coronaviral proteins important for the spread of these early strains (or their initial transmission to humans) for study and ultimately for targeting for drug inhibition; and, the use of these early SARS coronavirus strains in whole or in part in the development of vaccines to prevent SARS.

[007] Despite this need for an understanding of the earliest stages of the SARS virus in humans, to date the data on the evolution of the SARS coronavirus during these earliest outbreak stages are limited. For example, Ruan *et al.* (*Lancet* 361:1779 (2003)) compared the RNA nucleotide sequences of fourteen SARS-CoV sequences from the 2002-2003 outbreak, only one of which (GZ01, also referred to in the literature as GD01) dates to the early stages of this outbreak. Although these data allowed Ruan *et al.* to make a number of statements regarding nucleotide positions of the SARS-CoV RNA associated with different regional SARS outbreaks, in light of the paucity of data from the earliest stages of the 2002-2003 outbreak, few conclusions can be drawn from the data of Ruan *et al.* about the critical earliest stages of the evolution of SARS-CoV.

[008] There is thus a great need to obtain data regarding the earliest stages of the evolution of the SARS coronavirus in humans in order to understand, treat, and prevent SARS.

## 5 SUMMARY OF THE INVENTION

[009] The present invention satisfies the need to better understand the earliest stages of the evolution of the SARS coronavirus. Specifically, the present invention is directed to the characterization of different SARS-CoV strains occurring at different stages of a SARS outbreak, and particularly to previously uncharacterized earliest stage SARS-CoV strains,  
10 i.e., SARS-CoV strains obtained from the earliest infected patients in a SARS outbreak; and to the uses of these data for, e.g., disease diagnosis, drug development, and vaccine development.

[010] Thus one embodiment of the present invention is directed to the use of the data obtained in these sequence comparisons to identify insertions, deletions, and single  
15 nucleotide variations (SNV) within the SARS virus RNA that are characteristic for each stage of SARS virus, i.e., that serve as stage-specific molecular markers for the different strains characteristic of the earliest stages (including both “early-early” and “early” stages as defined elsewhere herein), middle, and late stages of an outbreak. The present invention also includes methods for using this information in determining the staging of SARS-CoV or  
20 SARS-CoV-like sequences of interest, as well as kits for conducting such analyses.

[011] A second embodiment of the present invention is directed to the SARS coronaviral nucleotide sequences characteristic of these stages, and particularly to the nucleotide sequences characteristic of earliest stage SARS-CoV sequences (i.e., “early-early” and “early” stages), as well as closely-related SARS-CoV-like sequences. These earliest  
25 stage SARS-CoV sequences and closely-related SARS-CoV-like sequences are novel, i.e., first-characterized in the present invention, and may be used as the basis for experiments designed to test the role of any of the proteins encoded by the coronaviral RNA sequence in species specificity and infectivity. Thus, for example, the availability of novel SARS-CoV sequences from the earliest-infected patients of the 2002-2003 SARS outbreak as well as  
30 from the 2003-2004 outbreak allow for the use of these sequences to produce proteins

characteristic of the virus in its earliest stages, i.e., its state shortly after crossing the species barrier from its non-human source into humans. These nucleotide sequences are supplemented by the availability in the present invention of novel SARS-CoV-like sequences, specifically the previously uncharacterized SARS-CoV-like sequences obtained from civet cats during the period of the 2003-2004 human SARS outbreak. As described elsewhere herein, these sequences are very similar to those of the earliest stage SARS-CoV sequences; therefore, the human-derived SARS-CoV sequences in combination with the civet-derived SARS-CoV-like sequences allow for the more complete study of the earliest stage coronaviral proteins and other viral properties.

[012] Finally, a third embodiment of the present invention is directed to the development of vaccines based on the properties of the novel earliest stage SARS-CoV sequences and closely-related SARS-CoV-like sequences of the present invention. Such vaccines take advantage of the unique properties observed for the earliest stage SARS-CoV, namely, the milder symptoms seen with these coronaviral infections, as well as the lack of human-human infectivity (see, e.g., Example 2 elsewhere herein).

### **BRIEF DESCRIPTION OF THE DRAWINGS**

[013] The foregoing summary, as well as the following detailed description of preferred embodiments of the present invention, will be better understood when read in conjunction with the appended drawings. For the purpose of illustrating the present invention, there are shown in the drawings embodiments which are presently preferred. It should be understood, however, that the present invention is not limited to the precise arrangements and instrumentalities shown.

[014] Figures 1A-J show the single nucleotide variations (SNVs) for 92 SARS-CoV and SARS-CoV-like coronaviral sequences. Specifically, Figures 1A-J show the alignment of 2 human SARS-CoV RNA sequences from the 2003-2004 outbreak (the "HP04" sequences), 3 SARS-CoV-like palm civet RNA sequences obtained in this same period (the "PC04" sequences), 2 SARS-CoV-like palm civet RNA sequences obtained during the period of the 2002-2003 outbreak (the "PC03" sequences), 14 early-stage SARS-CoV RNA sequences from the 2002-2003 outbreak (the "HP03-E" sequences), 15 middle-stage SARS-CoV RNA sequences from the 2002-2003 outbreak (the "HP03-M" sequences), and 56 late-

stage SARS-CoV RNA sequences from the 2002-2003 outbreak (the "HP03-L" sequences). Note that Figures 1A-E provide a list of all single nucleotide variations (SNVs) for the HP04, PC04, PC03, HP03-E, and HP03-M groups; the 56-member HP03-L group is described in the last 13 lines of Figures 1A-E and in Figures 1F-J. Early, middle, and late-stage

5 determinations were performed as described in Example 1. Sequence alignments were performed as described in Example 1, using the GZ02 SARS-CoV sequence (SEQ ID NO:1) as the reference sequence. Rows 1 and 2 of the figure show the protein-coding regions of the SARS-CoV RNA. Row 3 shows the SNVs for each position in the table, with the first of the two nucleotides show that of the GZ02 reference sequence. Thus, for example, position 508  
10 is either the T of the GZ02 sequence or other sequences as shown, or the G of the non-GZ02 sequences also indicated for this position in the figure. Row 4 shows the triplet codon containing this SNV, with the number at the end of the triplet codon indicating the exact position in the triplet which contains the SNV. Thus, for example, the "ggc1" shown in this row for position 508 indicates that the first position of the triplet corresponds to the 508 SNV,  
15 i.e., that the allowable variant codons corresponding to the 508 SNV are either TGC or GGC. Row 5 shows the single letter amino acid abbreviations corresponding to these two triplets, i.e., for the 508 SNV, C (corresponding to TGC) or G (corresponding to GGC). Row 6 shows the amino acid residue in the appropriate protein product of the coronaviral RNA corresponding to the triplet codon. Row 7 shows the nucleotide coordinate for the SNV, with  
20 the numbering indicated relative to the GZ02 reference sequence (SEQ ID NO:1). Finally, the vertically shaded columns at positions 17564, 21721, 22222, 23823, and 27827 are used to highlight these positions; see Example 1 and elsewhere in the text for a detailed discussion of the use of these positions in the present invention. Note that the "N" entries in the figure refer to nucleotides with undetermined identity.

25 [015] Figures 2A-H show a summary of the data of Figures 1A-J. Specifically, Figures 2A-H show the occurrence in some or all of each of the possible SARS-CoV or SARS-CoV-like groups of Figures 1A-J (i.e., HP04, PC04, PC03, HP03-E, HP03-M, and HP03-L) of each of the nucleotides shown in column 2 at the SNV position indicated in column 1. Thus, for example, Figure 2A shows that a T nucleotide at the 508 SNV occurs  
30 only in the HP03-E group, and only in some (but not all) of the members of this group. By contrast, Figure 2A shows that a T nucleotide at the 1909 SNV occurs in two groups, PC03 and HP03-E, with all members of the PC03 group having this value, but only some of the members of the HP03-E group having this value. Figures 2A-H also indicate SNV positions

which are characteristic for all members of a single group of Figures 1A-J, i.e., situations in which a particular nucleotide at a particular SNV position occurs in only one of the HP04, PC04, PC03, HP03-E, HP03-M, and HP03-L groups, and occurs in all members of that particular group. In these figures, "EE" indicates the "early-early" group referred to in the text, with inclusion in this group requiring that the particular nucleotide at a particular SNV position occur in all of the members of two groups, the HP04 and PC04 groups. In these figures, "E" indicates the "earliest stage" HP04/PC04/PC03/HP03-E group also referred to in the text, with inclusion in this group requiring that the particular nucleotide at a particular SNV position occur in all of the members of the four groups HP04, PC04, PC03, and HP03-E. Note that only a G nucleotide at the 23823 position is indicative of this four-member group.

[016] Figure 3 shows the predicted RNA secondary structure of the Orf7b-Orf8 region of the SARSCoV genome. SARS-CoV genotypic variations caused by major deletion events were observed on a number of occasions during the epidemic. All such deletions were confined to the Orf7b-Orf8 region. The genomic locations of the major deletions observed in this study are indicated on the predicted RNA secondary structures of the longest SARS-CoV genotype (left panel) and the genotype with the 29-nt deletion (right panel). The former genotype is represented by GZ02 (SEQ ID NO:1) while the latter is represented by TOR2 (accession number AY274119 available at the website [ncbi.nlm.nih.gov/entrez](http://ncbi.nlm.nih.gov/entrez)). This latter genotype predominated for the remainder of the epidemic from the middle phase onwards. For both panels, the illustrated region starts from 14 nucleotides upstream to the start of the predicted Orf7 to 14 nucleotides downstream to the end of Orf8. The illustrated region corresponds to nucleotide positions 27288 to 28161 on GZ02 (SEQ ID NO:1) and nucleotide positions 27259 to 28132 on TOR2. The prediction was made using the VIENNARNA:RNAfold software (available at the website [bioweb.pasteur.fr/](http://bioweb.pasteur.fr/)). GZ-B and GZ-C are two genotypes obtained from two Guangzhou patients with disease onset from mid-March but demonstrated a 39-nt deletion.

[017] Figure 4 shows the predicted coding sequence changes caused by the major deletion events in the Orf7b-Orf8 region of the SARS-CoV genome. The amino acid sequences of the Orf7b, Orf8 (8a and 8b) and N proteins as predicted for the major SARS-CoV deletion variants are listed in the figure. Corresponding nucleotide coordinates for each predicted open reading frame are based on the GZ02 SARS-CoV sequence (SEQ ID NO:1).

[018] Figure 5 shows the genotype clustering of SARS-CoV during the course of the epidemic. An unrooted phylogenetic tree of SARS-CoV was constructed from 61 human SARS-CoV genomes and two SARS-like coronavirus sequences from palm civets. Only those variant sequences (including deletions) that were present in at least two independent samples were used for tree construction. The map distance between individual sequences represents the extent of genotypic difference. The 5-nt motifs (see Example 1) that characterized the phylogenetically related genotypes are boxed. The genomic sequences are named in concordance with their GenBank nomenclature. Genotypes with major deletions are marked accordingly. All other genotypes (unmarked) had the 29-nt deletion. This 29-nt deletion was specifically marked for three genotypes, namely GZ-A, JMD, and GZ50, to indicate their special clustering within the early phase isolates.

[019] Figure 6 shows the genotype clustering of SARS-CoV and SARS-CoV-like coronaviruses from the outbreaks of 2002-2003 and 2003-2004. Specifically, an unrooted phylogenetic tree of SARS-CoV and SARS-CoV-like coronaviruses genomes was constructed based on the data of Examples 1 and 2. The map distance between individual sequences represents the extent of genotypic difference. To highlight the variations between neighboring clusters, the total number of single nucleotide variations (SNVs) as well as the numbers of synonymous and non-synonymous mutations (causing drastic amino acid changes) present in at least two independent samples are shown in the boxes.

[020] Figures 7A-C shows the phylogeny of the most variable genes, Spike (Figure 7A), sars3a (Figure 7B), and nsp3 (Figure 7C) in the SARS-CoV and SARS-CoV-like coronaviral samples from the early HP03-E cases of the 2002-2003 SARS outbreak, from the new HP04 cases of the 2003-2004 outbreak, and from the PC03 and PC04 palm civet sequences. The two numbers shown along each branch are the maximum-likelihood estimates of, respectively, the number of synonymous and nonsynonymous substitutions for each entire gene along that branch. In each tree, a different  $d_N/d_S$  ratio is assumed for each branch. The branch length is proportional to the total number of substitutions between sequences.

**DETAILED DESCRIPTION OF THE INVENTION**

[021] The present invention is directed to: the characterization of different SARS-CoV strains occurring at different stages of a SARS outbreak; the identification of stage-specific molecular markers characteristic of one or more of these different stages of a SARS outbreak; to the sequences of SARS-CoV and SARS-CoV like coronaviral strains from these different stages; to vectors, cells, and animals containing these sequences; to the use of the nucleotide and polypeptide sequences of these strains; and, to the development of vaccines based on these sequences.

[022] As used herein, the term “stage-specific molecular marker” (synonymously, “stage-specific marker,” “stage-specific nucleotide marker,” “molecular marker,” “marker,” etc.) refers to single or multiple nucleotide positions that are characteristic of one or more members of the different groups of SARS-CoV and SARS-CoV-like strains that occur at different stages of a SARS outbreak. As used herein, stage-specific molecular markers are intended to include both single nucleotide variants (SNVs), and also insertions and deletions in the SARS-CoV or SARS-CoV-like genome.

[023] A stage-specific molecular marker is said to be characteristic of one or more members of the different groups of SARS-CoV and SARS-CoV-like strains that occur at different stages of a SARS outbreak when it occurs in some but not all of these groups. As is shown in Figure 1 and diagrammatically in Figure 2, some of the stage-specific molecular markers of the invention occur in only one group, while others occur in more than one group. As these figures also show, a stage-specific molecular marker of the invention need not occur in all members of a particular group, and may instead occur in a subset of the members of a group, including as few as only one member of a particular group.

[024] Although the present invention is directed to the characterization of different SARS-CoV strains occurring at different stages of a SARS outbreak, SARS-CoV strains obtained from the earliest infected patients in a SARS outbreak are of particular interest herein. Thus, as is discussed in detail in Example 1, the present invention is based on a sequence analysis of the SARS-CoV strains obtained from patients in the early, middle, and late stages of the 2002-2003 SARS outbreak, where, for the purposes of this 2002-2003 outbreak: “early-stage” (synonymously, “early phase”) refers to the period from the first emergence of SARS to the first documented superspreader event; “middle-stage”

(synonymously, “middle phase”) refers to the ensuing events up to the first cluster of SARS cases in a hotel (Hotel M) in Hong Kong; and, “late-stage” (synonymously, “late phase”) refers to cases following this Hotel M cluster.

[025] As discussed in detail in Example 2, the present invention is also based on a  
5 sequence analysis of the SARS-CoV strains obtained from two patients of the 2003-2004 SARS outbreak. As Example 2 demonstrates, this additional sequence analysis allows for the definition of a “very early” stage (synonymously, “early-early” stage) of SARS coronavirus. As used herein the term “earliest stage” refers to the combination of “early-early” and “early” stage coronaviral strains, i.e., both “early-early” and “early” stage strains are intended to be  
10 subsumed within this term. Finally, as discussed in detail in Examples 1 and 2, the present invention is also based on palm civet SARS-CoV-like coronavirus sequences obtained during both the 2002-2003 and 2003-2004 SARS outbreak periods, with these sequences used to better define the nature of the virus in the earliest stages of infection in humans.

[026] Thus the present invention is based on SARS-CoV coronavirus sequences  
15 from both the 2002-2003 and 2003-2004 SARS outbreaks and, additionally, on palm civet SARS-CoV-like sequences from these same two periods. In the present invention, the following nomenclature is used to identify these four groups of sequences: “HP03” is used to refer to SARS-CoV strains obtained from human patients of the 2002-2003 outbreak; “HP04” is used to refer to SARS-CoV strains obtained from human patients of the 2003-2004  
20 outbreak; “PC03” is used to refer to SARS-CoV-like strains obtained from palm civets during the period of the 2002-2003 outbreak; and, “PC04” is used to refer to SARS-CoV-like strains obtained from palm civets during the period of the 2003-2004 outbreak. In addition, as described in Example 1, the SARS-CoV strains of the 2002-2003 outbreak (i.e., the HP03 strains) are further classified into “early,” “middle,” and “late” stages, which are represented  
25 herein as the “HP03-E,” “HP03-M,” and “HP03-L” strains, respectively.

[027] Although detailed analyses of these data are presented in Examples 1 and 2 as provided elsewhere herein, a number of points relevant to the present invention will be summarized here. These summary points are not intended as limiting of the present invention, and represent a subset of the aspects of the present invention that are discussed at  
30 greater length elsewhere herein, and in particular in Examples 1 and 2.

[028] First, the comparison of the RNA sequences of the HP03-E, HP03-M, HP03-L, and PC03 SARS-CoV and SARS-CoV-like coronaviral groups presented in Example 1 identifies a variety of stage-specific molecular markers characteristic of these groups. Examples of these markers are provided in both the deletion/insertion sequences and the single nucleotide variations (SNVs) of Figures 1 and 2, and are discussed in detail in the “SARS Coronavirus Stage-Specific Markers” section below. Of note is the fact that a pentet of SNV positions are particularly informative in such classifications, namely the SNVs at positions 17564, 21721, 22222, 23823, and 27827 of the SARS-CoV sequence. As for elsewhere in the present invention, these sequence positions are given relative to the GZ02 HP03-E reference sequence (SEQ ID NO:1), unless otherwise noted.

### Sequence Identifiers

[029] The present invention refers to a number of sequence identifiers. In this regard: SEQ ID NO:1 refers to the GZ02 reference sequence used as the basis for the SARS-CoV and SARS-CoV-like RNA sequences given throughout the the specification; SEQ ID NO:2 refers to the 29 nucleotide insertion corresponding to positions 27884-27912 of the GZ02 reference sequence; SEQ ID NO:3 refers to the 53 nucleotide deletion spanning nucleotides 27858-27883 and 27912-27939 of the GZ02 reference sequence; SEQ ID NO:4 refers to the nucleotide and corresponding amino acid sequences of the spike (“S”) protein, which occurs at nucleotides 21492-25259 of the GZ02 reference sequence; SEQ ID NO:5 is the S protein amino acid sequence; SEQ ID NO:6 is the sense PCR primer of Example 3; and, SEQ ID NO:7 is the anti-sense PCR primer of Example 3.

### SARS Coronavirus Stage-Specific Molecular Markers

[030] One embodiment of the present invention is directed to the use of the data obtained in the sequence comparisons of the novel SARS-CoV and SARS-CoV-like sequences of the present invention to identify insertions, deletions, and single nucleotide variations (SNV) that are characteristic for each stage of SARS virus, i.e., that serve as molecular markers for the different strains characteristic of the earliest stages (including both “early-early” and “early” stages), middle, and late stages of an outbreak.

[031] Thus as is described in Example 1, the present invention provides for the comparison of 63 SARS-CoV and SARS-CoV-like sequences, including the following previously uncharacterized sequences: GZ02 (SEQ ID NO:1), HGZ8L1-A, HSZ-Cc, HSZ-A, HSZ-Bb, HSZ-Cb, HSZ-Bc, HGZ8L1-B, ZS-A, ZS-B, ZS-C, HZS2-D, HZS2-E, HZS2-C, HZS2-Bb, HZS2-A (all HP03-E sequences); HGZ8L-2, HZS2-Fc, HZS2-Fb (all HP03-M sequences); and, GZ-D, GZ-B, GZ-C, CUHK-LC2, CUHK-LC3, CUHK-LC4, and CUHK-LC5 (all HP03-L sequences).

[032] As is described in Example 1 and shown in Figures 1-5, these comparisons provide a number of insertions/deletions (Figures 2-5) and SNVs (Figures 1-2 and 5) that may be used to classify any particular SARS-CoV or SARS-CoV-like sequence within the HP03-E, HP03-M, HP03-L, and PC03 groups provided in this Example. Although these data do not include the HP04 and PC04 data of Example 2, they are also applicable to these groups of SARS-CoV and SARS-CoV-like strains.

[033] Thus with regard to the insertions/deletions of Example 1, these data show an insertion of 29 nucleotides (CCTACTGGTTACCAACCTGAATGGAATAT; SEQ ID NO:2) in the Orf8 region (this insert is part of the GZ02 reference sequence (SEQ ID NO:1) at positions 27884-27912, i.e., immediately after position 27883 at which the insertion occurs; the Orf8 region itself spans nucleotides 21637-28147 of the GZ02 sequence) in seven of the HP03-E SARS-CoV strains (GZ02, HGZ8L1-A, HSZ-A, HSZ-Bb, HSZ-Bc, HSZ-Cb and HSZ-Cc; in Example 1 the HSZ-Bb and HSZ-Bc strains are summarized as HSZ-B, and the HSZ-Cb and HSZ-Cc strains are summarized as HSZ-C), i.e., the same sequence in this region as observed for the PC03 sequences originally characterized by Guan *et al.* (*Science* 302:276 (2003)). By the middle phase of the HP03 outbreak, however, the characterized SARS-CoV strains lack this 29 nucleotide insertion (see, e.g., Rota *et al.*, *Science* 300:1394 (2003); Marra *et al.*, *Science* 300:1399 (2003); Ruán *et al.*, *Lancet* 361:1779 (2003)). Thus the presence of the 29 nucleotide insertion of SEQ ID NO:2 at position 27883 may be used in the identification of a SARS-CoV sequence as an early-stage sequence, i.e., this 29 nucleotide sequence serves as an example of what is termed herein to be an "insertion early-stage-specific molecular marker."

