ELSEVIER

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

ELSEVIER

Journal of Clinical Virology 38 (2007) 19–26

JOURNAL OF CLINICAL VIROLOGY

www.elsevier.com/locate/jcv

Characterizing 56 complete SARS-CoV S-gene sequences from Hong Kong

Julian W. Tang[a], Jo L.K. Cheung[a], Ida M.T. Chu[a], Margaret Ip[a],
Mamie Hui[a], Malik Peiris[b], Paul K.S. Chan[a,c,*]

[a] *Department of Microbiology, School of Public Health, The Chinese University of Hong Kong,*
Prince of Wales Hospital, Shatin, Hong Kong SAR, China
[b] *Department of Microbiology, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong SAR, China.*
[c] *Centre for Emerging Infectious Diseases, School of Public Health, The Chinese University of Hong Kong,*
Prince of Wales Hospital, Shatin, Hong Kong SAR, China

Received 11 May 2006; received in revised form 18 September 2006; accepted 6 October 2006

Abstract

Background: The spike glycoprotein (S) gene of the severe acute respiratory syndrome-associated coronavirus (SARS-CoV) has been useful in analyzing the molecular epidemiology of the 2003 SARS outbreaks.
Objectives: To characterize complete SARS-CoV S-gene sequences from Hong Kong.
Study design: Fifty-six SARS-CoV S-gene sequences, obtained from patients who presented with SARS to the Prince of Wales Hospital during March–May 2003, were analysed using a maximum likelihood (ML) approach, together with 138 other (both human and animal) S-gene sequences downloaded from GenBank.
Results: The maximum-likelihood (ML) trees showed little evolution occurring within these 56 sequences. Analysis with the other sequences, showed three distinct SARS clusters, closely correlated to previously defined early, middle and late phases of the 2003 international SARS outbreaks. In addition, two new single nucleotide variations (SNVs), T21615A and T21901A, were discovered, not previously reported elsewhere.
Conclusions: The ML approach to the reconstruction of tree phylogenies is known to be superior to the more popular, less computationally and time-demanding neighbour-joining (NJ) approach. The ML analysis in this study confirms the previously reported SARS epidemiology analysed mostly using the NJ approach. The two new SNVs reported here are most likely due to the tissue-culture passaging of the clinical samples.
© 2006 Elsevier B.V. All rights reserved.

Keywords: SARS-CoV; S-gene; Phylogenetics; Maximum-likelihood; Neighbour-joining; Single nucleotide variation

1. Introduction

Since the severe acute respiratory syndrome (SARS) epidemic of 2003 around the world, many researchers have attempted to determine the natural reservoir of the SARS-associated coronavirus (SARS-CoV). Studies on the possible animal source of SARS-CoV have mainly focused on the Himalyan palm civet (*Paguma larvata*), though other animals (e.g. the raccoon dog, *Nyctereutes procyonoides*) have been shown to carry coronaviruses closely related to the SARS-CoV. Of note, these related animal coronaviruses possess a 29 base-pair sequence (position 27,869–27,897) in the

---

Abbreviations: SARS-CoV, severe acute respiratory syndrome-associated coronavirus; S-gene, spike glycoprotein gene; CS, clinical sample; TC, tissue culture; RT-PCR, reverse transcription polymerase chain reaction; SNV, single nucleotide variation; ML, maximum likelihood; NJ, neighbour-joining; CUHK, Chinese University of Hong Kong

[*] Corresponding author at: Department of Microbiology, School of Public Health, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China. Tel.: +852 2632 3333; fax: +852 2647 3227.
E-mail address: paulkschan@cuhk.edu.hk (P.K.S. Chan).

1386-6532/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jcv.2006.10.001

putative open reading frame (ORF) 11 (Marra et al., 2003), that is absent in most human SARS-CoV isolates (Guan et al., 2003; Song et al., 2005). A more recent study on SARS-CoV-like coronaviruses from palm civets and raccoon dogs from various live markets in China identified a series of single nucleotide variations (SNVs) in the SARS-CoV spike (S) glycoprotein gene (Kan et al., 2005). The authors suggested that these SNVs marked the transmission of SARS-CoV-like viruses from these animals into humans during various phases of the SARS epidemic. Most recently, SARS-like CoV have been found in bats though the S-gene homology with SARS-CoV is low, around 80%, and it is still uncertain as to whether this is the true natural reservoir for SARS-CoV (Lau et al., 2005; Li et al., 2005; Poon et al., 2005a,b).