[034] Also as described in Example 1, a lung biopsy of a patient from the middle phase of the 2002-2004 outbreak (i.e., HP03-M) was found to contain two SARS-CoV genotypes. In both genotypes the 29 nucleotide deletion was observed; unique to this patient,

however, was the presence of an additional 53 nucleotide deletion bracketing either side of the region of the 29 nucleotide deletion (i.e.,

AACCTCATGTGCTTGAAGATCCTTGTAAGGTACAACACTAGGGGTAATACTTA

(SEQ ID NO:3) spanning nucleotides 27858-27883 and 27912-27939 of the GZ02 sequence

(SEQ ID NO:1)) in some (but not all) of the SARS-CoV clones obtained from this patient and analyzed by sequencing. Specifically, 17 of 27 clones from this patient had sequences

lacking these 53 nucleotides, while the remaining 10 clones had the 53 nucleotides in this position, i.e., had the same sequence in this region as was observed for other SARS-CoV

HP03-M strains. Finally, an 82 nucleotide deletion representing both the 29 and 53

nucleotide deletions was found in four more of the HP03-E strains (ZS-A, ZS-B, ZS-C, and

HGZ8L1-B). Thus these 53 and 82 (i.e., 29+53) nucleotide deletions may potentially also be

used in the classification of SARS-CoV strains, i.e., for the staging of a SARS-CoV strain,

similarly to the staging described above for the 29 nucleotide sequence. Thus these deletions

serve as examples of what are termed herein to be “deletion early-stage-specific molecular

markers.”

[035] With regard to the use of the SNVs of the present invention in staging of SARS-CoV strains, both Examples 1 and 2 provide a variety of SNVs that are useful for such staging (i.e., serve as what are termed to be “SNV stage-specific molecular markers”), with

Example 2 supplementing the data of Example 1. Specifically, Example 2 provides an

analysis of 5 additional coronavirus sequences: 2 SARS-CoV sequences obtained from two

of the four patients of the 2003-2004 SARS outbreak in Guangzhou (GZ03-02 and GZ03-01);

and, 3 palm civet SARS-CoV-like sequences obtained from palm civets in the Guangzhou

food market during the same period (PC4-13, PC4-136, and PC4-227). As shown in Figure 6

and described in Example 2, these sequences are exceptional in that they are more closely

related to one another than they are to either of the two HP03-E or PC03 outgroups (i.e.,

GZ03-02, GZ03-01, PC4-13, PC4-136, and PC4-227 all group together on the unrooted

phylogenetic tree shown in Figure 6; see also Figures 7A-C for a similar analysis using the

Spike, sars3a, or nsp3 sequence regions, respectively). These data suggest an “early-early”

stage containing the PC04 and HP04 sequences; see elsewhere herein for a complete

discussion.

[036] Figures 1 and 2 summarize the combined data of Examples 1 and 2 regarding SNVs useful in the present invention. Specifically, Figure 1 provides the SNVs for 92

SARS-CoV and SARS-CoV-like coronaviral sequences, with all 6 groups identified in Examples 1 and 2 provided in this Figure (i.e., HP04, PC04, PC03, HP03-E, HP03-M, and HP03-L).

[037] The skilled artisan will recognize that any nucleotide that is present in some of these groups and absent in others will act as a molecular marker either in light of its presence in one or more groups or, alternatively, in light of its absence in one or more groups. Thus the occurrence of a C at position 4160 in the GZ02 reference sequence (SEQ ID NO:1) is an example of a molecular marker of use in identifying SARS-CoV sequences falling into the HP04 group, as is shown both in Figure 1 and in Figure 2. Referring to Figure 1, it is apparent that position 4160 is a C for both the HP04 sequences; for all other sequences in the remaining groups (PC04, PC03, HP03-E, HP03-M, and HP03-L), this position is an A. This pattern is also shown graphically in Figure 2, where only the column labeled "HP04" in the section labeled "Nucleotide Present in All Members of Group Below" is indicated, thereby indicating that this nucleotide (C) at this position (4160) is predictive of a sequence falling within the HP04 group of SARS-CoV strains. On the basis of the observed pattern in the data of Figures 1 and 2, it is clear that a new sequence containing a C in this position must be a member of the HP04 group of strains; alternatively, either the presence of an A in this position or the absence of a C would indicate the new sequence belonged to one of the 5 remaining groups of strains (i.e., PC04, PC03, HP03-E, HP03-M, or HP03-L).

[038] Figure 2 shows a number of nucleotides in this category, i.e., nucleotides which are present in all members of only one of the six groups HP04, PC04, PC03, HP03-E, HP03-M, and HP03-L, and which when present may therefore be used to indicate the membership of the sequence containing that nucleotide in the particular group which the nucleotide designates. Specifically, Figure 2 provides non-limiting examples of SNVs indicative of: HP04 (T at position 185; C at position 4160; T at position 14151; C at position 20885; and, T at position 27869); PC04 (T at position 17390); PC03 (T at position 3671; T at position 6456; A at position 22172; A at position 22273; C at position 23310; T at position 23593; G at position 23718; T at position 23752; G at position 24171); and, T at position 25544); HP03-E (G at position 4220); and, HP03-L (T at position 27827).

[039] Although the present invention contemplates the molecular markers described above as being particularly useful in identifying the stage of a SARS coronavirus strain or SARS-CoV-like strain, the invention is also directed to any SNV at any position that can be

used to assist in the identification (staging) of a SARS-CoV or SARS-CoV-like sequence. For example, the present invention specifically contemplates markers which, while associated only with a particular group of SARS coronavirus strains, occur only in some members of that group. A strain containing a C at position 1727, for example, is from the data of Figures 1 and 2 a member of the HP03-L group; however, unlike the markers discussed above, because only some members of the HP03-L group have a C in this position, the absence of a C in this position is not conclusive as to group membership.

[040] Also explicitly contemplated herein are molecular markers which occur in all members of multiple groups (e.g., the “EE” markers discussed below), as well as molecular markers which occur in some members of multiple groups (e.g., the occurrence of a T at position 9404), and markers which occur in some members of some groups and all members of others (e.g., a C at position 9479). As discussed above, the skilled artisan will recognize that the requirement for a suitable molecular marker useful for staging is satisfied by any insertion(s), deletion(s), SNV(s), or combination or combinations thereof which allow one of ordinary skill in the art to discriminate between the different groups of SARS-CoV and SARS-CoV-like strains. Thus, for example, a nucleotide that is present in all sequences of every group at a particular position is not suitable as a molecular marker, nor is a nucleotide that is found in at least some sequences in every group suitable as a marker.

[041] Although the present invention contemplates molecular markers for identifying membership in one of the six groups of strains PC04, PC03, HP03-E, HP03-M, and HP03-L identified in Examples 1 and 2, the present invention is also directed to the identification of two other groups of strains referred to in these Examples, namely, the “early-early” and “earliest stage” groups of the Examples. Thus as described in Example 2, based on the data obtained for the PC04 and HP04 SARS-CoV-like and SARS-CoV strains, the HP03-E strains should more appropriately be characterized as the “relatively late stage of the early phase,” with the PC04/HP04 groups together representing the “early-early” stage of the coronaviral evolution in humans. Markers suitable for identifying coronaviruses that are members of this “EE” group are shown in Figure 2, and include, but are not limited to: A at position 2013; T at position 2606; T at position 2760; C at position 3567; G at position 3584; A at position 4108; G at position 5142; C at position 5811; T at position 6990; T at position 7137; C at position 7881; C at position 9335; A at position 10977; C at position 12119; G at position 13760; A at position 14117; C at position 14300; T at position 17374; G at position

19406; T at position 21907; G at position 21931; T at position 22874; C at position 22906; G at position 22930; G at position 23719; T at position 23785; C at position 25031; A at position 25341; G at position 25356; A at position 25693; G at position 26436; T at position 27425; A at position 27893; and, G at position 29022. Similarly, “earliest stage” coronviral sequences are those falling in the group containing the HP04, PC04, PC03, and HP03-E sequences. An example of an SNV marker suitable for identifying this group is the G at position 23823 shown under the “E” column in Figure 2.

[042] As discussed in detail in Example 1, one set of SNVs of particular relevance in the present invention is the pentet of SNVs at positions 17564, 21721, 22222, 23823, and 27827 relative to the GZ02 reference sequence (SEQ ID NO:1) (see also Figure 5). These positions are shown by the corresponding shaded columns in Figure 1, and have the values shown in the Table 1 below for each of the groups HP04, PC04, PC03, HP03-E, HP03-M, and HP03-L (note that this table is based on the exclusion of the single G nucleotide occurring at position 17564 in the HP03-L strain QXC1, the single T nucleotide occurring at position 22222 in the HP03-E strain GZ-A, and, the unidentified nucleotide “N” occurring at position 23823 in the HSZ-A strain):

Table 1

| Gene       | Nsp13 | S     | S     | S     | Sars8a |
|------------|-------|-------|-------|-------|--------|
| Nucleotide | 17564 | 21721 | 22222 | 23823 | 27827  |
| AA         | 5767  | 77    | 244   | 778   | 17     |
| AA Switch  | E-D   | D-G   | T-I   | D-Y   | R-C    |
| HP04       | G     | A     | C     | G     | C      |
| PC04       | G     | A     | C     | G     | C      |
| PC03       | G     | A     | C     | G     | C      |
| HP03-E     | G     | A     | C     | G     | C      |
| HP03-M     | G     | A     | C     | T     | C      |
| HP03-L     | T     | G     | T     | T     | T      |

[043] Thus as is shown in Table 1, the following pentets of SNVs at these positions are associated with each of the six SARS-CoV and SARS-CoV-like groups of strains shown above: GACGC (HP04, PC04, PC03, HP03-E); GACTC (HP03-M); and, TGTTC (HP03-L). Thus this pentet may be used as the basis for distinguishing these groups.

[044] Alternatively, these positions may be used singly, or in various other combinations with each other or with other SNVs or insertions/deletions to classify SARS-CoV or SARS-CoV-like sequences. For example, the presence of a T at position 17564 is indicative of an HP03-L sequence, the presence of a T at position 23823 indicates membership in either the HP03-M or HP03-L groupings, etc.

[045] The present invention also contemplates molecular markers within the region of the SARS-CoV or SARS-CoV-like coronavirus genome encoding the coronaviral Spike ("S") protein (SEQ ID NOS:4 and 5; the S protein is encoded in the GZ02 reference sequence (SEQ ID NO:1) at nucleotides 21492-25259)). This protein is a major component of the interaction between the coronavirus and the host, and mutations in this protein can therefore be expected to closely correlated with the infectivity of the coronavirus. Based on the data of Figures 1 and 2, as well as the preceding discussion, SNVs at positions 22570, 22927-22928, 22930 and 23316-23317 of the nucleotide sequence of the SARS coronaviral genome encoding the S protein may also have utility as specific markers for staging SARS-CoV and SARS-CoV-like strains.

[046] Thus a C at position 22570 is characteristic of the HP04, PC04, and PC03 groups (see Figures 1 and 2), and therefore can be used to identify sequences falling into this set of strains. With regard to positions 22927-22928, this pair of SNVs represents a diad which is either AT (corresponding to Asn) for all SARS-affected individuals, or GA/AA (corresponding to Arg or Lys, i.e., to a positively-charged amino acid) for 4 of the 5 palm civet sequences determined (i.e., 22927-212928 is AT only for one palm civet sequence, the PC04 PC-227 sequence. See Figure 1). This observation suggests that the change from the predominant palm civet GA/AA diad to the AT diad found in all human cases may be important to the successful propagation of the coronavirus in humans. Therefore, it is likely that this diad at 22927-22928 will be useful as a molecular marker for staging and, possibly, as a proxy for viral infectivity or success in a human host.

[047] Similarly, at positions 23316-23317 there are only two observed diads: TT, which predominates in the "early-early" group PC04/HP04 (of the five sequences in this group, only one, that of the HP04 sequence GZ03-02 is GC instead of TT); and, GC which is present in all of the members of the PC03, HP03-E, HP03-M, and HP03-L groups. Thus, as for the 22927-22928 diad, the SNVs at 23316-23317 may be useful as a marker for staging or as a proxy for infectivity or survival/reproduction in a human host.

[048] Further evidence for the utility of the diad pairs discussed above is provided by modeling studies suggesting their proximity to regions of the S protein of functional significance. Thus the present invention contemplates those nucleotide positions corresponding to important functional regions of one or more of the SARS-CoV or SARS-CoV-like proteins as being particularly useful, in light of their likely evolutionary constraints and/or selection pressures. For the S protein, structural modeling suggests that the amino acids in the region of amino acids 311-487 (corresponding to nucleotides 22422-22951 of the reference GZ02 sequence (SEQ ID NO:1)) are involved in the interaction of this protein with its receptor protein, ACE2. Modeling studies also suggest that another region of the S protein at amino acids 577-613 (corresponding to nucleotides 22422-22951) is important in the formation of trimers of the S protein. Thus in light of the contemplated importance of such functional regions in the present invention, the role of these two diad regions in staging is likely even apart from the other evidence for such utility given above.

[049] Also with regard to the insertions/deletions and SNVs discussed above, although the present invention contemplates the use of any of the insertions/deletions and SNVs detailed in the preceding section, in the figures, or in the Examples, the skilled artisan will understand that some of these markers are expected to be of particular utility in the staging discussed in this section. For example, nucleotide changes that result in changed (non-synonymous) amino acids are expected to be under greater selection pressures than those which result in synonymous changes. Therefore it is likely that in some situations those nucleotide changes which result in non-synonymous amino acid changes will be of greater utility than those which cause synonymous changes. Similarly, nucleotide positions which encode amino acids of proteins of the coronavirus that are under selection pressure are also expected to be of potentially particular importance, as discussed above. Therefore, one of ordinary skill in the art of molecular marker selection would know to pick particular markers based not only on the extensive data provided herein, but also on the basis of such additional considerations as discussed.

[050] After identifying particular insertion(s)/deletion(s) or SNV(s) that are useful as stage-specific molecular markers as described above, it is necessary to determine the presence or absence of such stage-specific markers in a SARS-CoV or SARS-CoV-like sequence of interest.

[051] In this regard it will first be necessary to obtain the RNA sequence of interest, either directly from a patient as isolated SARS-CoV or SARS-CoV-like coronaviral RNA, or from cultures of cells infected with the SARS-CoV or SARS-CoV-like coronaviral strain from which the sequence of interest is to be obtained. Such purification and culturing methods are described elsewhere herein, and are routine to one of ordinary skill in the art.

[052] After obtaining the RNA sequence of interest, an analysis can be conducted directly, using various RNA-based methods such as RT-PCR, or via an isolated cDNA produced using reverse transcriptase in conditions that are also well-known to the skilled artisan. In the latter case the cDNA may be analyzed by a variety of techniques discussed below.

[053] In light of the above discussion, and in view of subsequent discussions of the proteins contemplated herein, it is therefore clear that the present invention encompasses isolated or substantially purified nucleic acid or protein compositions. An "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof, is 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. Preferably, an "isolated" nucleic acid is free of 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 that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, 5%, (by dry weight) of contaminating protein. When the protein of the invention or biologically active portion thereof is recombinantly produced, preferably, culture medium represents less than about 30%, 20%, 10%, or 5% (by dry weight) of chemical precursors or non-protein of interest chemicals.

[054] As discussed, one method of analysis contemplated herein is based on the direct analysis of the SARS-CoV or SARS-CoV-like coronaviral RNA using techniques including RT-PCR, while another method of analysis contemplated is an analysis based on PCR amplification of the isolated cDNA obtained from the RNA by reverse transcription.

[055] In a PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from coronaviral cDNA. Methods for designing PCR primers or RT-PCR primers and PCR amplification are generally known in the art and are disclosed in Sambrook et al. (1989) *Molecular Cloning. A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, N.Y.). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially-mismatched primers, and the like.

[056] In PCR amplification, sense and anti-sense primers are hybridized to the sequence of interest, which is then amplified in an amplification reaction. Thus the present invention contemplates the use of various hybridization techniques for PCR reactions, as well as for other analytical methods known to one of ordinary skill.

[057] In hybridization, all or part of a known nucleotide sequence is used as a probe that selectively hybridizes to other corresponding nucleotide sequences, e.g., a SARS-CoV or SARS-CoV-like coronaviral cDNA sequence. In general hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, as appropriate, and may be labeled with a detectable group such as  $^{32}\text{P}$ , or any other detectable marker. Thus, for example, probes for hybridization can be made by labeling synthetic oligonucleotides based on the coronaviral sequences of the invention. Methods for preparation of probes for hybridization and PCR are generally known in the art and are disclosed in Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, N.Y.).

[058] An important parameter in hybridizations is the specificity of hybridization between the template and probes. Thus to achieve specific hybridization under a variety of conditions, such probes include sequences that are unique to the desired region of the coronaviral sequence, and are preferably at least about 10 nucleotides in length, and most preferably at least about 20 nucleotides in length. In PCR reactions, such probes may be used to amplify corresponding coronaviral sequence regions of interest, or as a diagnostic assay to determine the presence of particular sequence regions or individual nucleotides in a coronaviral template nucleotide sequence.

[059] Hybridizations may be carried out under different conditions of stringency, for example under stringent conditions. By "stringent conditions" or "stringent hybridization conditions" is intended conditions under which a probe will hybridize to its target sequence to a detectably greater degree than to other sequences (e.g., at least 2-fold over background).

5 Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences that are 100% complementary to the probe can be identified (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally,  
10 a probe is less than about 1000 nucleotides in length, preferably less than 500 nucleotides in length.

[060] Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes (e.g., 10 to  
15 50 nucleotides) and at least about 60 °C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37 °C, and a wash in 1x to 2x SSC (20x SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55 °C.  
20 Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37 °C, and a wash in 0.5x to 1x SSC at 55 to 60 °C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37 °C, and a wash in 0.1x SSC at 60 to 65 °C. Duration of hybridization is generally less than about 24 hours, usually about 4 to about 12 hours.

25 [061] Specificity is typically the function of post-hybridization washes, with the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the  $T_m$  can be approximated from the equation of Meinkoth and Wahl (*Anal. Biochem.* 138:267 (1984)):  $T_m = 81.5\text{ °C} + 16.6 (\log M) + 0.41 (\% \text{ GC}) - 0.61 (\% \text{ form}) - 500/L$ ; where M is the molarity of monovalent cations, % GC is the percentage of guanosine and  
30 cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes

to a perfectly matched probe.  $T_m$  is reduced by about 1 °C for each 1% of mismatching; thus,  $T_m$ , hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with about 90% identity are sought, the  $T_m$  can be decreased 10 °C. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point ( $T_m$ ) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4 °C lower than the thermal melting point ( $T_m$ ); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10 °C lower than the thermal melting point ( $T_m$ ); and, low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20 °C lower than the thermal melting point ( $T_m$ ). Using the equation, hybridization and wash compositions, and desired  $T_m$  those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a  $T_m$  of less than 45 °C (aqueous solution) or 32 °C (formamide solution), it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 (Elsevier, N.Y.); and Ausubel et al., eds. (1995) *Current Protocols in Molecular Biology*, Chapter 2 (Greene Publishing and Wiley-Interscience, New York). See also Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, N.Y.).

[062] In order to determine the presence or absence of a molecular marker in the SARS-CoV or SARS-CoV-like sequence of interest it will additionally be necessary to determine the position of each nucleotide in the SARS-CoV or SARS-CoV-like sequence of interest relative to the positions of the stage-specific molecular markers described herein, which are themselves given relative to the GZ02 reference sequence (SEQ ID NO:1). In this regard the skilled artisan will understand that it will be necessary to align the SARS-CoV or SARS-CoV-like sequence of interest with the GZ02 reference sequence, using any of the well-known methods for such alignments, as are described elsewhere herein. Such an alignment will allow for the correlation of each position in the SARS-CoV or SARS-CoV-like sequence of interest with the corresponding position of the GZ02 sequence.

[063] Although the present invention contemplates a variety of embodiments for detecting insertion(s)/deletion(s) or SNV(s) of interest, one non-limiting embodiment of particular interest is a kit for analyzing these insertion(s)/deletion(s) or SNV(s). Thus based on the present disclosure, it is possible to specifically PCR-amplify regions of the SARS-CoV or SARS-CoV-like coronaviral genome that contain the insertion(s)/deletion(s) or SNV(s) of interest, and then analyze the resulting material by sequencing or other methods known to the skilled artisan. Example 3 provides for one such kit.

### SARS Coronavirus Nucleotide Sequences

[064] Another embodiment of the present invention is directed to the SARS coronaviral nucleotide sequences characteristic of the HP04, PC04, PC03, HP03-E, HP03-M, and HP03-L groups, and particularly to the nucleotide sequences characteristic of earliest stage SARS-CoV sequences (i.e., "early-early" and "early" stages), as well as closely-related SARS-CoV-like sequences. The availability of these sequences allows for the more complete study of the characteristics of the SARS-CoV or SARS-CoV-like coronaviruses during their evolution, and particularly during their earliest stages of their evolution, i.e., shortly after crossing the species barrier from its non-human source into humans. These earliest stages are of particular interest in the present invention, in light of the reduced virulence and infectivity of coronaviruses from these stages (e.g., the HP04 strains of 2002-2004; see Example 2).

[065] Thus the present invention is directed to sequences characteristic of the HP04, PC04, PC03, HP03-E, HP03-M, and HP03-L groups. These sequences may be obtained by substituting the SNVs provided in Figure 1 and the insertions/deletions provided elsewhere herein into the GZ02 reference sequence provided in SEQ ID NO:1. Thus, for example, with regard to SNVs, the sequence of GZ03-02 may be obtained by substituting into the GZ02 reference sequence a T nucleotide at position 185, a G nucleotide at position 508, a C nucleotide at position 1302, etc., as determined by a comparison of the rows in Figure 1 corresponding to the GZ02 and GZ03-02 sequences.

[066] In one aspect, the present invention is directed to the exact sequences of the SARS-CoV and SARS-CoV-like coronaviral strains provided in Figure 1, which may be obtained as described above. In another aspect, the present invention is directed to sequences related to these sequences either by % identity (synonymously, "% similarity"), by the

presence of particular nucleotide(s), insertion(s), and/or deletion(s) at particular positions, or by all of these criteria.

[067] With regard to % identity, the following terms are used to describe the sequence relationships between two or more nucleic acids, polynucleotides, or polypeptides: “reference sequence”; “comparison window”; “sequence identity”; “percentage of sequence identity”; and, “substantial identity.” Note that this discussion is explicitly intended to encompass both the nucleotide sequences discussed in this section, and the polypeptides of the next section.

[068] Thus as used herein, “reference sequence” is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence. Thus reference sequences of the present invention include SARS-CoV and SARS-CoV-like sequences, as well as subsets of these sequences, such as fragments or variants. By “fragment” is intended a portion of a nucleotide or amino acid sequence of the present invention; by “variants” is intended substantially similar sequences.

[069] As used herein, “comparison window” makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches. The present invention contemplates that analogous considerations will apply to polypeptide sequences.

[070] Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent sequence identity between any two sequences can be accomplished using a mathematical algorithm. Preferred, non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988) CABIOS 4:11-17; the local homology algorithm of Smith *et al.* (1981) Adv. Appl. Math. 2:482; the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443-453; the search-

for-similarity-method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; and, the algorithm of Karlin and Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264, modified as in Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5877.

[071] Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA).

Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. (1988) Gene 73:237-244 (1988); 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. The ALIGN program is based on the algorithm of Myers and Miller (1988) supra. A PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used with the ALIGN program when comparing amino acid sequences. The BLAST programs of Altschul et al (1990) J. Mol. Biol. 215:403 are based on the algorithm of Karlin and Altschul (1990) supra. BLAST nucleotide searches can be performed with the BLASTN program, score=100, wordlength=12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the invention. BLAST protein searches can be performed with the BLASTX program, score=50, wordlength=3, to obtain amino acid sequences homologous to a protein or polypeptide of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. (1997) Nucleic Acids Res. 25:3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul et al. (1997) supra. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTX for proteins) can be used. See [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). Alignment may also be performed manually by inspection.

[072] For purposes of the present invention, comparison of nucleotide or protein sequences for determination of percent sequence identity to the promoter sequence or the antipathogenic sequences disclosed herein is preferably made using the Clustal W program

(Version 1.7 or later) with its default parameters or any equivalent program. By “equivalent program” is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment  
5 generated by the preferred program.

[073] As used herein, “sequence identity” or “identity” in the context of two nucleic acid or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that  
10 residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature  
15 of the substitution. Sequences that differ by such conservative substitutions are said to have “sequence similarity” or “similarity.” Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative  
20 substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, Calif.).

[074] As used herein, “percentage of sequence identity” means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the  
25 portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions,  
30 dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

[075] The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70% sequence identity, preferably at least 80%, more preferably at least 90%, and most preferably at least 95%, compared to a reference sequence using one of the alignment programs described using standard parameters.

5 One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 60%, more preferably at least 70%, 80%, 90%, and most  
10 preferably at least 95%.

[076] An additional indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. However, stringent conditions  
15 encompass temperatures in the range of about 1 °C to about 20 °C lower than the  $T_m$ , depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic  
20 code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

[077] The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70% sequence identity to a reference sequence,  
25 preferably 80%, more preferably 85%, most preferably at least 90% or 95% sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443-453. An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the  
30 second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. Peptides that are

“substantially similar” share sequences as noted above except that residue positions that are not identical may differ by conservative amino acid changes.

[078] With regard to the coronavirus nucleotide sequences of the invention, as discussed above such sequences are contemplated to include both the exact sequences presented herein (e.g., the strains in Figure 1), and also sequences that are related to these sequences by the presence of particular nucleotide(s), insertion(s), or deletion(s) at particular positions, by % identity, or by all of these criteria.

[079] Thus, in one embodiment, the present invention contemplates sequences that are fixed at one or more of the SNV positions of Figure 1, and/or the insertion/deletions characterized elsewhere herein, but that are allowed to vary at all other non-fixed positions. While it is theoretically possible for such sequences to be fixed at only a single position, it is preferable to have a sequence fixed at multiple positions, in order to limit the size of the resulting genus of RNA species defined by this mixture of fixed and variable positions.

[080] In this regard, one embodiment of the present invention is directed to genres of sequences that have some or all of the hallmark characteristics of one of the groups of SARS-CoV and SARS-CoV-like strains of the invention (e.g., HP04, PC04, HP03-E, HP03-M, HP03-L, “EE,” and “E,” as defined elsewhere herein). Thus for example, the present invention contemplates a genus of SARS-CoV RNA sequences (and viruses resulting from assembly with such sequences, cells containing such viruses, etc., as described below) defined by having a base SARS-CoV RNA sequence such as the GZ02 reference sequence (SEQ ID NO:1) which has been modified at some or all of the appropriate positions so as to possess some or all of the characteristic SNVs, insertions, and deletions of a particular group or groups of strains of the invention.

[081] Thus for example the appropriate positions in the GZ02 sequence may be modified to contain some or all of the nucleotides shown in Figure 2 as being characteristic of the “early-early” (“EE”) group of HP04/PC04. Thus the GZ02 sequence would have its usual sequence throughout its length, apart from one or more of the following positions, which would be assigned as follows: A at position 2013; T at position 2606; T at position 2760; C at position 3567; G at position 3584; A at position 4108; G at position 5142; C at position 5811; T at position 6990; T at position 7137; C at position 7881; C at position 9335; A at position 10977; C at position 12119; G at position 13760; A at position 14117; C at position 14300; T

at position 17374; G at position 19406; T at position 21907; G at position 21931; T at position 22874; C at position 22906; G at position 22930; G at position 23719; T at position 23785; C at position 25031; A at position 25341; G at position 25356; A at position 25693; G at position 26436; T at position 27425; A at position 27893; and, G at position 29022.

5           [082] Similarly, the pentet of positions provided in Table 1 may also be used as the basis for creating such genres of sequences. Consider, for example, a GZ02 sequence with position 17564 = T, 21721 = G, 22222 = T, 23823 = T, and 27827 = T. Solely with regard to these SNVs (i.e., omitting from consideration any insertions/deletions), this sequence possesses some of the hallmarks of an HP03-L sequence; i.e., any classification of this novel  
10           sequence based on a test of this pentet of nucleotides would indicate it to fall within the HP03-L category.

          [083] As mentioned above, although the nucleotide sequences of all of the groups defined herein are of interest, the earliest stage groups are of particular interest in light of the reduced virulence and infectivity of coronaviruses from these stages (e.g., the HP04 strains of  
15           2002-2004; see Example 2). Thus the present invention is particularly directed to the nucleotide sequences characteristic of these groups, e.g., nucleotide sequences containing the insertions/deletions/SNVs which are the hallmarks for the "early-early" (HP04/PC04) and/or "early" (HP04/PC04/PC03/HP03-E) stages. Thus sequences with, for example, the pentet shown in Table 1 as being characteristic of the "early" stages (position 17564 = G, 21721 =  
20           A, 22222 = C, 23823 = G, and 27827 = C) and with the base sequence of, e.g., GZ02, are preferred. In this example, sequences with this pentet of fixed positions, and with the remaining positions those of, e.g., GZ02, but varied so that the sequence is of high % identity to the base GZ02 sequence, are particularly preferred, as described below.

          [084] The present invention contemplates that the embodiments given above may  
25           include both those embodiments in which the base sequence is unvaried except for the hallmark insertions/deletions/SNVs etc. introduced as described, and those embodiments in which those positions of the base sequence which are not fixed by the hallmark insertions/deletions/SNVs etc. are allowed to vary. In this latter case, the present invention contemplates as particularly useful those genres of RNA species in which this variation of  
30           the base sequence is limited, i.e., to situations where the % identity between the novel sequence obtained by variation (and containing the fixed positions) is relatively high. Thus for example, the present invention contemplates situations in which the % identity between

the non-fixed positions of the base sequence and the original base sequence is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9%. This % identity may be judged by an alignment over the entire length of the SARS-CoV or SARS-CoV-like RNA sequence (i.e., over the  
5 approximately 29,000 bases of the RNA sequence), or it may be determined over a shorter length of the sequence, for example, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, etc. (i.e., continuing by increments of 10 nucleotides up to the maximum length of the  
10 RNA). This % identity may be calculated by one of the algorithms described elsewhere herein; alternatively, it may be calculated as the number of different nucleotides per 100 nucleotides, such that a % identity of 99.9% would refer to no more than 1 nucleotide difference per 1000 nucleotides.

[085] The present invention contemplates not only the sequences discussed above,  
15 but also the assembly of SARS-CoV or SARS-CoV-like coronaviruses containing these sequences, as well as vectors containing all or parts (fragments) of these sequences, as well as cells or animals containing these vectors or sequences. Methods for producing these constructs, cells, animals, etc., are well-known to the skilled artisan. For an example of assembly of a SARS coronavirus particle from a template cDNA see, e.g., Yount *et al.* (*Proc.*  
20 *Natl. Acad. Sci. U.S.A.* 100:12995 (2003)).

### SARS Coronavirus Proteins

[086] Another embodiment of the present invention is directed to amino acid sequences (synonymously, "polypeptides") encoded by the SARS-CoV or SARS-CoV-like  
25 nucleotide sequences of the present invention, as well as to vectors containing these amino acid-encoding nucleotide sequences, cells containing these vectors, and animals containing these nucleotide sequences, vectors, or cells.

[087] Any of the proteins encoded by the SARS-CoV or SARS-CoV-like coronavirus are contemplated for use in the present invention. For SARS-CoV the following  
30 proteins have been identified: orf1ab, orf1a, S, sars3a, sars3b, E, M, sars6, sars7a, N, sars9b, and nspl-16. The beginning and ending nucleotides for each of these proteins relative to the

GZ02 reference sequence (SEQ ID NO:1) are as follows: orflab (265-13398,13398-21485); orfla (265-13413); S (21492-25259); sars3a (25268-26092); sars3b (25689-26153); E (26117-26347); M (26398-27063); sars6 (26913-27265); sars7a (27273-27641); sars7b (27638-27772); N (27779-29417); sars9b (28159-28455); nsp1 (265-804); nsp2 (805-2718);  
 5 nsp3 (2719-8484); nsp4 (8485-9984); nsp5 (9985-10902); nsp6 (10903-11772); nsp7 (11773-12021); nsp8 (12022-12615); nsp9 (12616-12954); nsp10 (12955-13371); nsp11 (13372-13410); nsp12 (13372-13398, 13398-16166); nsp13 (16167-17969); nsp14 (17970-19550); nsp15 (19551-20588); and, nsp16 (920589-21482).

[088] Analogous to the SARS-CoV and SARS-CoV-like nucleotide sequences of the invention discussed in the previous section, the amino acid sequences of the present invention include the exact amino acid sequences obtained by introducing the changes shown in Figure 1 into the amino acid sequences encoded by the GZ02 reference sequence (SEQ ID NO:1). The amino acid sequences of the invention also include amino acid sequences that are related to these exact sequences, but are different as a result of the introduction or removal of  
 15 insertions/deletions/SNVs into the SARS-CoV or SARS-CoV-like coronaviral RNA that encodes these proteins, as well as additional changes added to other (non-fixed) positions which preserve a high % identity between the novel protein sequence and the sequence of the protein encoded in the original GZ02 reference sequence (i.e., the base amino acid sequence).

[089] Thus for example, the present invention contemplates situations in which the  
 20 % identity between the non-fixed positions of the novel amino acid sequence and the original (base) amino acid sequence is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9%. This % identity may be judged by an alignment over the entire length of the amino acid sequence, or it may be determined over a shorter length of the sequence, for example, 10, 20, 30, 40, 50, 60, 70,  
 25 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, etc. (i.e., continuing by increments of 10 amino acid residues up to the maximum length of the polypeptide of interest). This % identity may be calculated by one of the algorithms described elsewhere herein; alternatively, it may also be calculated  
 30 as the number of different amino acids per 100 nucleotides, such that a % identity of 99% would refer to no more than 1 amino acid difference per 1000 nucleotides.

[090] One protein of particular interest in the present invention is the S protein. As discussed previously, this protein mediates the interaction of the coronavirus with a host cellular receptor. Therefore this protein is of high importance in both the virulence and infectivity of a SARS coronavirus. Since the HP03-L stage is correlated with generally increased infectivity and virulence, particularly in relation to the HP04 and other earliest stage sequences, the present invention specifically contemplates S protein amino acid

sequences that contain the hallmark amino acids that correspond to each of these stages. With regard to the pentet of Table 1, for example, 3 of the SNVs listed fall within the region of the nucleotide sequence encoding the S protein (21721, 22222, and 23823). Thus the

present invention contemplates S proteins characteristic of either the HP04/PC04/PC03/HP03-E group (i.e., with residue 77 = D, residue 244 = T, and residue 778 = D), the HP03-L group (i.e., with residue 77 = G, residue 244 = I, and residue 778 = Y), or the (intermediate) HP03-M group (i.e., with residue 77 = D, residue 244 = T, and residue 778 = Y). These proteins are fixed at these positions; at the remaining positions they may either be fixed (e.g., corresponding to the base sequence) or they may be varied so as to preserve a high % identity to the base sequence, as discussed above.

[091] As discussed elsewhere herein, the SNVs at positions 22570, 22927-22928, 22930 and 23316-23317 of the nucleotide sequence of the SARS coronaviral genome encoding the S protein may be used as staging markers. In addition, the corresponding changes in amino acid as a result of mutations at these positions (namely 360 F → S, 479 N → R or K, 480 D → G, 609 A → L) may cause reduction in infectivity of SARS virus without affecting the immunogenicity of the S protein. Therefore the present invention also contemplates S protein sequences containing these amino acids in fixed positions, with the other positions either those of the base polypeptide sequence, or the base polypeptide sequence varied as described elsewhere herein (e.g., having a set % identity to the original base polypeptide sequence).

[092] The amino acid positions of the S protein described above provide examples of what are referred to herein as "stage-specific amino acids," i.e., amino acids which serve to indicate stage-specificity in a manner analogous to the stage-specific nucleotide markers described previously. Thus a "stage-specific amino acid" is an amino acid that is encoded by the SARS-CoV or SARS-CoV-like genomes of some, but not all, of the groups of SARS-CoV or SARS-CoV-like strains presented elsewhere herein. Table 1 provides four examples

of such stage-specific amino acids: the D amino acid residue corresponding to amino acid position 5767 of Orflab, which is specific for the HP03-L stage; the G amino acid residue corresponding to amino acid position 77 of the S protein, which is specific for the HP03-L stage; the I amino acid residue corresponding to amino acid position 244 of the S protein, which is specific for the HP03-L stage; and, the C amino acid residue corresponding to amino acid position 17 of the sars8a protein, which is specific for the HP03-L stage.

[093] Also contemplated as an additional parameter affecting the allowed variation of a protein sequence is the activity of that sequence. That is, in addition to a requirement that the non-fixed positions of a sequence be varied only to the extent that the sequence retains a particular % identity to the original sequence, the present invention contemplates the possibility of an independent or additional requirement that the sequence be allowed to vary only to the extent that it retains the functionality of the original polypeptide, or, if the sequence in question is a fragment of the polypeptide, that it retains the activity characteristic of that original fragment.

[094] Activity, as contemplated herein, will depend upon the specific protein or portion thereof, and will therefore be assayed by whatever method is commonly used for that particular protein or protein portion. For the S protein, for example, the skilled artisan will understand that appropriate assays include those measuring interaction with the host, or assays that reflect one component activity of the entire protein associated with a particular region of the protein of interest, e.g., interaction with ACE2, etc.

### SARS Coronavirus Vaccines

[095] Another embodiment of the present invention is directed to the development of vaccines for the SARS-CoV coronavirus. Thus as discussed elsewhere herein, and particularly in Example 2, the earliest stage SARS coronavirus strains characterized in the present invention are in general less virulent or contagious than are later stage strains of the SARS coronavirus. This observation suggests the particularly great utility of earliest stage coronavirus in the development of vaccines for the disease. To date, however, vaccine candidates have in general been developed from later stage SARS coronaviral sequences, which are both more readily available and easier to grow than earliest stage strains. Thus one aspect of the present invention involves the production of vaccines based on these earliest

strains, including whole-virus vaccines, and vaccines based on components of these strains, e.g., the S proteins characteristic of these earliest stages of the coronavirus.

[096] As stated above, there are a number of advantages to the use of earliest stage SARS coronaviral strains, or components thereof, in the development of vaccines. First, although middle- or late-stage SARS coronaviral strains are easy to culture, their high virulence and infectivity constitute a high risk for their safe use. In contrast, earliest stage strains, although difficult to culture, are less virulent and contagious, and therefore have an advantageous safety profile. In addition, studies have suggested that the immunity elicited by these earliest stage strains is likely sufficient to neutralize later-occurring strains.

[097] The present invention contemplates vaccines based on whole-coronavirus vaccines, including live-attenuated and inactivated coronavirus vaccines. The present invention also contemplates vaccines based on various components of the virus, e.g., based on the S protein. Also contemplated are vaccines based on antibodies against the coronavirus, or component or components thereof (see, e.g., ter Meulen *et al.*, *Lancet* 363:2139 (2004)). Particularly contemplated in the present invention are those vaccines based on earliest stage coronaviruses, or components thereof.

[098] In the whole-coronavirus vaccines of the present invention, the coronavirus is mixed with the appropriate adjuvant, diluents, and carriers. Physiologically acceptable media that can be used include, but are not limited to, appropriate isoosmotic solutions and phosphate buffers. Vaccines based on components of the coronavirus, such as those based on the earliest stage S protein sequence, as described in the preceding section, are particularly contemplated herein. The construction of a vector containing the nucleotide sequence encoding an earliest phase S protein (i.e., an S protein with residue 77 = D, residue 244 = T, and residue 778 = D; see, e.g., Table 1 above) is described elsewhere herein and would be known to one of ordinary skill in the art. See also Bukreyev *et al.*, *Lancet* 363:2122 (2004).

[099] The term "adjuvant" refers to a compound or mixture that enhances the immune response to an antigen. An adjuvant can serve as a tissue depot that slowly releases the antigen and also as a lymphoid system activator that non-specifically enhances the immune response (Hood *et al.*, *Immunology*, Second Ed., 1984, Benjamin/Cummings: Menlo Park, Calif., p. 384). Often, a primary challenge with an antigen alone, in the absence of an adjuvant, will fail to elicit a humoral or cellular immune response. Adjuvants include, but are

not limited to, complete Freund's adjuvant, incomplete Freund's adjuvant, saponin, mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil or hydrocarbon emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-  
5 Guerin) and *Corynebacterium parvum*.

[0100] Selection of an adjuvant depends on the subject to be vaccinated. Preferably, a pharmaceutically acceptable adjuvant is used. For example, a vaccine for a human should avoid oil or hydrocarbon emulsion adjuvants, including complete and incomplete Freund's adjuvant. One example of an adjuvant suitable for use with humans is alum (alumina gel). In  
10 a specific embodiment, *infra*, the vaccine of the present invention is administered intramuscularly in alum. Alternatively, the vaccine of the present invention can be administered subcutaneously, intradermally, intraperitoneally, or via other acceptable vaccine administration routes.

[0101] A vaccine formulation may be administered to a subject *per se* or in the form  
15 of a pharmaceutical or therapeutic composition. Pharmaceutical compositions comprising the adjuvant of the invention and an antigen may be manufactured by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Pharmaceutical compositions may be formulated in conventional manner using one or more physiologically acceptable carriers, diluents,  
20 excipients or auxiliaries which facilitate processing of the antigens of the invention into preparations which can be used pharmaceutically. Proper formulation is dependent upon the route of administration chosen. For purposes of this application, "physiologically acceptable carrier" encompasses carriers that are acceptable for human or animal use without relatively harmful side effects (relative to the condition being treated), as well as diluents, excipients or  
25 auxiliaries that are likewise acceptable. Systemic formulations include those designed for administration by injection, e.g. subcutaneous, intradermal, intramuscular or intraperitoneal injection. For injection, the vaccine preparations may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, phosphate buffered saline, or any other physiological saline buffer. The solution may contain  
30 formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the proteins may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

[0102] Determination of an effective amount of the vaccine formulation for administration is well within the capabilities of those skilled in the art. An effective dose can be estimated initially from in vitro assays. For example, a dose can be formulated in animal models to achieve an induction of an immune response using techniques that are well known in the art. One having ordinary skill in the art could readily optimize administration to all animal species based on results described herein. Dosage amount and interval may be adjusted individually. For example, when used as a vaccine, the vaccine formulations of the invention may be administered in about 1 to 3 doses for a 1-36 week period. Preferably, 1 or 2 doses are administered, at intervals of about 3 weeks to about 4 months, and booster vaccinations may be given periodically thereafter. Alternative protocols may be appropriate for individual animals. A suitable dose is an amount of the vaccine formulation that, when administered as described above, is capable of raising an immune response in an immunized animal sufficient to protect the animal from an infection for at least 4 to 12 months. In general, the amount of the antigen present in a dose ranges from about 1 pg to about 100 mg per kg of host, typically from about 10 pg to about 1 mg, and preferably from about 100 pg to about 1 pg. Suitable dose range will vary with the route of injection and the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

**Example 1: Molecular Evolution of the SARS Coronavirus During the Course of the 2002-2003 SARS Epidemic**

[0103] Severe acute respiratory syndrome (SARS) first emerged in Guangdong Province, China. Subsequently, the SARS coronavirus (SARS-CoV) was identified as the causative agent (Fouchier *et al.*, *Nature* 423:240 (2003); Ksiazek *et al.*, *N. Engl. J. Med.* 348:1953 (2003); Drosten *et al.*, *N. Engl. J. Med.* 348:1967 (2003); Rota *et al.*, *Science* 300:1394 (2003); Marra *et al.*, *Science* 300:1399 (2003)). It remains a challenge to establish the relationship between observed genomic variations and the biology of SARS (Rota *et al.*, *Science* 300:1394 (2003); Marra *et al.*, *Science* 300:1399 (2003); Vogel, *Science* 300:1062 (2003); Ruan *et al.*, *Lancet* 361:1779 (2003); Guan *et al.*, *Science* 302:276 (2003)). Recent molecular epidemiological studies have identified characteristic variant sequences in SARS-CoV for tracking disease transmission (Ruan *et al.*, *Lancet* 361:1779 (2003); Tsui *et al.*, *N. Engl. J. Med.* 349:187 (2003); Chim *et al.*, *Lancet* 362:1807 (2003); Chiu *et al.*, *N. Engl. J. Med.* 349:1875(2003)). Evidence suggests that SARS-CoV emerged from nonhuman sources

(Guan *et al.*, *Science* 302:276 (2003); Rest and Mindell, *Infect. Genet. Evol.* 3:219 (2003)). In this study, epidemiological and genetic evidence for viral adaptation to human beings was obtained through molecular investigations of the characteristic viral lineages found in China (Zhong *et al.*, *Lancet* 362:1353 (2003)).