Guan et al. (2004) described the genetic variation of the first 2149 bp of the S-gene, mainly focusing on the S1 (amino) region. This study compared local SARS-CoV S1 gene sequences collected from 137 Hong Kong SARS patients from February to April 2003, with 27 other sequences then available from GenBank, using neighbour-joining (NJ) phylogenetic analysis. The authors concluded that the international SARS epidemics were caused by closely related SARS-CoVs. However, some strains isolated from Hong Kong in the early phase of epidemic, were distinct from the international outbreak strain, and may have represented transitory strains of SARS-CoV (Chim et al., 2004). Furthermore, analysis of SARS-CoV strains subsequently isolated from China suggested that this virus may have been circulating there, unidentified, for some time before the international epidemic occurred (Chinese SARS Molecular Epidemiology Consortium, CSMEC, 2004; Zhao et al., 2004).

In order to further characterize the evolution of SARS-CoV, S-gene sequences of isolates from patients in Hong Kong were analysed using a maximum likelihood (ML) method, and any new SNVs documented.

2. Materials and methods

2.1. Patients, sample collection and processing

Patients with laboratory-confirmed SARS-CoV infection admitted to the New Territories East Cluster Hospitals in Hong Kong were included (Chan et al., 2004). For each patient, a stored original clinical sample (CS), or its primary isolate from tissue culture (TC) when the original sample has been exhausted, was retrieved for this study. The nucleotide sequence of the whole S-gene was obtained by direct sequencing. Briefly, the extracted RNA was amplified with two sets of overlapping primers using Superscript^{TM} III One-Step RT-PCR System with Platinum *Taq* DNA polymerase (Invitrogen, Life Technology, Carlsbad, CA). Purified PCR products were sequenced using six internal sequencing primers using BigDye$^{®}$ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Their sequences were edited and aligned using the Seqscape v2.1.1 software (Applied Biosystems, Foster City, CA).

All 56 Hong Kong S-gene sequences obtained in this study were deposited on the National Center for Biotechnology Information Genome Database (GenBank, accession numbers: DQ412574–DQ412629).

2.2. SARS-CoV S-gene phylogenetic analysis

The S-gene sequences were aligned using the Clustal X's multiple alignment algorithm, then manually checked and edited using Bioedit v7.0.4.1 (Tippmann, 2004). After alignment and editing, the total length of the S-gene sequences was 3759 bp. A maximum likelihood (ML) tree were then constructed using the program PAUP* v4.0b10 (Swofford, 2001) under an optimum model of evolution as selected by the program Modeltest v3.7 (Posada and Crandall, 1998). The robustness of the trees' topology was statistically assessed by bootstrap analysis, with a minimum of 1000 rounds of replication. To put these 56 CUHK S-gene sequences in context, 138 other S-gene sequences were downloaded from the GenBank and these 194 S-gene sequences were drawn, using the same methods above. Both trees were rooted against a civet cat sequence (CIV007, AY572034), and plotted using NJPlot (Perrière and Gouy, 1996).

3. Results

3.1. Patients, sample collection and processing

A total of 56 SARS patients were included in this study (mean age 54.8, S.D. 21.9 years, 21 males). These patients all became ill during March–May 2003. Samples for analysis were collected after a mean of 5.9 (S.D. 4.8) days after illness onset. Table 1 shows the specimen type and numbers used in this study, 11 were original clinical samples (CS) and 45 were primary tissue culture isolates (TC).

3.2. SARS-CoV S-gene phylogenetic analysis

The ML trees in Figs. 1 and 2 were both constructed using a Kimura three-parameters model of base substitution with unequal base frequencies, a proportion of invariant sites and

Table 1
Types of specimens used for S-gene sequencing

Specimen type	No. of specimens tested	Clinical sample (CS)	Tissue culture isolate (TC)
Respiratory[a]	44	4	40
Stool	5	5	0
Urine	2	2	0
Tissue[b]	5	0	5
Total	56	11	45

[a] Includes nasopharyngeal aspirate, throat and nasal swabs.

[b] Includes four lung and one terminal ileum biopsy samples.