5           [0104] On the basis of epidemiological investigations, the course of the 2002-2003 epidemic was divided into early ("HP03-E"), middle ("HP03-M"), and late ("HP03-L") phases. The early phase is defined as the period from the first emergence of SARS to the first documented superspreader event (SSE) (Zhong *et al.*, *Lancet* 362:1353 (2003)). The middle  
10       phase refers to the ensuing events up to the first cluster of SARS cases in hotel M in Hong Kong (Tsang *et al.*, *N. Engl. J. Med.* 348:1977 (2003)). Cases following this cluster fall into the late phase.

          [0105] The early phase was initially characterized by a series of seemingly independent cases. Eleven index cases that had arisen locally in the absence of any contact history were identified from different geographical locations within Guangdong Province.  
15       This phenomenon was observed from the retrospectively identified SARS index patient from the city of Foshan (onset date, 16 November 2002) (Zhong *et al.*, *Lancet* 362:1353 (2003)) through to an index patient from the city of Dongguan (onset date, 10 March 2003). All of these cases were confined to regions directly west of Guangzhou, the capital city of Guangdong Province, and to the city of Shenzhen in the south, with no cases being reported to  
20       the north or east of Guangzhou. This region, the Pearl River Delta, has enjoyed rapid economic development since the late 1970s, leading to the adoption of culinary habits requiring exotic animals. Seven of these 11 cases had documented contact with wild animals. In contrast to the apparently independent seeding of the earliest cases, the rest of the epidemic was characterized by SSEs and clusters of cases that were epidemiologically linked (Chim *et al.*,  
25       *Lancet* 362:1807 (2003); Chiu *et al.*, *N. Engl. J. Med.* 349:1875(2003); Zhong *et al.*, *Lancet* 362:1353 (2003); Tsang *et al.*, *N. Engl. J. Med.* 348:1977 (2003); Lee *et al.*, *N. Engl. J. Med.* 348:1986 (2003)).

          [0106] The first major SARS outbreak occurred in a hospital, HZS-2, in the city of Guangzhou, beginning on 31 January 2003 where an SSE was identified to be associated with  
30       more than 130 primary and secondary infections, of which 106 were hospital-acquired cases. Doctor A, a nephrologist who worked in this hospital, visited Hong Kong and stayed in Hotel M on 21 February 2003. Other visitors to the hotel later became infected with SARS-CoV

(Zhong *et al.*, *Lancet* 362:1353 (2003); Tsang *et al.*, *N. Engl. J. Med.* 348:1977 (2003)). This led to the transmission of SARS to Vietnam, Canada, Singapore, and the United States (Centers for Disease Control and Prevention, *Morb. Mortal. Wkly. Rep.* 52, 241 (2003)), with two further SSEs in Hong Kong, each resulting in the virus being transmitted to >100 contacts  
 5 (Chim *et al.*, *Lancet* 362:1807 (2003); Lee *et al.*, *N. Engl. J. Med.* 348:1986 (2003)).

[0107] Genomic sequence data for SARS-CoV were largely derived from isolates linked to the Hotel M cluster (Vogel, *Science* 300:1062 (2003)), hence they were predominantly from the late phase of the epidemic. 29 SARS-CoV genomic sequences were determined from 22 patients from Guangdong Province with disease onset dates in all three  
 10 phases of the epidemic, and from two patients from the late phase in Hong Kong. To eliminate mutational noise, it was assumed that sequence variants associated with common ancestry, but not arising in cell culture, should be seen in multiple isolates (Ruan *et al.*, *Lancet* 361:1779 (2003)). Meanwhile, critical genomic variations or complete genome sequences of certain virus isolates were verified by sequencing the reverse transcription  
 15 polymerase chain reaction (RT-PCR) products derived directly from patient specimens. The genomic sequences obtained were compared with 32 human SARS-CoV sequences and two SARS-like coronavirus sequences from Himalayan palm civets (*Paguma larvata*) available at GenBank (see the website [ncbi.nlm.nih.gov/entrez](http://ncbi.nlm.nih.gov/entrez)). Thus the following 63 sequences were compared: SZ3, SZ16, GZ02 (SEQ ID NO:1), GD01, HGZ8L1-A, HSZ-Cc, HSZ-A, HSZ-  
 20 Bb, HSZ-Cb, HSZ-Bc, GZ50, GZ-A, JMD, HGZ8L1-B, ZS-A, ZS-B, ZS-C, BJ04, BJ03, BJ02, BJ01, CUHK-W1, HZS2-D, HZS2-E, HZS2-C, HGZ8L2, HZS2-Bb, HZS2-A, HZS2-Fc, HZS2-Fb, TWC, Sin2679, ZJ01, HSR, TW1, HKU-39849, GZ-D, Urbani, Sin2748, Sin2677, Sin2500, Frankfurt, Sin2774, CUHK-Su10, CUHK-LC1, CUHK-AG01, CUHK-AG02, CUHK-AG03, TWH, TC1, TWY, TWS, TWK, TWJ, TC3, TC2, GZ-B, GZ-C,  
 25 TOR2, CUHK-LC2, CUHK-LC3, CUHK-LC4, and CUHK-LC5.

[0108] Only two major genotypes predominated during the early phase of the epidemic. Five isolates were found to contain a 29-nucleotide (nt) sequence that is absent in most of the publicly available SARS-CoV sequences, whereas another four isolates showed a previously unreported 82-nt deletion in the same region of the genome, Orf8 (Snijder *et al.*, *J. Mol. Biol.* 331, 991 (2003)) (see Figures 3 and 4). The former sequence is represented by the  
 30 GZ02 isolate (SEQ ID NO:1), and is used as the reference for annotation throughout this study. All of the isolates exhibiting this sequence (GZ02, HGZ8L1-A, HSZ-A, HSZ-B, and

HSZ-C; see Figure 5) were obtained from patients with contact histories traceable to some of the earliest independent cases in Guangzhou and were not detected in any of the later isolates. It is noteworthy that this sequence with the 29-nt segment is identical to the genomic sequence of coronaviruses isolated from animals in a Shenzhen live animal market (Guan *et al.*, *Science* 302:276 (2003)).

[0109] Three of the SARS-CoV genome sequences (ZS-A, ZS-B, and ZS-C; see Figure 5) with the 82-nt deletion were obtained from samples of very early cases from Zhongshan city. This 82-nt deletion was further confirmed by RT-PCR directly on an additional stool sample. A sequence with an identical 82-nt deletion has also been observed in coronaviruses isolated from farmed civets in Hubei Province, China. Specifically, SARS-like coronaviruses were isolated from palm civets farmed domestically in Hubei Province, China, by Hu *et al.* at the Wuhan Institute of Virology, Chinese Academy of Sciences. Partial genome sequencing revealed an 82-nt deletion within the Orf8 region, which is identical to that found in human SARS-CoV isolates from the early patients of Zhongshan, Guangdong Province, China. Contamination can be ruled out because no human SARS-CoV isolate with the 82-nt deletion has ever been found in that institute or has been isolated in that region of China. It is thus interesting to note that both sequences of the early phase were identified from other mammalian hosts. They provided a link to support the notion that early human infection of SARS-CoV may have originated from wild animals (Guan *et al.*, *Science* 302:276 (2003); Rest and Mindell, *Infect. Genet. Evol.* 3:219 (2003)).

[0110] In contrast to the early phase, a SARS-CoV sequence with the 29-nt deletion was observed during the middle phase that dominated the viral population for the rest of the epidemic (Rota *et al.*, *Science* 300:1394 (2003); Marra *et al.*, *Science* 300:1399 (2003); Ruan *et al.*, *Lancet* 361:1779 (2003)). Although this shift in genome size might be due to chance, deletion events appeared to be overrepresented in the Orf8 region. A fourth sequence with the 82-nt deletion was obtained from a Guangzhou patient (HGZ8L1-B), who was infected in the same ward as one of the patients where the longest sequence was obtained (HGZ8L1-A) (see above). Furthermore, a lung biopsy of a patient from the middle phase was found to contain two SARS-CoV genotypes, with the 29-nt and the 82-nt deletions, respectively. Remarkably, another genotype with a 415-nt deletion resulting in the loss of the whole Orf8 region was isolated and confirmed in two Hong Kong patients with disease onset from mid-May 2003 (see Figures 3 and 5). Specifically, the SARS-CoV sequence with the 415-nt deletion

(CUHK-LC2, CUHK-LC3, CUHK-LC4, and CUHK-LC5) was obtained from two SARS patients whose disease was linked to a late cluster of SARS cases in Hong Kong. Both patients had disease onset in mid-May 2003. The CUHK-LC2 sequence was initially obtained from the culture isolate of a throat wash specimen of an infected hospital health care worker and was later confirmed from the same specimen directly. CUHK-LC3, CUHK-LC4, and CUHK-LC5 were obtained from three different nasal swab specimens both directly and from the culture supernatants of an elderly patient who acquired SARS in the same hospital.

[0111] Because the majority of deletions observed in the SARS-CoV genome occurred in the Orf8 region with no apparent effect on the survival of the virus, it is tempting to suggest that this region is either noncoding or coding for a functionally unimportant putative protein. On the other hand, it is interesting to note that antiparallel reverse symmetrical sequences were readily predicted around the deletion sites (see Figure 3), which might account for the high deletion rates in this region. Whether such hairpin structures actually play a role in regulating either RNA replication or mRNA transcription in SARS-CoV is a subject for future studies.

[0112] Besides the deletion variants, 299 single-nucleotide variations (SNVs) were detected among the 63 sequences. Eighty-five of these variant loci were seen in more than one of the human SARS-CoV sequences. Among them, 52 were predicted to cause amino acid changes (nonsynonymous variations). When the epidemiologically determined transmission paths and SNV genotype data are combined, markers for genotypes characteristic of different lineages are evident (see Figure 5).

[0113] Viruses of the early phase have the characteristic motif of G:A:C:G:C at the GZ02 reference nucleotide residues 17,564, 21,721, 22,222, 23,823, and 27,827, with the bold SNVs matching the C:G:C:C motif identified previously (Ruan *et al.*, *Lancet* 361:1779 (2003)) (see Figure 5). This motif is shared by almost all early Guangzhou and Zhongshan isolates together with the animal SARS-like coronavirus isolates (SZ3 and SZ16) (Guan *et al.*, *Science* 276 (2003)). Along with the disappearance of viruses containing the 29-nt segment, the middle phase of the epidemic was characterized by the occurrence of genotypes with the G:A:C:T:C motif (see Figure 5). All of the middle-phase genotypes demonstrate this common motif but can be further classified into two variant groups on the basis of other SNVs. One group was represented by the isolates related to the Hospital HZS-2 outbreak (HZS2-A, HZS2-B, HZS2-C, HZS2-D, HZS2-E, and HGZ8L-2). The other group was

represented by the Hong Kong CUHK-W1 isolate that originated from Shenzhen (Tsui *et al.*, *N. Engl. J. Med.* 349:187 (2003)) together with the early Beijing isolates BJ01, BJ02, and BJ03, traceable to Guangdong. The transition between the characteristic motifs of the early and middle phases represented a G→T transversion at nucleotide residue 23,823 and is  
 5 predicted to cause an Asp →Tyr change at amino acid residue 778 of the spike (S) protein (see Figure 1).

[0114] An additional A→G transition at nucleotide 21,721 (see Figure 5) was identified in one isolate from a secondarily infected patient from Hospital HZS-2 with disease onset on 7 February 2003 (HZS2-Fc) (see Figure 5). This sequence was additionally  
 10 confirmed by direct sequencing of the RT-PCR product from an oropharyngeal swab of this patient (HZS2-Fb). This mutation is predicted to cause an Asp<sup>77</sup> →Gly amino acid switch in the S protein (fig. S4), and the G:G:C:T:C motif is so far genotypically the closest sequence to that of the Hotel M outbreak (T:G:T:T:T) (see Figure 5) (Tsang *et al.*, *N. Engl. J. Med.* 348:1977 (2003)). Epidemiologically, this patient is potentially linked to the Hotel M  
 15 outbreak through her contact with Doctor A during the first 3 days of illness. Thus, Doctor A was possibly infected with this viral variant.

[0115] Additionally, one G→T transversion and two C→T transitions at nucleotide residues 17,564, 22,222, and 27,827 are observed in the Hotel M-associated SARS-CoV genotypes (see Figure 5). These SNVs are predicted to cause amino acid switches in the  
 20 nonstructural polyprotein (Glu<sup>1389</sup> →Asp), the S protein (Thr<sup>244</sup> →Ile), and Orf8a (Arg<sup>17</sup> →Cys), respectively. This T:G:T:T:T motif is shared by the sequences of all isolates infected from and after the Hotel M cluster (Ruan *et al.*, *Lancet* 361:1779 (2003)), including the Hong Kong Amoy Gardens isolates (Chim *et al.*, *Lancet* 362:1807 (2003)) and the more recent isolates from Zhejiang (ZJ01), Taiwan (Chiu *et al.*, *N. Engl. J. Med.* 349:1875 (2003)), and  
 25 Guangdong (GZ-B, GZ-C, and GZ-D) (see Figure 5). This motif is also conserved in the late 415-nt deletion variant in Hong Kong with the exception of nucleotide 27,827, which falls within the deleted segment discussed previously. Thus, surprisingly few genotypes predominated during the late phase of the epidemic.

[0116] The characteristically high mutation rate of RNA viruses (Lai and Holmes, in  
 30 *Fields Virology*, Knipe and Howley, Eds. (Lippincott Williams & Wilkins, New York, ed. 4, 2001), chap. 35) may give rise to strains with increased virulence (Brown *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 98, 6883 (2001)) that can either escape host defenses (Seo *et al.*, *Nature*

*Med.* 8:950 (2002)) or change their tissue tropism (Rasschaert *et al.*, *J. Gen. Virol.* 71:2599 (1990)). In this regard, it was observed that the neutral mutation rate for SARS-CoV during this epidemic was almost constant, and was estimated to be  $8.26 \times 10^{-6}$  ( $\pm 2.16 \times 10^{-6}$ ) nt<sup>-1</sup> day<sup>-1</sup>. This is similar to the values obtained for known RNA viruses and is about one-third  
5 that for the human immunodeficiency virus (Li *et al.*, *Mol. Biol. Evol.* 5:313 (1988); Drake and Holland, *Proc. Natl. Acad. Sci. U.S.A.* 96:13910 (1999)). In contrast to the constant rate of synonymous variations, the nonsynonymous mutation rates were variable for the three epidemic phases. The predicted domains of the S protein responsible for viral host receptor recognition or internalization (Luo *et al.*, *J. Virol.* 73:8152 (1999)) were those that underwent  
10 the most extensive amino acid substitutions.

[0117] Between the coronavirus sequences of the palm civets (SZ3 or SZ16) and each of the human SARS-CoV sequences, the ratios of the rates of nonsynonymous to synonymous changes (Ka/Ks) for the S gene sequences were always greater than 1, indicating an overall positive selection pressure. However, pairwise analysis of the Ka/Ks for the genotypes in  
15 each epidemic group shows that the average Ka/Ks for the early phase was significantly larger than that for the middle phase, which in turn was significantly larger than the ratio for the late phase, which in fact was significantly less than 1. These data indicate that the S gene showed the strongest positive selection pressures initially, with subsequent purifying selections and eventual stabilization. For Orf1a, a pattern similar to that for the S gene was  
20 observed. In contrast, Orf1b (nt coordinate: 13,398 to 21,485) seems to be undergoing purifying selection during the whole course of the epidemic. Indeed, it is the most conserved genomic region of SARS-CoV (Ruan *et al.*, *Lancet* 361:1779 (2003)).

[0118] The present analysis thus suggests that adaptive pressures operated on the SARS-CoV genome but stabilized during the late phase of the epidemic with the emergence  
25 of a predominant genotype. Alternatively, sampling bias for cases related to SSEs (Bush *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 97:6974 (2000)) may distort the data. Such strategies may, however, be justifiable from a public health perspective, as the viral genotypes associated with the SSEs are the most epidemiologically important. To explore the possibility of bias, the date for the most recent common ancestor of the samples available was estimated. On the  
30 basis of the observed neutral mutation rate, this date was estimated to lie in mid-November 2002 (95% confidence interval: early June 2002 and late December 2002). This result is consistent with the onset date of 16 November 2002 for the earliest index patient from Foshan

(Zhong *et al.*, *Lancet* 362:1353 (2003)) and supports the finding that the early, middle, and late phase genotypes represent different stages of evolution of the same viral lineage. This is further evident from the remarkable correlation between the molecular clustering and epidemiological grouping of the genotypes throughout the epidemic (see Figure 5).

5 [0119] In tracing the molecular evolution of SARS-CoV in China, it was observed that the epidemic started and ended with deletion events, together with a progressive slowing of the nonsynonymous mutation rates and a common genotype that predominated during the latter part of the epidemic. The mechanistic explanation for the selective adaptation and purification processes that led to such genomic evolutionary changes in SARS-CoV requires  
10 further work (Ewald, *J. Urban Health* 75:480 (1998)). Nonetheless, this study has provided valuable clues to aid further investigation of this remarkable evolutionary tale.

[0120] In summary, then, in this Example sixty-one SARS coronavirus genomic sequences derived from the early, middle, and late phases of the severe acute respiratory syndrome (SARS) epidemic were analyzed together with two viral sequences from palm  
15 civets. Genotypes characteristic of each phase were discovered, and the earliest genotypes were similar to the animal SARS-like coronaviruses. Major deletions were observed in the Orf8 region of the genome, both at the start and the end of the epidemic. The neutral mutation rate of the viral genome was constant but the amino acid substitution rate of the coding sequences slowed during the course of the epidemic. The spike protein showed the strongest  
20 initial responses to positive selection pressures, followed by subsequent purifying selection and eventual stabilization.

## **Materials and Methods**

### **A. Epidemiological investigations**

25 [0121] Official epidemiological records of the Guangdong Center for Disease Control and Prevention (GDCDCP), which represented an aggregate of the regular SARS epidemiology reports submitted by the local Centers for Disease Control and Prevention of individual cities, were reviewed. The contact and clinical histories of all of the early  
seemingly independent index cases and several key cases (*e.g.* HZS2-F) were reconfirmed  
30 either by review of hospital patient records or direct interview with the patients and/or the

physicians-in-charge. In particular, eleven index cases from seven cities located in the Pearl River Delta region of Guangdong Province (Fig.1 and fig. S1), which occurred prior to the first superspreader event of a Guangzhou hospital, HZS-2, were investigated in detail

[0122] The majority of the specimens were collected by the virologists of GDCDCP, with the remaining samples collected by the staffs of local hospital or Guangzhou Center for Disease Control and Prevention.

## B. Sequencing strategy and procedures

[0123] Viral RNA templates were isolated either from the culture supernatants of VeroE6 cells that showed cytopathic effects or directly from patients' specimens of SARS cases (including serum, stool, oropharyngeal swabs, nasal pharyngeal aspirates or autopsy lung tissues). RNA was extracted with the QIAamp viral RNA mini kit (Qiagen, Valencia, CA, USA) or TRIZOL Reagent (GIBCOBRL). The double-strand cDNA was synthesized with the SuperScript II cDNA system (Invitrogen, Carlsbad, CA, USA) or RNA PCR Kit (AMV) Ver.2.1 (Takara, Dalian China). To amplify the genomic sequences of the SARS-CoV, 53 sets of nested primers were designed based on the TOR2 sequence. The nested PCR fragments were directly sequenced in both forward and reverse directions on the ABI-3700 DNA sequencer (Applied Biosystems, Foster City, CA, USA) with 2- to 4-fold redundancy. For GZ02, PCR primers were designed to cover the whole genome in every 1kb interval with 200bp overlap with the adjacent fragment based on the TOR2 sequence. PCR products were sequenced using ABI BigDye Terminator Cycle Sequencing Kit on ABI-377. All of the nucleotide sequence variations of GZ02, which differ from that of the human SARS-CoV sequences available at GenBank as of June 2003, in particular, TOR2 and GZ01 (the sequence of an independent viral isolate from the same patient as GZ02 and currently renamed as GD01) sequences (including the 29-bp segment), were re-sequenced from RNA extractions from the same lung tissue specimen of that patient and the 5' end sequence was completed. The PHRED/PHRAP/CONSED software (University of Washington, Seattle, WA, USA; available at the website phred.org) was used for base calling, assembly, and editing. The assembled genome sequence was checked manually for accuracy and the regions with poor quality were re-sequenced. For data analysis, the nucleotide coordinate of GZ02 (SEQ ID NO:1) was used as a reference.

### C. Sequence Alignments

[0124] Sequence alignments were generated using CLUSTALW 1.83 with the Gonnet nuclear acid comparison matrix for the sequences analyzed.

5

#### Example 2: Cross-host Evolution of SARS Coronavirus in Palm Civet and Humans

[0125] The coronaviruses isolated from a number of Himalayan palm civets (*Paguma larvata*) and a raccoon dog (*Nyctereutes procyonoides*) at a Shenzhen food market during the severe acute respiratory syndrome (SARS) epidemic of 2003 (April) were named SARS coronavirus-like coronaviruses ("SARS-CoV-like" coronaviruses) when it was observed that they displayed 99.8% sequence homology with the human SARS-CoV (Guan *et al.*, *Science* 302:276 (2003)). However, recent analyses of palm civet and other wild animals in the Guangzhou food market of late 2003 suggested that in these animals SARS-CoV-like coronaviral loads and sequence variations were greater than those observed previously in these animals.

[0126] The molecular investigation of the characteristic viral lineages of the 2002-2003 Chinese SARS epidemic discussed in Example 1 above provides epidemiological and genetic evidence for viral adaptation to human beings. Although an animal origin of the infection has been suggested, neither direct evidence nor clues about the molecular mechanisms that enable the virus to switch hosts have been available. In this Example, the sequence data of viruses obtained from recent (2003-2004) human and palm civet infections were used to delineate the characteristics of the cross-host evolution of the SARS-CoV over a short period of time. This is an essential step for understanding the genetic process of adaptation of the SARS coronavirus to humans, and is critically important to developing procedures for treating and ultimately preventing SARS.

[0127] Between December 16, 2003 and January 8, 2004, a total of 4 patients were hospitalized in the city of Guangzhou, Guangdong Province, China, with flu-like syndromes later diagnosed as confirmed SARS. No patient had contact with other SARS cases nor had contact with each other. However, all of these patients had direct contact history with wild

animals (palm civets and house rat) in geographically restricted areas. They all had very mild symptoms, much less severe than those displayed by most patients during the previous epidemic. None of their close contacts were infected. In summary, these four cases occurred independently and seemed to have little infectivity towards other human beings.

5 [0128] Specimens were collected during the 2003-2004 outbreak in Guangzhou, with nearly complete SARS-CoV viral genome sequences from the first and the second of the four human patients. Related SARS-CoV-like coronaviruses were similarly obtained from two palm civets of Guangzhou food market and one sample from an animal cage at a restaurant in the area (Restaurant TDL).

10 [0129] The viral sequences of the 2003-2004 outbreak were compared with those identified from the 2002-2003 epidemic, as shown in Figure 6. All of them were characterized as bearing the 29 bp segment marker in orf8a as in the viruses of PC03 and the Guangzhou lineage of HP03E, *e.g.*, GZ02 (See Example 1: see also Chinese SARS Molecular Epidemiology Consortium, *Science* 303:1666 (2004)). A total of 202 single-nucleotide  
15 variations (SNVs) with multiple occurrences were identified. Beside 2 non-coding variations and 72 synonymous variations, 89 of the remaining 128 nonsynonymous mutations may cause drastic amino acid changes in the viral proteins.

[0130] The phylogenetic analysis of these viral sequences demonstrated that the genomes of the SCoV from HP04 were almost identical to those of the SARS-CoV-like  
20 coronaviruses from PC04 (see Figure 6). There were 33 SNVs detected among the viruses from PC04 and HP04, which accounts for only 0.11% of the viral genome. In contrast, a total of 77 SNVs was detected among the SCoV genomic sequences of HP03E and PC03, accounting for nearly 0.26% of the the viral genome. Remarkably, no SNV distinguishes the genotype of SCoV from that of the SCoV-like viruses, although 17 out of the 202 SNVs were  
25 only observed in animals. Thus, structurally, there is little difference to distinguish these two viruses and functionally, concerning the direct animal contact history of the current patients, it is likely that the same virus can infect both palm civet and human. SARS is indeed a zoonotic disease.

[0131] Comparing the genomic sequence variations between PC03 and PC04, the  
30 difference was significant. There were 82 SNVs detected among the viruses from palm civets, which accounts for nearly 0.28% of the viral genome. This variation ratio is even

higher than that observed between PC03 and HP03-E (see above). To explore further this remarkable observation, the phylogeny of the three most significantly variable protein coding sequences (CDSs), Spike (S), sars3a and nsp3 among palm civets and human patients of the two epidemic was analyzed using the maximum likelihood estimation (Yang, *Mol. Biol. Evol.* 15:568(1998)) (see Figures 7A-C).

[0132] As shown in Fig 7A, the S gene sequence has evolved particularly fast, under positive Darwinian selection. First, from the ancestor node of PC03 to the node of HP03-E, the nonsynonymous/synonymous (A/S) ratio is 11.8/0, which corresponds to  $\omega=\infty$  ( $\omega$ : ratio of nonsynonymous and synonymous rates). This confirmed the previous conclusion of Example 1 (see also Chinese SARS Molecular Epidemiology Consortium, *Science* 303:1666 (2004)) that during the virus transmission from palm civet to human, the S gene experienced strong positive selection and improvement in order to adapt to its human host. Within the HP03-E, in most branches, a very high A/S ratio was observed, again suggesting that the S gene was still evolving, having not yet reached its maximum adaptation to human.

[0133] Second, from the ancestor node of PC03 to the node of PC04, the A/S ratio is 18.2/2.1 ( $\omega=2.68$ ). This demonstrates that the S gene is also under positive selection pressure during animal to animal transmission. In this regard, it is likely that the SARS-CoV of the current epidemic has evolved to a more virulent form in palm civets, *i.e.*, better adapted to its host. This is supported by the fact that it was much easier to obtain SARS-CoV samples for whole genome sequencing from Guangzhou food market during 2003-2004 period than during the 2002-2003 epidemic (data not show). Interestingly, the ancestor nodes of PC04 and HP04 are the same, indicating that unlike during the 2002-2003 epidemic, these viruses did not have a chance to diverge for enough time, although in the patient GZ03-02, they already accumulated some amino acid changes (A/S=6/1). Because PC03 and PC04 do not belong to the same lineage, these comparisons also implied that the transmission of the virus from animal to human did occur independently.

[0134] The significant difference for the level of genomic variation between PC03/HP03-E and PC04/HP04 should be stressed along with the difference of the infectivity of the human SARS cases. In sharp contrast to the 2002-2003 epidemic, no further infection was reported for any of the HP04 cases, while the earliest retrospectively identified SARS patients of the 2002-2003 epidemic infected 2 close contacts directly and 3 others secondarily (see Example 1; see also Chinese SARS Molecular Epidemiology Consortium, *Science*

303:1666 (2004)). Therefore, the so called “early phase” of the 2002-2003 SARS, i.e., HP03-E, should be more accurately described as a “relatively late stage of the early phase,” as they have already shown fairly severe symptoms and significant infectivity towards human contacts.

5           [0135] Although the HP03E and HP04 were not from the same lineage of SCoV, the A/S ratio between their ancestor nodes is 9.4/2.1 ( $\omega=1.39$ ). This high ratio further confirmed the idea of positive selection in the early human infection period and implied that HP04, to certain extent, represented the “true” early phase of the virus, i.e., the “very early” or “early-early” phase. In parallel, the A/S or  $\omega$  ratio decreased during the course of the epidemic,  
10           which suggests that purifying selection began to dominate when the virus had adapted to the host, as discussed in Example 1 above.

          [0136] It has been known that the sars3a CDS was expressed and translated into a membrane associated viral structure protein (Zeng *et al.*, *J. Mol. Biol.* in press), while it evolves rapidly, in synergy with the S protein. Therefore, it is not surprising that its  
15           evolution showed the same trifurcating tree for the four epidemic groups as that for the S gene, as is shown in Figure 7B. Combining the three lineages, connecting the ancestor nodes of the four groups, the A/S ratio is 7/0 (there is no synonymous changes) between group HP03-E and HP04. In contrast, there is no single change among palm civets and human beings of the current epidemic. Therefore, it seems that amino acid changes in sars3a are  
20           critical to enable human to human (or even animal to animal) transmission and may result in increasing infectiousness during the early period of human to human transmission.

          [0137] The phylogenetic tree of nsp3 (see Figure 7C) is largely different from that of Spike or sars3a. The PC03 is very close to HP03-E but relatively more diverging from those of new cases. This suggests that this gene may be under different evolutionary forces from  
25           that for Spike and sars3a genes. In the lineage connecting the ancestor node of HP03-E and HP04 (or PC04), the A/S ratio is only 4.1/6.2 ( $\omega=0.227$ ), which does not show any positive selection signature. It is worth pointing out that in new cases, there is one mutation at nucleotide 6295 leading to a stop codon in the nsp3 CDS of the orf1a, which might account for the mild clinical symptoms and the apparent weak infectivity of this episode.

30           [0138] The Spike protein is responsible for receptor binding and thus, is the fastest evolving protein of SARS-CoV over the epidemic from animal to human. Out of the 17

SNVs observed in animals, 10 are located in the S gene. Among them, 7 were observed in the current epidemic, 1 was observed in the previous epidemic, and 2 were observed in both. With 7 more S gene sequences from samples of the third patient of the current epidemic and of 6 palm civets from Guangzhou food market added for analysis, no further changes were found in the SNV patterns. The three dimensional structure of the S protein simulated by Niccolai *et al.* (Bernini *et al.*, submitted) was used to better understand the molecular mechanism driving the mutations of the S gene over the course of the epidemic. Although mutations are dispersed over the whole protein, *i.e.*, the S1 and S2 domains, the exposed top or the buried interior, *etc.*, the majority of the mutations are located in the S1 domain (31 out of 48 total SNVs), particularly in the region predicted to constitute the ACE2 receptor binding site, 11 SNVs corresponding to 10 amino acids. Among them, except for 2 synonymous variations, 7 out of the 9 non-synonymous mutations may cause drastic amino acid changes. Two of these non-synonymous mutations, at positions 22422 and 22549, occurred during the 2002-2003 epidemic and represented evolutionary steps in human hosts, while the remaining 5 fell into 3 categories. First, mutations at the second and third nucleotides (22927 and 22928) of codon 479 may cause changes corresponding to 3 different amino acid residues (K, R, or N). Although all of these codons were found in the palm civet samples, only the *aat* codon for N was found in all the human samples as well as some 2003-2004 palm civet samples. Second, the *c*→*t* switch of nt 22570 causing the S→F mutation of codon 360 distinguishes the virus of 2002-2003 epidemic (HP03) from all the other viruses isolated from palm civet (PC03 and PC04) as well as human patients of the 2003-2004 outbreak (HP04). Third, the *g*→*a* switch of nt 22930 causing the G→D mutation of codon 480 distinguishes the virus of 2002-2003 (PC03 and HP03) from those of 2003-2004 (PC04 and HP04), regardless of sources.

[0139] Although concerted mutation events are extremely rare, the second case was observed for nts 23316 and 23317 encoding amino acid residue 609, which is predicted to be buried at the interface of S proteins. This *tta*→*gca* switch causing an L→A mutation is one of a few nonsynonymous mutations that nearly distinguishes the virus of 2002-2003 from those of 2003-2004, disregarding either their human or animal sources. This event is the more remarkable because it also goes in the direction of G+C enrichment, a feature that is usually extremely rare in viruses, for metabolic reasons (Rocha and Danchin, *Trends Genet.* 18:291 (2002)). This indicates that this change, which may modify the relative orientation of the S1

and S2 domains, plays a major role either in the stability of the protein, or in its ability to interact with its target receptor.

[0140] In summary, the unfortunate recurrence of SARS at the end of year 2003 provided an opportunity to witness the variation/adaptation behaviour of the etiological agent of the disease. The new SARS-CoV did not derive from the preceding episode, but very likely from a common ancestor, which does not harbor the deletion that marks most of the virulent forms of SCoV for the 2002-2003 epidemic. The fates of the virus inside the human host and in palm civets are similar, i.e., the virus is not yet adapted to its new hosts, making it fast-evolving (possibly into more highly contagious and/or virulent forms); and, in general, the infection is mild. Therefore, humans working with wild animals are often seropositive for the SARS-CoV without noticeable severe symptoms (see Guan *et al.*, *Science* 302:276 (2003)). These data point to a common source of disease lingering in the environment that is presumably both adapted to its natural host and able to come into contact with humans and/or animals. It may have a fairly high probability of mutation under favorable conditions to a form causing SARS in humans. This situation is expected to yield an unusual epidemic pattern, since a proportion of humans may have been immunized against an innocuous form of the virus, so that distribution of the disease, when it happens, is expected to be highly uneven. These data thus strongly suggest the need for further research on the discovery of coronaviruses in animals, in particular in the Guangdong region.

### Example 3: SNV Diagnostic Kit

[0141] Coronaviruses are isolated and cultured from patient samples, including samples obtained from patient phlegm, excrement or tissues. Coronaviral RNA is prepared from these cultures, and cDNA is obtained using reverse transcription. Alternatively, cDNA may be directed obtained from patient samples by reverse transcription without intermediate culturing. This procedure will be particularly important for SARS-CoV or SARS-CoV-like coronaviral strains which are difficult to culture (e.g., which propagate poorly on VeroE6 cells).

[0142] PCR reactions are then performed on the cDNA obtained as above, with primers chosen depending upon the insertion(s)/deletion(s) or SNV(s) to be analyzed. For SNVs at positions 22222, 22570, 22927-22928, 22930, 23316-23317, and 23823, for

example, one set of suitable primers is the sense primer  
GCACCCACCTGCTCTTAATTGTTATTGGC (SEQ ID NO:6) and the anti-sense primer  
TATTAAAGAGCAAGTCCTCAATAAAAGACC (SEQ ID NO:7). The selection of such  
primers is based on standard considerations used for PCR amplifications, as would be well-  
5 known to one of ordinary skill in the art of such amplifications.

[0143] In order to conduct the PCR reaction, primers are diluted to 1  $\mu\text{mol}/\mu\text{l}$  in a  
solution containing the template cDNA. Amplified PCR fragments are purified and analyzed  
by sequencing or gel electrophoresis as appropriate.

10 [0144] While the present invention has been described with reference to its preferred  
embodiments, one of ordinary skill in the relevant art will understand that the present  
invention is not intended to be limited by these preferred embodiments, and is instead  
contemplated to include all embodiments consistent with the spirit and scope of the present  
invention as defined by the appended claims.

# **DEMANDES OU BREVETS VOLUMINEUX**

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS  
COMPREND PLUS D'UN TOME.**

**CECI EST LE TOME \_\_1\_\_ DE \_\_2\_\_**

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

---

# **JUMBO APPLICATIONS / PATENTS**

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE  
THAN ONE VOLUME.**

**THIS IS VOLUME \_\_1\_\_ OF \_\_2\_\_**

NOTE: For additional volumes please contact the Canadian Patent Office.

## WHAT IS CLAIMED IS:

1. A method for detecting an early-stage SARS coronavirus comprising determining the presence of at least one early-stage-specific molecular marker in a viral RNA,  
5 wherein said early-stage-specific molecular marker is selected from the group consisting of: an insertion early-stage-specific molecular marker, a deletion early-stage-specific molecular marker, and at least one single nucleotide variation (SNV) early-stage-specific molecular marker.
- 10 2. The method of claim 1, wherein at least one SNV early-stage-specific molecular marker is selected from the group consisting of: C at position 4220, G at position 17564, N at position 17564, A at position 21721, N at position 21721, C at position 22222, T at position 22222, N at position 22222, G at position 23823, C at position 23823, A at position 23823, C at position 27827, and N at position 27827.
- 15 3. The method of Claim 1, wherein the SNV early-stage-specific molecular marker is selected from the consisting of:
  - a. a G at position 17564, an A at position 21721, a C at position 22222, a G at position 23823, and a C at position 27827;
  - 20 b. a G at position 17564, an A at position 21721, a T at position 22222, a G at position 23823, and a C at position 27827;
  - c. an N at position 17564, an N at position 21721, an N at position 22222, a G at position 23823, and an N at position 27827; and,
  - d. an N at position 17564, an N at position 21721, an N at position 22222, a G,  
25 C, or A at position 23823, and an N at position 27827.
- 30 4. The method of Claim 1, wherein the SNV early-stage-specific molecular marker is selected from the group consisting of: a T at position 508, a C at position 1206, a T at position 1221, a T at position 1909, a T at position 2759, a C at position 3326, a T at position 3570, a C at position 3626, a T at position 3962, an A at position 5251, a T at position 5963, a T at position 6612, an A at position 2929, a C at position 7070, a T at position 7842, an A at position 8031, a G at position 8502, a C at position 8559, a T at position 8815, an A at position 8946, a T at position 9095, a C at position 9176, a T at

position 9404, a G at position 9448, a C at position 9479, an A at position 10029, a C at position 10587, an A at position 12725, a T at position 13862, a G at position 16541, a C at position 17131, a T at position 17723, a G at position 19838, a T at position 19882, an A at position 20840, an A at position 20992, a T at position 21479, a T at position 21637, a G at position 21715, a C at position 22145, a T at position 22207, an A at position 22422, a G at position 22517, a G at position 22522, a C at position 24566, a G at position 24978, a T at position 25230, an A at position 25521, an A at position 25566, a C at position 25779, a C at position 25808, a T at position 25844, an A at position 26032, a C at position 26586, a T at position 27243, a T at position 28089, a T at position 28193, and a C at position 29725.

5. The method of claim 1, wherein the early-stage-specific molecular marker comprises an insertion of SEQ ID NO:2 at position 27883.
6. The method of claim 1, wherein the early-stage-specific molecular marker comprises deletion SEQ ID NO:3 between positions 27858-27883 and 27912-27939.
7. A vaccine that generates an immune response to a SARS-CoV or SARS-CoV-like coronavirus in a mammalian host, comprising a therapeutically effective amount of early stage SARS-CoV virus having at least 95% similarity to the nucleotide sequence of SEQ ID NO:1 in a pharmacologically acceptable carrier.
8. The vaccine of Claim 1, wherein the early stage SARS Co-V whole virus further comprises at least one single nucleotide variation from SEQ ID NO:1 selected from the group of: C at position 4220, G at position 17564, N at position 17564, A at position 21721, N at position 21721, C at position 22222, T at position 22222, N at position 22222, G at position 23823, C at position 23823, A at position 23823, C at position 27827, and N at position 27827.
9. The vaccine according to Claim 7, wherein the SARS CoV virus comprises at least one single nucleotide variation selected from the group consisting of: a T at position 508, a C at position 1206, a T at position 1221, a T at position 1909, a T at position 2759, a C at position 3326, a T at position 3570, a C at position 3626, a T at position 3962, an A at position 5251, a T at position 5963, a T at position 6612, an A at

position 2929, a C at position 7070, a T at position 7842, an A at position 8031, a G at position 8502, a C at position 8559, a T at position 8815, an A at position 8946, a T at position 9095, a C at position 9176, a T at position 9404, a G at position 9448, a C at position 9479, an A at position 10029, a C at position 10587, an A at position 12725, a T at position 13862, a G at position 16541, a C at position 17131, a T at position 17723, a G at position 19838, a T at position 19882, an A at position 20840, an A at position 20992, a T at position 21479, a T at position 21637, a G at position 21715, a C at position 22145, a T at position 22207, an A at position 22422, a G at position 22517, a G at position 22522, a C at position 24566, a G at position 24978, a T at position 25230, an A at position 25521, an A at position 25566, a C at position 25779, a C at position 25808, a T at position 25844, an A at position 26032, a C at position 26586, a T at position 27243, a T at position 28089, a T at position 28193, and a C at position 29725.

10. The vaccine of Claim 8, wherein the at least one single nucleotide variation from SEQ ID NO:1 is further selected from the group consisting of:

a) a G at position 17564, an A at position 21721, a C at position 22222, a G at position 23823, and a C at position 27827;

b) a G at position 17564, an A at position 21721, a T at position 22222, a G at position 23823, and a C at position 27827;

c) an N at position 17564, an N at position 21721, an N at position 22222, a G at position 23823, and an N at position 27827; and,

d) an N at position 17564, an N at position 21721, an N at position 22222, a G, C, or A at position 23823, and an N at position 27827.

11. The vaccine of claim 7, wherein the early stage SARS-CoV virus is inactivated.

12. The vaccine of claim 7, wherein the early stage SARS-CoV virus is attenuated.

13. A polynucleotide comprising an early stage SARS-CoV or SARS-CoV-like viral nucleotide sequence having at least 95% similarity to the nucleotide sequence of SEQ ID NO: 1.

14. The polynucleotide of Claim 13, further comprising at least one single nucleotide variation in SEQ ID NO: 1 selected from the group consisting of: C at position 4220, G at position 17564, N at position 17564, A at position 21721, N at position 21721, C at position 22222, T at position 22222, N at position 22222, G at position 23823, C at position 23823, A at position 23823, C at position 27827, and N at position 27827.

15. The polynucleotide of Claim 14, wherein the at least one single nucleotide variation in SEQ ID NO: 1 is further selected from the group consisting of:

- a. a G at position 17564, an A at position 21721, a C at position 22222, a G at position 23823, and a C at position 27827;
- b. a G at position 17564, an A at position 21721, a T at position 22222, a G at position 23823, and a C at position 27827;
- c. an N at position 17564, an N at position 21721, an N at position 22222, a G at position 23823, and an N at position 27827; and,
- d. an N at position 17564, an N at position 21721, an N at position 22222, a G, C, or A at position 23823, and an N at position 27827.

16. The polynucleotide according to Claim 13, wherein the SARS CoV virus includes at least one single nucleotide variation selected from the group consisting of: a T at position 508, a C at position 1206, a T at position 1221, a T at position 1909, a T at position 2759, a C at position 3326, a T at position 3570, a C at position 3626, a T at position 3962, an A at position 5251, a T at position 5963, a T at position 6612, an A at position 2929, a C at position 7070, a T at position 7842, an A at position 8031, a G at position 8502, a C at position 8559, a T at position 8815, an A at position 8946, a T at position 9095, a C at position 9176, a T at position 9404, a G at position 9448, a C at position 9479, an A at position 10029, a C at position 10587, an A at position 12725, a T at position 13862, a G at position 16541, a C at position 17131, a T at position 17723, a G at position 19838, a T at position 19882, an A at position 20840, an A at position 20992, a T at position 21479, a T at position 21637, a G at position 21715, a C at position 22145, a T at position 22207, an A at position 22422, a G at position 22517, a G at position 22522, a C at position 24566, a G at position 24978, a

T at position 25230, an A at position 25521, an A at position 25566, a C at position 25779, a C at position 25808, a T at position 25844, an A at position 26032, a C at position 26586, a T at position 27243, a T at position 28089, a T at position 28193, and a C at position 29725.

5

17. An amino acid sequence encoded by the polynucleotide of Claim 13.

18. An amino acid sequence encoded by the polynucleotide of Claim 14.

10

19. An amino acid sequence encoded by the polynucleotide of Claim 15.

20. A polypeptide comprising the amino acid sequence of Claim 17.

21. A polypeptide comprising the amino acid sequence of Claim 18.

15

22. A polypeptide comprising the amino acid sequence of Claim 19.

23. An immunodiagnostic for detecting a SARS-CoV or SARS-CoV-like coronavirus comprising at least two of the polypeptides of Claim 20 of at least two SARS CoV or SARS-CoV-like coronaviral serotypes.

20

24. An immunodiagnostic for detecting a SARS-CoV or SARS-CoV-like coronavirus comprising at least two of the polypeptides of Claim 21 of at least two SARS CoV or SARS-CoV-like coronaviral serotypes.

25

25. An immunodiagnostic for detecting SARS-CoV or SARS-CoV-like coronavirus comprising at least two of the polypeptides of Claim 22 of at least two SARS CoV or SARS-CoV-like coronaviral serotypes.

30

26. An immunodiagnostic kit for detecting SARS-CoV or SARS-CoV-like coronavirus in a test subject comprising:

- a. a polypeptide selected from the group consisting of : the polypeptide of Claim 20, the polypeptide of Claim 21 and the polypeptide of Claim 22; and

- b. antibodies directed against an epitope in early stage SARS-CoV or SARS-CoV-like coronavirus.

27. The immunodiagnostic kit according to Claim 26, wherein the epitope further  
5 comprises single nucleotide variations of SEQ ID NO:1 present in at least one of  
nucleotide positions 17564, 21721, 2222, 23823 and 27827.

28. The immunodiagnostic kit according to Claim 19, wherein the epitope further  
10 comprises single nucleotide variations in the Spike protein of SEQ ID NO: 4 present  
in at least one of nucleotide positions 21721, 22222, and 23823.

29. The immunodiagnostic kit according to Claim 28, wherein the epitope further  
comprises an amino acid mutation from D to Y at position 778 of the Spike protein of  
SEQ ID NO: 5.

15 30. A vector encoding an amino acid sequence comprising a polypeptide selected from  
the group consisting of: the polypeptide of Claim 20, the polypeptide of Claim 21 and  
the polypeptide of Claim 22.

20 31. A cell comprising the vector of Claim 30.

32. An animal comprising at least one of the vector of Claim 20 and the cell of Claim 31.

25 33. A method for detecting an early-stage SARS coronavirus comprising determining the  
presence of at least one early-stage-specific amino acid in SARS coronaviral S  
protein, wherein said early-stage-specific amino acid is selected from the group  
consisting of: S at position 360, R or K at position 479, G at position 480, and L at  
position 609 of the SARS coronaviral S protein.

FIG 1A

|      |       | Orf1a |    |    |    |    |    |    |    |    |    |       |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|------|-------|-------|----|----|----|----|----|----|----|----|----|-------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
|      | Phase | nspl2 |    |    |    |    |    |    |    |    |    | nspl3 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Year |       | nspl1 | gt | ct | ct | tc | ag | tc | gc | ct | gc | ca    | ga | ct | gt | ag | ct | tc | at | ag | ct | gc | ac | ga | tg | ag | ac | tc | ct | ta | ag | tc | ct | ag | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc | tc | ct | gc |

[illegible]

F16 IC

| SNV      |  | Year |  | Phase |  | Codon |  | AA switch |  | AA residue # |  | nt coordinate |  |
|----------|--|------|--|-------|--|-------|--|-----------|--|--------------|--|---------------|--|
| GZ03-02  |  | HP04 |  | C     |  | C     |  | C         |  | C            |  | G             |  |
| GZ03-01  |  | HP04 |  | C     |  | C     |  | C         |  | C            |  | G             |  |
| gzl227   |  | PC04 |  | C     |  | C     |  | C         |  | C            |  | G             |  |
| gzl136   |  | PC04 |  | C     |  | C     |  | C         |  | C            |  | G             |  |
| GZL13    |  | PC04 |  | C     |  | C     |  | C         |  | C            |  | G             |  |
| SZ16     |  | PC03 |  | C     |  | C     |  | C         |  | C            |  | G             |  |
| SZ3      |  | PC03 |  | C     |  | C     |  | C         |  | C            |  | G             |  |
| HGZ8L1-1 |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| GZ02     |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| GD01     |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| ZS-C     |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| ZS-B     |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| ZS-A     |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| GZ50     |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| JMD      |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| HSZ-A    |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| HSZ-Bb   |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| HSZ-Cc   |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| HSZ-Cb   |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| HSZ-Bc   |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| GZ-A     |  | HP03 |  | E     |  | C     |  | G         |  | C            |  | C             |  |
| HZS2-Bb  |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| HSZ2-A   |  | HP03 |  | M     |  | N     |  | G         |  | C            |  | C             |  |
| HGZ8L1-1 |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| CUHK-W   |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| HZS2-D   |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| HZS2-E   |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| HZS2-C   |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| BJ03     |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| BJ02     |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| BJ01     |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| HGZ8L2   |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| BJ04     |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| NS1      |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| HZS2-Fb  |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| HZS2-Fc  |  | HP03 |  | M     |  | C     |  | G         |  | C            |  | C             |  |
| QXC1     |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| GZ-D     |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| GZ-C     |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| GZ-B     |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| TOR2     |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| URBANI   |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| TC2      |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| TWY      |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| TWS      |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| TWK      |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| TWJ      |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| TC3      |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |
| TC1      |  | HP03 |  | L     |  | C     |  | G         |  | C            |  | C             |  |

FIG 1D

|               |      |       |       |       |       |       |       |       |       | Sars3a |       |       |       |       |       |       |       |       |       | Sars3b |       |       |       |       |       |       |       |       |       |   |
|---------------|------|-------|-------|-------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---|
|               |      |       |       |       |       |       |       |       |       |        |       |       |       |       |       |       |       |       |       |        |       |       |       |       |       |       |       |       |       |   |
| SNV           | Year | Phase | ag    | cg    | ct    | tg    | tc    | gc    | gt    | gc     | gc    | gc    | gc    | gc    | gc    | gc    | gc    | gc    | gc    | gc     | gc    | gc    | gc    | gc    | gc    | gc    | gc    | gc    | gc    |   |
| Codon         |      |       | gac   | acc   | acc   | gac   | gac   | gac   | gac   | gac    | gac   | gac   | gac   | gac   | gac   | gac   | gac   | gac   | gac   | gac    | gac   | gac   | gac   | gac   | gac   | gac   | gac   | gac   | gac   |   |
| AA switch     |      |       | D-G   | T-S   | S-S   | S-A   | V-V   | S-P   | A-S   | A-V    | D-E   | S-S   | T-A   | C-C   | D-D   | E-K   | R-K   | H-H   | C-S   | W-Q    | K-Q   | W-R   | A-A   | P-P   | S-S   |       |       |       |       |   |
| AA residue #  |      |       | 480   | 487   | 561   | 577   | 596   | 607   | 609   | 609    | 613   | 665   | 701   | 743   | 743   | 748   | 754   | 765   | 778   | 894    | 964   | 1001  | 1025  | 1148  | 1163  | 1180  | 1208  | 1247  | 1259  |   |
| nt coordinate |      |       | 22930 | 22951 | 23174 | 23220 | 23279 | 23310 | 23316 | 23317  | 23330 | 23485 | 23593 | 23718 | 23719 | 23735 | 23752 | 23785 | 23823 | 24171  | 24383 | 24493 | 24566 | 24933 | 24978 | 25031 | 25114 | 25230 | 25286 |   |
| GZ03-02       | HP04 | G     | G     | C     | T     | T     | T     | T     | G     | C      | T     | C     | C     | A     | G     | A     | C     | T     | G     | A      | G     | C     | C     | C     | C     | C     | C     | C     | C     | G |
| GZ03-01       | HP04 | G     | G     | C     | T     | T     | T     | T     | T     | A      | C     | C     | A     | G     | A     | C     | T     | G     | A     | G      | C     | C     | C     | C     | C     | C     | C     | C     | C     | G |
| gzl227        | PC04 | G     | G     | C     | T     | T     | T     | T     | T     | A      | C     | C     | A     | G     | A     | C     | T     | G     | A     | G      | C     | C     | C     | C     | C     | C     | C     | C     | C     | G |
| gzl136        | PC04 | G     | G     | C     | T     | T     | T     | T     | T     | A      | C     | C     | A     | G     | A     | C     | T     | G     | A     | G      | C     | C     | C     | C     | C     | C     | C     | C     | C     | G |
| GZL13         | PC04 | G     | G     | C     | T     | T     | T     | T     | T     | A      | C     | C     | A     | G     | A     | C     | T     | G     | A     | G      | C     | C     | C     | C     | C     | C     | C     | C     | C     | G |
| SZ16          | PC03 | A     | G     | C     | T     | T     | C     | G     | C     | T      | C     | T     | G     | C     | A     | T     | C     | G     | G     | T      | G     | C     | C     | C     | C     | C     | C     | C     | C     | G |
| SZ3           | PC03 | A     | G     | C     | T     | T     | C     | G     | C     | T      | C     | T     | G     | C     | A     | T     | C     | G     | G     | G      | C     | C     | C     | C     | C     | C     | C     | C     | C     | G |
| HGZ8L1-A      | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| GZ02          | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| GD01          | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| ZS-C          | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| ZS-B          | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| ZS-A          | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| GZ50          | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| JMD           | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HSZ-A         | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | N     | N     | N     | N     | N     | N     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HSZ-Bb        | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HSZ-Cc        | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HSZ-Cb        | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HSZ-Bc        | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| GZ-A          | HP03 | E     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HZS2-Bb       | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HSZ2-A        | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HGZ8L1-B      | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| CUHK-W1       | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HZS2-D        | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HZS2-E        | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HZS2-C        | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| BJ03          | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| BJ02          | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| BJ01          | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HGZ8L2        | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| BJ04          | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| NS1           | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HZS2-Fb       | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| HZS2-Fc       | HP03 | M     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| QXC1          | HP03 | L     | A     | C     | C     | G     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| GZ-D          | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| GZ-C          | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| GZ-B          | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| TOR2          | HP03 | L     | A     | C     | C     | G     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| URBANI        | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| TC2           | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| TWY           | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| TWS           | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| TWK           | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| TWJ           | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| TC3           | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |
| TC1           | HP03 | L     | A     | C     | C     | T     | T     | T     | G     | C      | T     | T     | C     | A     | C     | A     | C     | C     | G     | A      | G     | C     | C     | A     | T     | C     | C     | C     | C     | G |

5/24

FIG 1E

|          |       | E         |              | M             |       | Sars6 |       | Sars7a |       | Sars8a |       | Sars8b |       | Sars9b |       | Z     |       |
|----------|-------|-----------|--------------|---------------|-------|-------|-------|--------|-------|--------|-------|--------|-------|--------|-------|-------|-------|
|          |       | Year      |              | Phase         |       |       |       |        |       |        |       |        |       | sars9b |       |       |       |
| SNV      | Codon | AA switch | AA residue # | nt coordinate | ct    | at    | ga    | ag     | ct    | ag     | ct    | ca     | ct    | ct     | ct    | ag    | gc    |
|          |       |           |              |               | 26203 | 26205 | 26410 | 26436  | 26586 | 27091  | 27425 | 27889  | 28008 | 28193  | 28200 | 28268 | 28950 |
| GZ03-02  | HP04  |           |              |               | C     | C     | A     | G      | T     | T      | C     | C      | A     | C      | C     | C     | G     |
| GZ03-01  | HP04  |           |              |               | C     | C     | A     | G      | T     | T      | C     | C      | A     | C      | C     | C     | G     |
| gzl227   | PC04  |           |              |               | C     | C     | A     | G      | T     | T      | C     | C      | A     | C      | C     | C     | G     |
| gzl136   | PC04  |           |              |               | C     | C     | A     | G      | T     | T      | C     | C      | A     | C      | C     | C     | G     |
| GZL13    | PC04  |           |              |               | C     | C     | A     | G      | T     | T      | C     | C      | A     | C      | C     | C     | G     |
| SZ16     | PC03  |           |              |               | C     | C     | A     | A      | T     | T      | C     | C      | A     | C      | C     | C     | A     |
| SZ3      | PC03  |           |              |               | C     | C     | A     | A      | T     | T      | C     | C      | A     | C      | C     | C     | A     |
| HGZ8L1-A | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| GZ02     | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| GD01     | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| ZS-C     | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| ZS-B     | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| ZS-A     | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| GZ50     | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| JMD      | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HSZ-A    | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HSZ-Bb   | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HSZ-Cc   | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HSZ-Cb   | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HSZ-Bc   | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| GZ-A     | HP03  |           |              |               | E     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HZS2-Bb  | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HSZ2-A   | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HGZ8L1-B | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| CUHK-W1  | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HZS2-D   | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HZS2-E   | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HZS2-C   | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| BJ03     | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| BJ02     | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| BJ01     | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HGZ8L2   | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| BJ04     | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| NS1      | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HZS2-Fb  | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| HZS2-Fc  | HP03  |           |              |               | M     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| QXC1     | HP03  |           |              |               | L     | C     | C     | A      | G     | A      | T     | T      | C     | C      | A     | C     | A     |
| GZ-D     | HP03  |           |              |               | L     | C     | C     | A      | N     | N      | N     | N      | N     | N      | N     | N     | N     |
| GZ-C     | HP03  |           |              |               | L     | C     | C     | A      | G     | A      | T     | T      | C     | G      | C     | A     | C     |
| GZ-B     | HP03  |           |              |               | L     | C     | C     | A      | G     | A      | T     | T      | C     | G      | C     | A     | C     |
| TOR2     | HP03  |           |              |               | L     | C     | C     | A      | G     | A      | T     | T      | C     | A      | C     | A     | C     |
| URBANI   | HP03  |           |              |               | L     | C     | C     | A      | G     | A      | T     | T      | C     | A      | C     | A     | C     |
| TC2      | HP03  |           |              |               | L     | T     | C     | A      | G     | A      | T     | T      | C     | A      | C     | A     | C     |
| TWY      | HP03  |           |              |               | L     | T     | C     | A      | G     | A      | T     | T      | C     | A      | C     | A     | C     |
| TWS      | HP03  |           |              |               | L     | T     | C     | A      | G     | A      | T     | T      | C     | A      | C     | A     | C     |
| TWK      | HP03  |           |              |               | L     | T     | C     | A      | G     | A      | T     | T      | C     | A      | C     | A     | C     |
| TWJ      | HP03  |           |              |               | L     | T     | C     | A      | G     | A      | T     | T      | C     | A      | C     | A     | C     |
| TC3      | HP03  |           |              |               | L     | T     | C     | A      | G     | A      | T     | T      | C     | A      | C     | A     | C     |
| TC1      | HP03  |           |              |               | L     | C     | C     | A      | G     | A      | T     | T      | C     | A      | C     | A     | C     |

F16 IF

[illegible]

| SNV | Codon | AA switch | AA residue # | nt coordinate |
|-----|-------|-----------|--------------|---------------|
| ct  | ccc3  | P-P       | 5709         | 17390         |
| ct  | gc12  | A-V       | 5704         | 17374         |
| ct  | lla2  | S-L       | 5623         | 17131         |
| ct  | cca1  | S-S       | 5542         | 16890         |
| lg  | ac13  | T-L       | 5426         | 16541         |
| ag  | cca3  | P-P       | 5354         | 16325         |
| ct  | lnc3  | V-Y       | 4948         | 14807         |
| lc  | lgl3  | C-C       | 4679         | 14300         |
| ct  | alc1  | L-F       | 4629         | 14151         |
| gg  | agg3  | R-R       | 4618         | 14117         |
| ct  | gac3  | D-D       | 4533         | 13862         |
| lg  | gc13  | A-A       | 4499         | 13760         |
| ct  | gac3  | D-D       | 4361         | 13347         |
| lc  | gic2  | V-A       | 3952         | 12219         |
| ag  | aag2  | K-R       | 3843         | 11792         |
| ac  | aac2  | N-T       | 3818         | 11717         |
| ct  | ggc3  | G-G       | 3750         | 11514         |
| ct  | lac3  | V-Y       | 3743         | 11493         |
| ct  | alc3  | I-L       | 3728         | 11448         |
| la  | cll3  | L-L       | 3571         | 10977         |
| ca  | gac3  | D-E       | 3488         | 10728         |
| ac  | aca3  | T-L       | 3441         | 10587         |
| ag  | ggg3  | G-G       | 3255         | 10029         |
| ct  | gcc2  | A-V       | 3197         | 9854          |
| ct  | gla2  | A-V       | 3072         | 9479          |
| lg  | lll1  | L-L       | 3062         | 9448          |
| ct  | glc2  | A-V       | 3047         | 9404          |
| lc  | gla2  | V-A       | 3024         | 9335          |
| ct  | gla2  | A-V       | 2971         | 9176          |
| ct  | lcl2  | S-F       | 2949         | 9110          |
| lc  | ac12  | L-L       | 2944         | 9095          |
| at  | gc13  | A-A       | 2894         | 8946          |
| lc  | clg1  | L-L       | 2851         | 8815          |
| gl  | gla1  | V-L       | 2770         | 8572          |
| ct  | la13  | V-Y       | 2765         | 8559          |
| g   | lgl3  | W-C       | 2746         | 8502          |
| la  | cll3  | L-L       | 2589         | 8031          |
| ct  | gc12  | A-V       | 2552         | 7919          |
| lc  | ag13  | S-S       | 2539         | 7861          |
| ct  | lac3  | V-Y       | 2537         | 7875          |
| ct  | lgc3  | C-C       | 2526         | 7842          |

FIG 1H

|  |  | nsf1.3 |  | nsf1.4 |  | nsf1.5 |  | nsf1.6 |  | nsf1.7 |  | nsf1.8 |  | nsf1.9 |  | nsf1.10 |  | nsf1.11 |  | nsf1.12 |  | nsf1.13 |  | nsf1.14 |  | nsf1.15 |  | nsf1.16 |  | nsf1.17 |  | nsf1.18 |  | nsf1.19 |  | nsf1.20 |  | nsf1.21 |  | nsf1.22 |  | nsf1.23 |  | nsf1.24 |  | nsf1.25 |  | nsf1.26 |  | nsf1.27 |  | nsf1.28 |  | nsf1.29 |  | nsf1.30 |  | nsf1.31 |  | nsf1.32 |  | nsf1.33 |  | nsf1.34 |  | nsf1.35 |  | nsf1.36 |  | nsf1.37 |  | nsf1.38 |  | nsf1.39 |  | nsf1.40 |  | nsf1.41 |  | nsf1.42 |  | nsf1.43 |  | nsf1.44 |  | nsf1.45 |  | nsf1.46 |  | nsf1.47 |  | nsf1.48 |  | nsf1.49 |  | nsf1.50 |  | nsf1.51 |  | nsf1.52 |  | nsf1.53 |  | nsf1.54 |  | nsf1.55 |  | nsf1.56 |  | nsf1.57 |  | nsf1.58 |  | nsf1.59 |  | nsf1.60 |  | nsf1.61 |  | nsf1.62 |  | nsf1.63 |  | nsf1.64 |  | nsf1.65 |  | nsf1.66 |  | nsf1.67 |  | nsf1.68 |  | nsf1.69 |  | nsf1.70 |  | nsf1.71 |  | nsf1.72 |  | nsf1.73 |  | nsf1.74 |  | nsf1.75 |  | nsf1.76 |  | nsf1.77 |  | nsf1.78 |  | nsf1.79 |  | nsf1.80 |  | nsf1.81 |  | nsf1.82 |  | nsf1.83 |  | nsf1.84 |  | nsf1.85 |  | nsf1.86 |  | nsf1.87 |  | nsf1.88 |  | nsf1.89 |  | nsf1.90 |  | nsf1.91 |  | nsf1.92 |  | nsf1.93 |  | nsf1.94 |  | nsf1.95 |  | nsf1.96 |  | nsf1.97 |  | nsf1.98 |  | nsf1.99 |  | nsf1.100 |  | nsf1.101 |  | nsf1.102 |  | nsf1.103 |  | nsf1.104 |  | nsf1.105 |  | nsf1.106 |  | nsf1.107 |  | nsf1.108 |  | nsf1.109 |  | nsf1.110 |  | nsf1.111 |  | nsf1.112 |  | nsf1.113 |  | nsf1.114 |  | nsf1.115 |  | nsf1.116 |  | nsf1.117 |  | nsf1.118 |  | nsf1.119 |  | nsf1.120 |  | nsf1.121 |  | nsf1.122 |  | nsf1.123 |  | nsf1.124 |  | nsf1.125 |  | nsf1.126 |  | nsf1.127 |  | nsf1.128 |  | nsf1.129 |  | nsf1.130 |  | nsf1.131 |  | nsf1.132 |  | nsf1.133 |  | nsf1.134 |  | nsf1.135 |  | nsf1.136 |  | nsf1.137 |  | nsf1.138 |  | nsf1.139 |  | nsf1.140 |  | nsf1.141 |  | nsf1.142 |  | nsf1.143 |  | nsf1.144 |  | nsf1.145 |  | nsf1.146 |  | nsf1.147 |  | nsf1.148 |  | nsf1.149 |  | nsf1.150 |  | nsf1.151 |  | nsf1.152 |  | nsf1.153 |  | nsf1.154 |  | nsf1.155 |  | nsf1.156 |  | nsf1.157 |  | nsf1.158 |  | nsf1.159 |  | nsf1.160 |  | nsf1.161 |  | nsf1.162 |  | nsf1.163 |  | nsf1.164 |  | nsf1.165 |  | nsf1.166 |  | nsf1.167 |  | nsf1.168 |  | nsf1.169 |  | nsf1.170 |  | nsf1.171 |  | nsf1.172 |  | nsf1.173 |  | nsf1.174 |  | nsf1.175 |  | nsf1.176 |  | nsf1.177 |  | nsf1.178 |  | nsf1.179 |  | nsf1.180 |  | nsf1.181 |  | nsf1.182 |  | nsf1.183 |  | nsf1.184 |  | nsf1.185 |  | nsf1.186 |  | nsf1.187 |  | nsf1.188 |  | nsf1.189 |  | nsf1.190 |  | nsf1.191 |  | nsf1.192 |  | nsf1.193 |  | nsf1.194 |  | nsf1.195 |  | nsf1.196 |  | nsf1.197 |  | nsf1.198 |  | nsf1.199 |  | nsf1.200 |  | nsf1.201 |  | nsf1.202 |  | nsf1.203 |  | nsf1.204 |  | nsf1.205 |  | nsf1.206 |  | nsf1.207 |  | nsf1.208 |  | nsf1.209 |  | nsf1.210 |  | nsf1.211 |  | nsf1.212 |  | nsf1.213 |  | nsf1.214 |  | nsf1.215 |  | nsf1.216 |  | nsf1.217 |  | nsf1.218 |  | nsf1.219 |  | nsf1.220 |  | nsf1.221 |  | nsf1.222 |  | nsf1.223 |  | nsf1.224 |  | nsf1.225 |  | nsf1.226 |  | nsf1.227 |  | nsf1.228 |  | nsf1.229 |  | nsf1.230 |  | nsf1.231 |  | nsf1.232 |  | nsf1.233 |  | nsf1.234 |  | nsf1.235 |  | nsf1.236 |  | nsf1.237 |  | nsf1.238 |  | nsf1.239 |  | nsf1.240 |  | nsf1.241 |  | nsf1.242 |  | nsf1.243 |  | nsf1.244 |  | nsf1.245 |  | nsf1.246 |  | nsf1.247 |  | nsf1.248 |  | nsf1.249 |  | nsf1.250 |  | nsf1.251 |  | nsf1.252 |  | nsf1.253 |  | nsf1.254 |  | nsf1.255 |  | nsf1.256 |  | nsf1.257 |  | nsf1.258 |  | nsf1.259 |  | nsf1.260 |  | nsf1.261 |  | nsf1.262 |  | nsf1.263 |  | nsf1.264 |  | nsf1.265 |  | nsf1.266 |  | nsf1.267 |  | nsf1.268 |  | nsf1.269 |  | nsf1.270 |  | nsf1.271 |  | nsf1.272 |  | nsf1.273 |  | nsf1.274 |  | nsf1.275 |  | nsf1.276 |  | nsf1.277 |  | nsf1.278 |  | nsf1.279 |  | nsf1.280 |  | nsf1.281 |  | nsf1.282 |  | nsf1.283 |  | nsf1.284 |  | nsf1.285 |  | nsf1.286 |  | nsf1.287 |  | nsf1.288 |  | nsf1.289 |  | nsf1.290 |  | nsf1.291 |  | nsf1.292 |  | nsf1.293 |  | nsf1.294 |  | nsf1.295 |  | nsf1.296 |  | nsf1.297 |  | nsf1.298 |  | nsf1.299 |  | nsf1.300 |  | nsf1.301 |  | nsf1.302 |  | nsf1.303 |  | nsf1.304 |  | nsf1.305 |  | nsf1.306 |  | nsf1.307 |  | nsf1.308 |  | nsf1.309 |  | nsf1.310 |  | nsf1.311 |  | nsf1.312 |  | nsf1.313 |  | nsf1.314 |  | nsf1.315 |  | nsf1.316 |  | nsf1.317 |  | nsf1.318 |  | nsf1.319 |  | nsf1.320 |  | nsf1.321 |  | nsf1.322 |  | nsf1.323 |  | nsf1.324 |  | nsf1.325 |  | nsf1.326 |  | nsf1.327 |  | nsf1.328 |  | nsf1.329 |  | nsf1.330 |  | nsf1.331 |  | nsf1.332 |  | nsf1.333 |  | nsf1.334 |  | nsf1.335 |  | nsf1.336 |  | nsf1.337 |  | nsf1.338 |  | nsf1.339 |  | nsf1.340 |  | nsf1.341 |  | nsf1.342 |  | nsf1.343 |  | nsf1.344 |  | nsf1.345 |  | nsf1.346 |  | nsf1.347 |  | nsf1.348 |  | nsf1.349 |  | nsf1.350 |  | nsf1.351 |  | nsf1.352 |  | nsf1.353 |  | nsf1.354 |  | nsf1.355 |  | nsf1.356 |  | nsf1.357 |  | nsf1.358 |  | nsf1.359 |  | nsf1.360 |  | nsf1.361 |  | nsf1.362 |  | nsf1.363 |  | nsf1.364 |  | nsf1.365 |  | nsf1.366 |  | nsf1.367 |  | nsf1.368 |  | nsf1.369 |  | nsf1.370 |  | nsf1.371 |  | nsf1.372 |  | nsf1.373 |  | nsf1.374 |  | nsf1.375 |  | nsf1.376 |  | nsf1.377 |  | nsf1.378 |  | nsf1.379 |  | nsf1.380 |  | nsf1.381 |  | nsf1.382 |  | nsf1.383 |  | nsf1.384 |  | nsf1.385 |  | nsf1.386 |  | nsf1.387 |  | nsf1.388 |  | nsf1.389 |  | nsf1.390 |  | nsf1.391 |  | nsf1.392 |  | nsf1.393 |  | nsf1.394 |  | nsf1.395 |  | nsf1.396 |  | nsf1.397 |  | nsf1.398 |  | nsf1.399 |  | nsf1.400 |  | nsf1.401 |  | nsf1.402 |  | nsf1.403 |  | nsf1.404 |  | nsf1.405 |  | nsf1.406 |  | nsf1.407 |  | nsf1.408 |  | nsf1.409 |  | nsf1.410 |  |
|--|--|--------|--|--------|--|--------|--|--------|--|--------|--|--------|--|--------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|
|--|--|--------|--|--------|--|--------|--|--------|--|--------|--|--------|--|--------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|---------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|----------|--|

F/6 1I

|           |       |           |              |               | Sars3a |       | Sars3b    |              |               |   |   |   |   |   |   |   |   |   |   |
|-----------|-------|-----------|--------------|---------------|--------|-------|-----------|--------------|---------------|---|---|---|---|---|---|---|---|---|---|
| SNV       | Codon | AA switch | AA residue # | nt coordinate | SNV    | Codon | AA switch | AA residue # | nt coordinate |   |   |   |   |   |   |   |   |   |   |
| gt        | leg3  | S-S       |              | 26129         | gt     | leg3  | S-S       |              | 26129         |   |   |   |   |   |   |   |   |   |   |
| ac        | cca3  | P-P       | 121          | 26050         | ac     | cca3  | P-P       | 121          | 26050         |   |   |   |   |   |   |   |   |   |   |
| at        | gcc3  | A-A       | 115          | 26032         | at     | gcc3  | A-A       | 115          | 26032         |   |   |   |   |   |   |   |   |   |   |
| ta        | agg1  | V-E       | 52           | 25844         | ta     | agg1  | V-E       | 52           | 25844         |   |   |   |   |   |   |   |   |   |   |
| ac        | aaa1  | K-Q       | 40           | 25808         | ac     | aaa1  | K-Q       | 40           | 25808         |   |   |   |   |   |   |   |   |   |   |
| ca        | gaa2  | K-E       | 31           | 25779         | ca     | gaa2  | K-E       | 31           | 25779         |   |   |   |   |   |   |   |   |   |   |
| ta        | gat3  | D-E       | 2            | 25693         | ta     | gat3  | D-E       | 2            | 25693         |   |   |   |   |   |   |   |   |   |   |
| tg        | tg11  | C-G       | 121          | 25628         | tg     | tg11  | C-G       | 121          | 25628         |   |   |   |   |   |   |   |   |   |   |
| ga        | ggc2  | G-D       | 100          | 25566         | ga     | ggc2  | G-D       | 100          | 25566         |   |   |   |   |   |   |   |   |   |   |
| ct        | cat1  | H-Y       | 93           | 25544         | ct     | cat1  | H-Y       | 93           | 25544         |   |   |   |   |   |   |   |   |   |   |
| ta        | cig2  | L-Q       | 85           | 25521         | ta     | cig2  | L-Q       | 85           | 25521         |   |   |   |   |   |   |   |   |   |   |
| ta        | lgc1  | C-S       | 81           | 25508         | ta     | lgc1  | C-S       | 81           | 25508         |   |   |   |   |   |   |   |   |   |   |
| ag        | cat2  | H-R       | 30           | 25366         | ag     | cat2  | H-R       | 30           | 25366         |   |   |   |   |   |   |   |   |   |   |
| ca        | ccc2  | P-H       | 25           | 25341         | ca     | ccc2  | P-H       | 25           | 25341         |   |   |   |   |   |   |   |   |   |   |
| ga        | aga2  | R-K       | 11           | 25299         | ga     | aga2  | R-K       | 11           | 25299         |   |   |   |   |   |   |   |   |   |   |
| ta        | ttt1  | L-F       | 7            | 25286         | ta     | ttt1  | L-F       | 7            | 25286         |   |   |   |   |   |   |   |   |   |   |
| ct        | cic1  | L-F       | 1247         | 25230         | ct     | cic1  | L-F       | 1247         | 25230         |   |   |   |   |   |   |   |   |   |   |
| ct        | gcc2  | A-V       | 1208         | 25114         | ct     | gcc2  | A-V       | 1208         | 25114         |   |   |   |   |   |   |   |   |   |   |
| tc        | att3  | L-I       | 1160         | 25031         | tc     | att3  | L-I       | 1160         | 25031         |   |   |   |   |   |   |   |   |   |   |
| ga        | aaa1  | E-K       | 1163         | 24978         | ga     | aaa1  | E-K       | 1163         | 24978         |   |   |   |   |   |   |   |   |   |   |
| ct        | cht1  | D-D       | 1148         | 24933         | ct     | cht1  | D-D       | 1148         | 24933         |   |   |   |   |   |   |   |   |   |   |
| ct        | lgl3  | C-C       | 1025         | 24566         | ct     | lgl3  | C-C       | 1025         | 24566         |   |   |   |   |   |   |   |   |   |   |
| gt        | agg2  | H-M       | 1001         | 24493         | gt     | agg2  | H-M       | 1001         | 24493         |   |   |   |   |   |   |   |   |   |   |
| gt        | leg3  | S-S       | 964          | 24363         | gt     | leg3  | S-S       | 964          | 24363         |   |   |   |   |   |   |   |   |   |   |
| ag        | acc1  | L-A       | 894          | 24171         | ag     | acc1  | L-A       | 894          | 24171         |   |   |   |   |   |   |   |   |   |   |
| gt        | ttt1  | D-Y       | 776          | 23823         | gt     | ttt1  | D-Y       | 776          | 23823         |   |   |   |   |   |   |   |   |   |   |
| ct        | gct2  | A-V       | 765          | 23785         | ct     | gct2  | A-V       | 765          | 23785         |   |   |   |   |   |   |   |   |   |   |
| ct        | gct2  | A-V       | 754          | 23752         | ct     | gct2  | A-V       | 754          | 23752         |   |   |   |   |   |   |   |   |   |   |
| ag        | gcc3  | A-A       | 748          | 23735         | ag     | gcc3  | A-A       | 748          | 23735         |   |   |   |   |   |   |   |   |   |   |
| cg        | aca2  | T-R       | 743          | 23719         | cg     | aca2  | T-R       | 743          | 23719         |   |   |   |   |   |   |   |   |   |   |
| ag        | aca1  | L-A       | 743          | 23718         | ag     | aca1  | L-A       | 743          | 23718         |   |   |   |   |   |   |   |   |   |   |
| ct        | tea2  | S-L       | 701          | 23563         | ct     | tea2  | S-L       | 701          | 23563         |   |   |   |   |   |   |   |   |   |   |
| lc        | tta2  | L-S       | 665          | 23485         | lc     | tta2  | L-S       | 665          | 23485         |   |   |   |   |   |   |   |   |   |   |
| ta        | gat3  | D-E       | 613          | 23330         | ta     | gat3  | D-E       | 613          | 23330         |   |   |   |   |   |   |   |   |   |   |
| ct        | gcc2  | A-V       | 609          | 23317         | ct     | gcc2  | A-V       | 609          | 23317         |   |   |   |   |   |   |   |   |   |   |
| gt        | gca1  | A-S       | 609          | 23316         | gt     | gca1  | A-S       | 609          | 23316         |   |   |   |   |   |   |   |   |   |   |
| lc        | tc11  | S-P       | 607          | 23310         | lc     | tc11  | S-P       | 607          | 23310         |   |   |   |   |   |   |   |   |   |   |
| lc        | glt3  | V-V       | 596          | 23279         | lc     | glt3  | V-V       | 596          | 23279         |   |   |   |   |   |   |   |   |   |   |
| lg        | tc11  | S-A       | 577          | 23220         | lg     | tc11  | S-A       | 577          | 23220         |   |   |   |   |   |   |   |   |   |   |
| ct        | lcc3  | S-S       | 561          | 23174         | ct     | lcc3  | S-S       | 561          | 23174         |   |   |   |   |   |   |   |   |   |   |
| cg        | acc2  | T-S       | 487          | 22951         | cg     | acc2  | T-S       | 487          | 22951         |   |   |   |   |   |   |   |   |   |   |
| ag        | gat2  | D-G       | 480          | 22930         | ag     | gat2  | D-G       | 480          | 22930         |   |   |   |   |   |   |   |   |   |   |
| TW8       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TWH       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TW7       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TW6       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TW11      | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TW10      | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| GD69      | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| HSR       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TWC       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| SIN2748   | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| SIN2677   | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| SIN2500   | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| PUMC03    | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| Sino1-11  | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| PUMC02    | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| Sino3-11  | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| PUMC01    | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| SIN2679   | HP03  | L         | A            | C             | T      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| Sin850    | HP03  | L         | A            | C             | T      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| Sin849    | HP03  | L         | A            | C             | T      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| Sin847    | HP03  | L         | A            | C             | T      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| Sin848    | HP03  | L         | A            | C             | T      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| SIN845    | HP03  | L         | A            | C             | T      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| Sin3765V  | HP03  | L         | A            | C             | T      | T     | C         | T            | G             | C | T | T | C | A | C | A | C | A | C |
| Sin852    | HP03  | L         | A            | C             | T      | T     | C         | T            | G             | C | T | T | C | A | C | A | C | A | C |
| AS        | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| CUHK-SU10 | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| SIN2774   | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| FRANKFURT | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| FRA       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| SoD       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| HKU39849  | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| WHU       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| CUHK-LC1  | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TW5       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TW4       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TW3       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TW2       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| TW1       | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| CUHK-LC2  | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| CUHK-LC3  | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| CUHK-LC4  | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |
| CUHK-LC5  | HP03  | L         | A            | C             | C      | T     | T         | T            | G             | C | T | T | T | C | A | C | A | C | C |

File 15

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

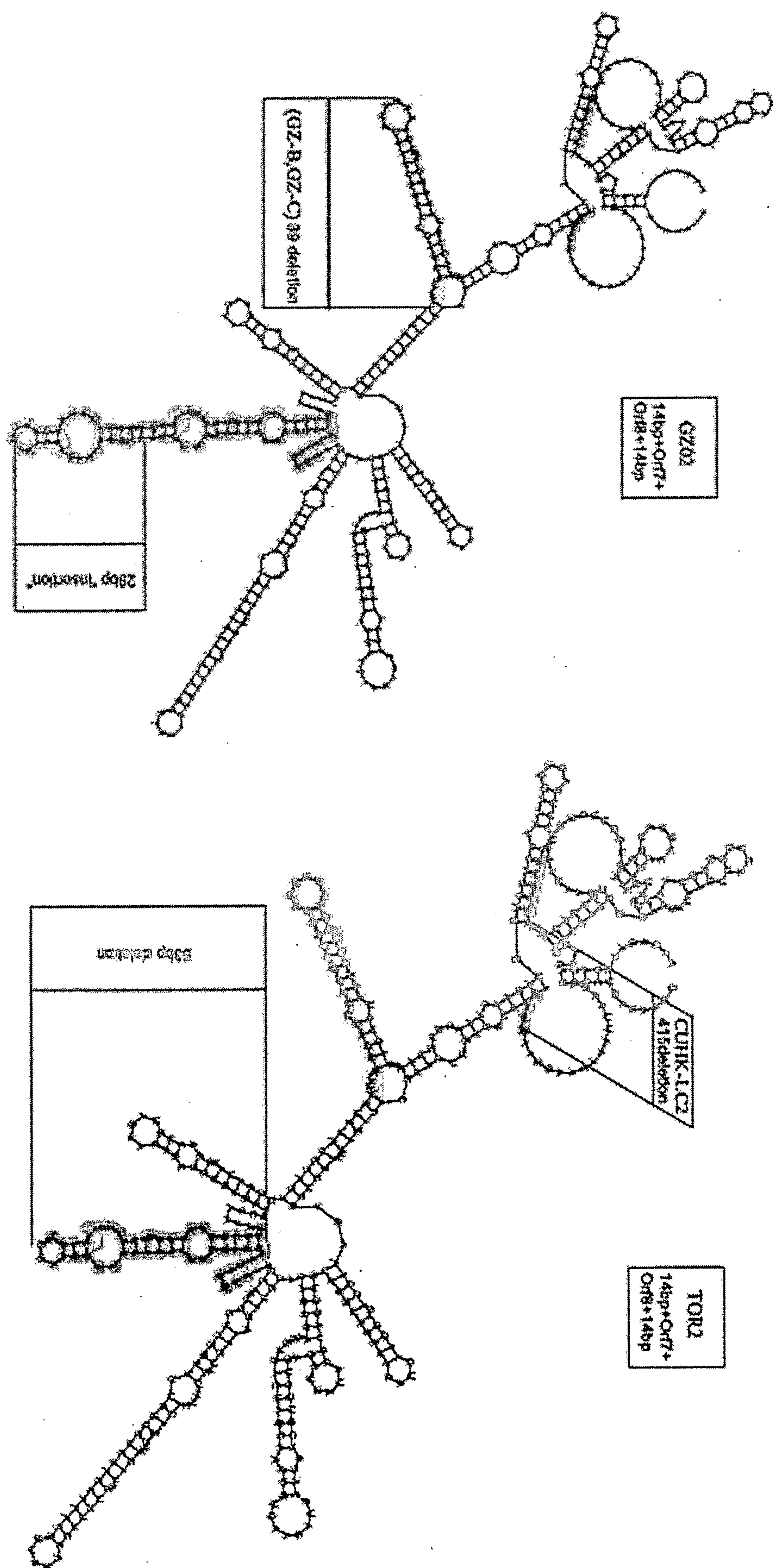


FIG. 3

Figs. 4

|                                   | Orf7b                                                                                             | Orf8a                                                                                                                                                                                           | Orf8b                                                                                                                                                            | Nucleocapsid Protein                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|-----------------------------------|---------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| GZ02                              | 27635: 27769<br>MNELTLIDFYLCFLA<br>FLFLVLIIMLIIFWF<br>SLEIQDLEEPCTKV*                             | 27776: 28144 (Sars8)<br>MKLLIVLTICISLCS<br>CIRTVQRCASNKP<br>HYLEDPCPT<br>GYQPEWNI<br>RNTGNTYSTA<br>WLCALGKYL<br>PFRHWTM<br>VQTCIPNVT<br>INCQDPAG<br>CALIARCWY<br>LHEGHTA<br>AFK<br>DVLVYLNKRTN* | 27861: 28144 sars8b<br>unknown (sars8b)<br>MCLKILVRYNTGNTYSTA<br>WLCALGKYL<br>PFRHWTM<br>VQTCIPNVT<br>INCQDPAG<br>CALI<br>ARCWY<br>LHEGHTA<br>AFRDL<br>VYLNKRTN* | 28146: 29414 sars9a nucleocapsid protein<br>MSDNGPQSNQRSAPRITFCGPTDSTDNQNGCRNGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPFGQVPIINTNSGPDQIGYYR<br>RATRVRGCGDKAKELSPRWYFYLLGTGPEASLPYGANKEGIYWVATEGALNTPKHIGTRNPNMAATVQLPQCTLLPKGFVAEGSR<br>GGSQASSRSSSRSGNSRNSTPGSSRGNSPARMAASCGCETALALLDLRLNQLESKVSCKGQDQCTVTKKSAEASKPRQKRTATK<br>QYVNTQAFGRGPEQTQCNFGDQDLIRGCTDYKHWPIAQFAPASASAFGMSRIGMEVTPSGTWLTYHGAIKLDKDPQFKDNVILLNK<br>HIDAYKTFPPTPEPKDKKKTDEAQLPQRQKKQPTVTLPAADMDFSRQLQNSMSGASADSTQA*         |
| TOR2<br>29 dels                   | 27635: 27769<br>MNELTLIDFYLCFLA<br>FLFLVLIIMLIIFWF<br>SLEIQDLEEPCTKV*                             | 27776: 27924<br>sars8a unknown<br>(sars8a)<br>MKLLIVLTICISLCS<br>CIRTVQRCASNKP<br>HYLEDPCXVQH                                                                                                   | 27861: 28144 sars8b<br>unknown (sars8b)<br>MCLKILVRYNTGNTYSTA<br>WLCALGKYL<br>PFRHWTM<br>VQTCIPNVT<br>INCQDPAG<br>CALI<br>ARCWY<br>LHEGHTA<br>AFRDL<br>VYLNKRTN* | 28146: 29414 sars9a nucleocapsid protein<br>MSDNGPQSNQRSAPRITFCGPTDSTDNQNGCRNGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPFGQVPIINTNSGPDQIGYYR<br>RATRVRGCGDKAKELSPRWYFYLLGTGPEASLPYGANKEGIYWVATEGALNTPKHIGTRNPNMAATVQLPQCTLLPKGFVAEGSR<br>GGSQASSRSSSRSGNSRNSTPGSSRGNSPARMAASCGCETALALLDLRLNQLESKVSCKGQDQCTVTKKSAEASKPRQKRTATK<br>QYVNTQAFGRGPEQTQCNFGDQDLIRGCTDYKHWPIAQFAPASASAFGMSRIGMEVTPSGTWLTYHGAIKLDKDPQFKDNVILLNK<br>HIDAYKTFPPTPEPKDKKKTDEAQLPQRQKKQPTVTLPAADMDFSRQLQNSMSGASADSTQA*         |
| HGZBL1-B<br>82 dels               | 27635: 27769<br>MNELTLIDFYLCFLA<br>FLFLVLIIMLIIFWF<br>SLEIQDLEEPCTKV*                             | 27776: 27963<br>MKLLIVLTICISLCS<br>CIRTVQRCASNALL<br>GFVL*                                                                                                                                      | *                                                                                                                                                                | 28146: 29414 sars9a nucleocapsid protein<br>MSDNGPQSNQRSAPRITFCGPTDSTDNQNGCRNGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPFGQVPIINTNSGPDQIGYYR<br>RATRVRGCGDKAKELSPRWYFYLLGTGPEASLPYGANKEGIYWVATEGALNTPKHIGTRNPNMAATVQLPQCTLLPKGFVAEGSR<br>GGSQASSRSSSRSGNSRNSTPGSSRGNSPARMAASCGCETALALLDLRLNQLESKVSCKGQDQCTVTKKSAEASKPRQKRTATK<br>QYVNTQAFGRGPEQTQCNFGDQDLIRGCTDYKHWPIAQFAPASASAFGMSRIGMEVTPSGTWLTYHGAIKLDKDPQFKDNVILLNK<br>HIDAYKTFPPTPEPKDKKKTDEAQLPQRQKKQPTVTLPAADMDFSRQLQNSMSGASADSTQA*         |
| CUHK-LC2<br>415 dels              | 27635: 27769<br>MNELTLIDFYLCFLA<br>FLFLVLIIMLIIFWF<br>SLEIQDLEEPCTKV*                             | *                                                                                                                                                                                               | *                                                                                                                                                                | 27635: 29414<br>MNELTLIDFYLCFLAFLFLVLIIMLIINEQIKMSDNGPQSNQRSAPRITFCGPTDSTDNQNGCRNGARPKQRRPQGLPNNTASWFTAL<br>TQHGKEELRFPFGQVPIINTNSGPDQIGYYRRAATRVRGCGDKAKELSPRWYFYLLGTGPEASLPYGANKEGIYWVATEGALNTPKH<br>IGTRNPNMAATVQLPQCTLLPKGFVAEGSRGGSQASSRSSSRSGNSRNSTPGSSRGNSPARMAASCGCETALALLDLRLNQLESKY<br>SGKQDQCTVTKKSAEASKPRQKRTATKQYVNTQAFGRGPEQTQCNFGDQDLIRGCTDYKHWPIAQFAPASASAFGMSRIGME<br>VTPSGTWLTYHGAIKLDKDPQFKDNVILLNKHIDAYKTFPPTPEPKDKKKTDEAQLPQRQKKQPTVTLPAADMDFSRQLQNSMS<br>GASADSTQA* |
| GZ-C, GZ-B<br>39 dels<br>+29 dels | 27638: 27844<br>MNELTLIDFYLCFLA<br>FLFLVLIIMLIIFWF<br>SLEIQDLEEPCTKVS<br>LCSCTVQRCASN<br>KPIVLEDP | *                                                                                                                                                                                               | *                                                                                                                                                                | 28146: 29414 sars9a nucleocapsid protein<br>MSDNGPQSNQRSAPRITFCGPTDSTDNQNGCRNGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPFGQVPIINTNSGPDQIGYYR<br>RATRVRGCGDKAKELSPRWYFYLLGTGPEASLPYGANKEGIYWVATEGALNTPKHIGTRNPNMAATVQLPQCTLLPKGFVAEGSR<br>GGSQASSRSSSRSGNSRNSTPGSSRGNSPARMAASCGCETALALLDLRLNQLESKVSCKGQDQCTVTKKSAEASKPRQKRTATK<br>QYVNTQAFGRGPEQTQCNFGDQDLIRGCTDYKHWPIAQFAPASASAFGMSRIGMEVTPSGTWLTYHGAIKLDKDPQFKDNVILLNK<br>HIDAYKTFPPTPEPKDKKKTDEAQLPQRQKKQPTVTLPAADMDFSRQLQNSMSGASADSTQA*         |

FIG. 5

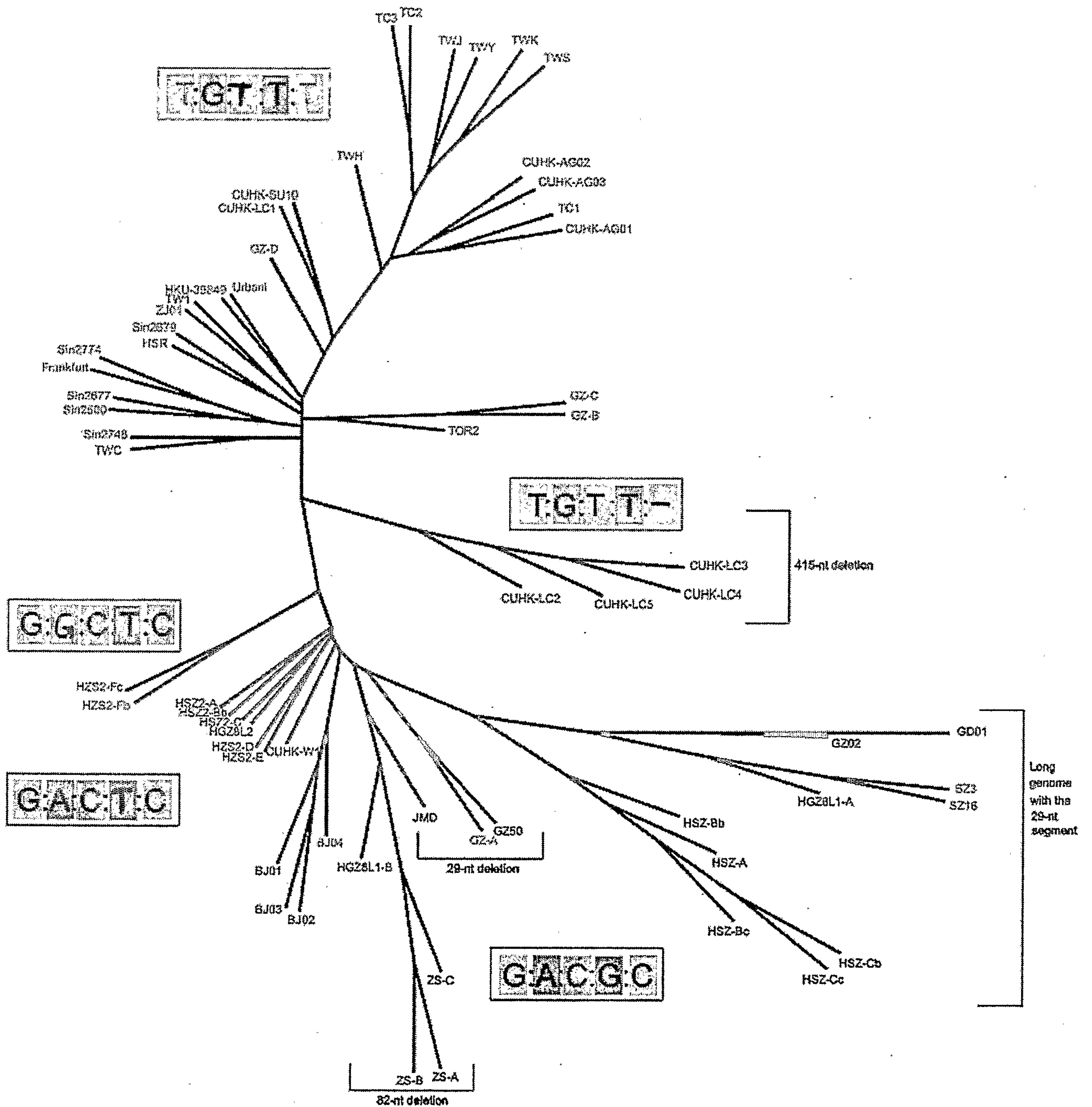


FIG. 6

22/24

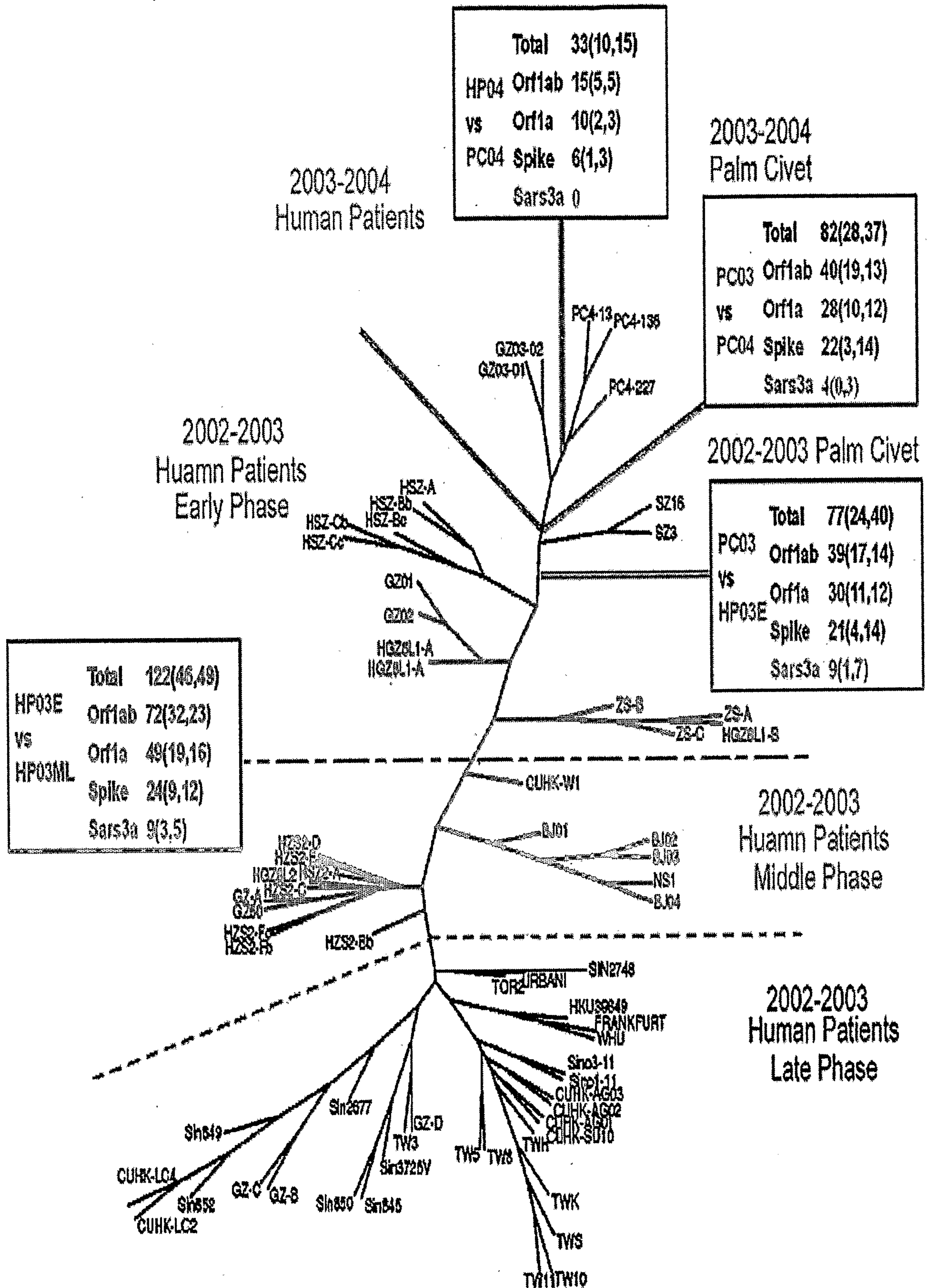


FIG. 7A

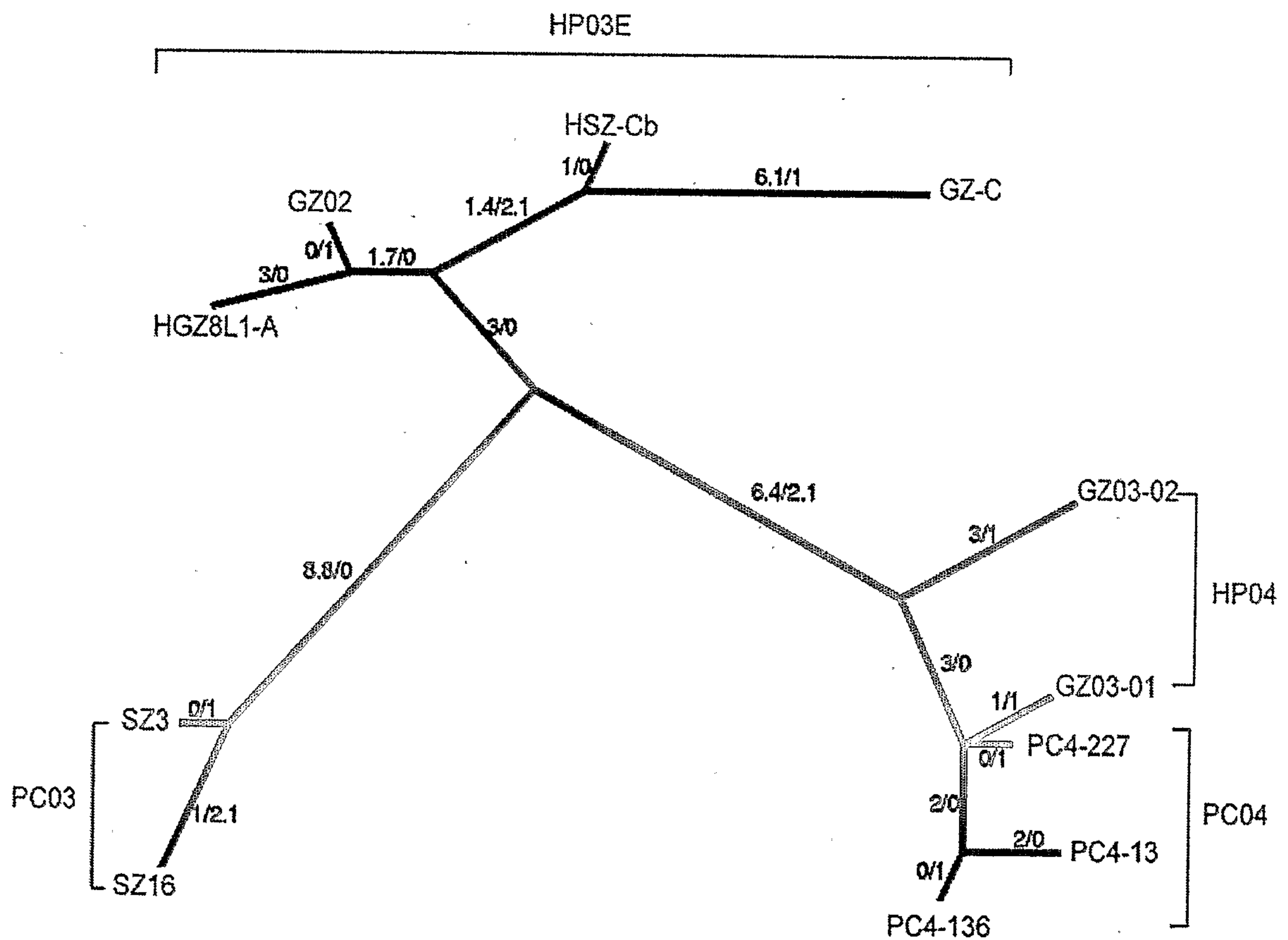


FIG. 7B

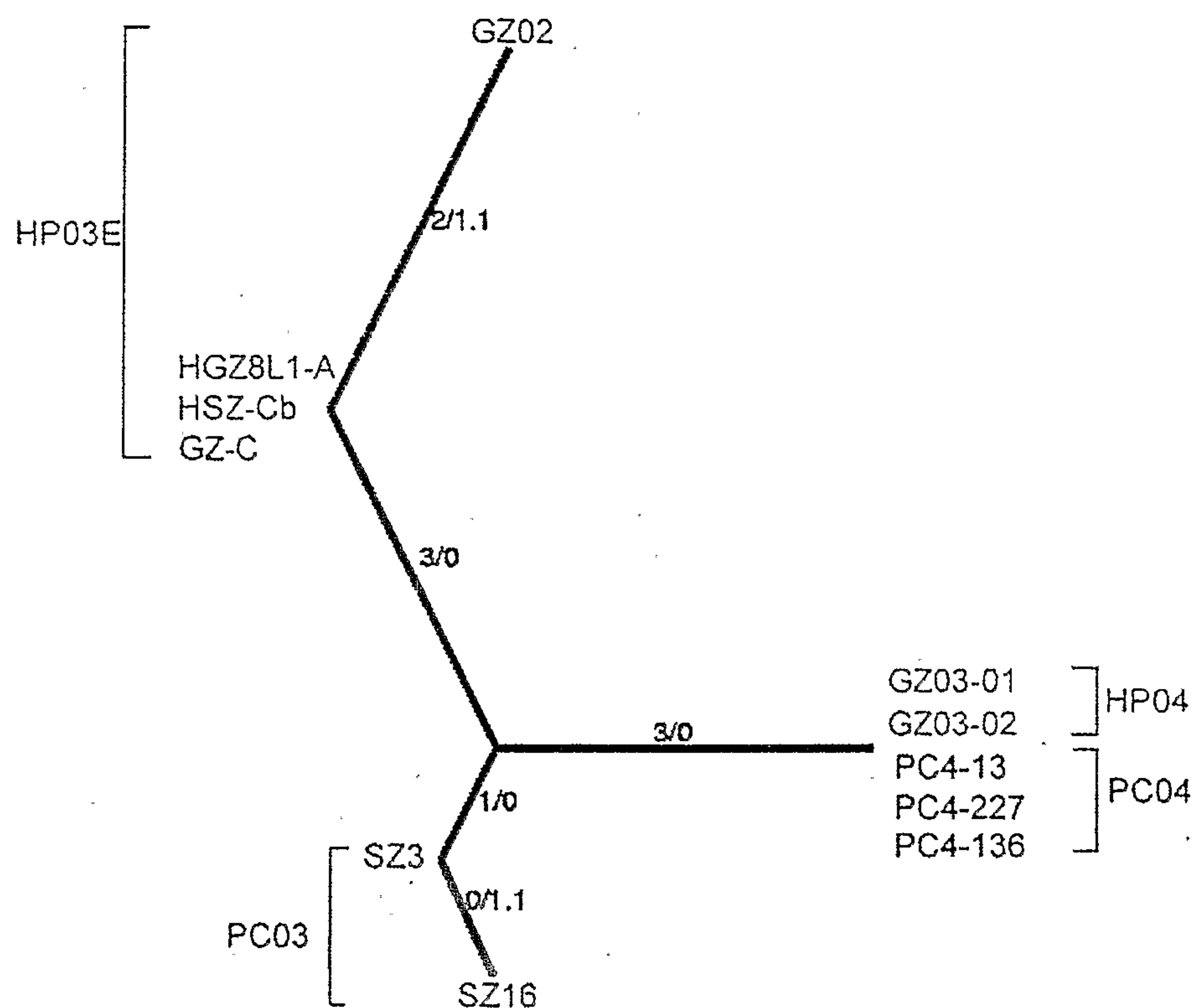


FIG. 7C