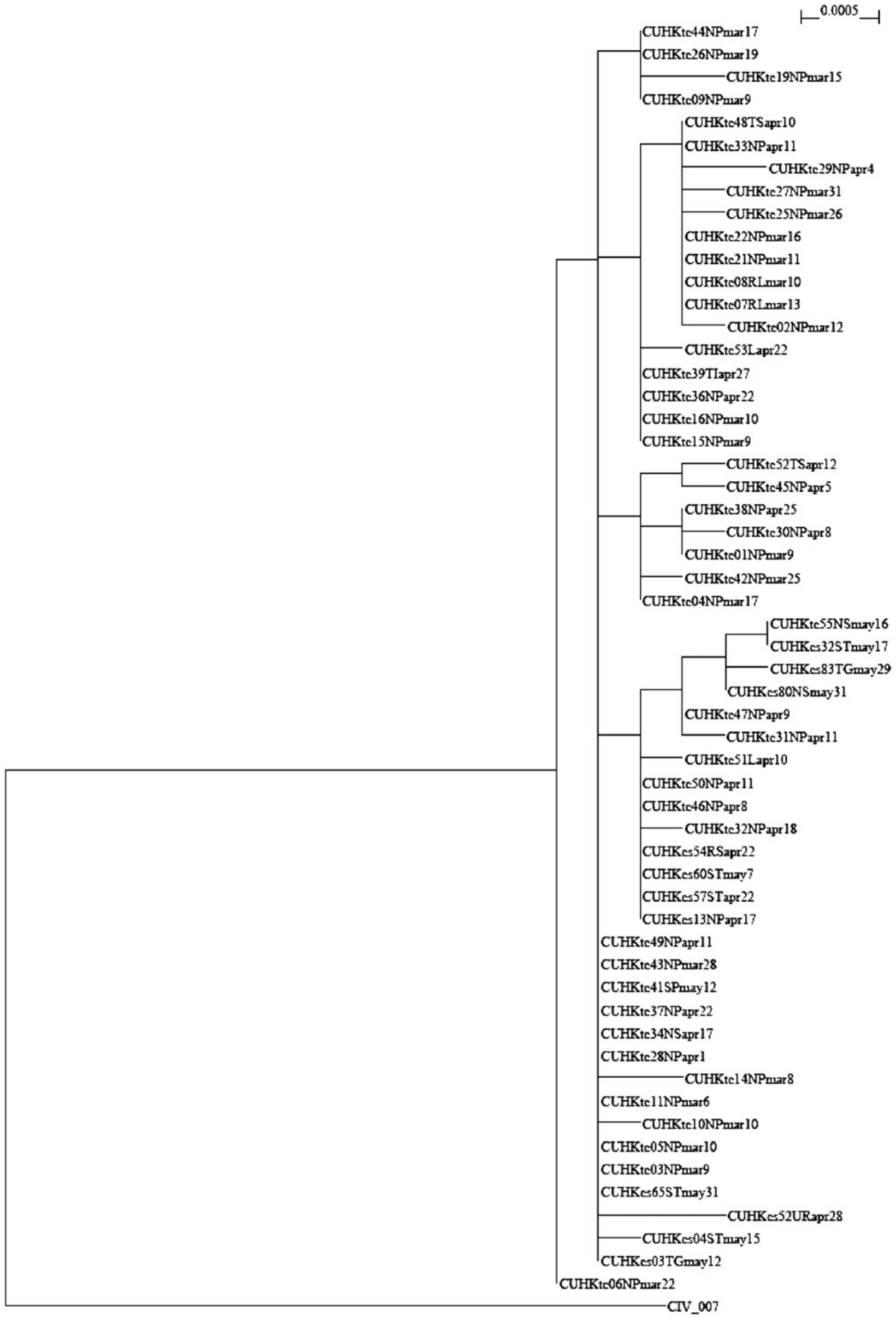
0.0005
CUHKtc44NPmar17
CUHKtc26NPmar19
CUHKtc19NPmar15
CUHKtc09NPmar9
CUHKtc48TSapr10
CUHKtc33NPapr11
CUHKtc29NPapr4
CUHKtc27NPmar31
CUHKtc25NPmar26
CUHKtc22NPmar16
CUHKtc21NPmar11
CUHKtc08RLmar10
CUHKtc07RLmar13
CUHKtc02NPmar12
CUHKtc53Lapr22
CUHKtc39TIapr27
CUHKtc36NPapr22
CUHKtc16NPmar10
CUHKtc15NPmar9
CUHKtc52TSapr12
CUHKtc45NPapr5
CUHKtc38NPapr25
CUHKtc30NPapr8
CUHKtc01NPmar9
CUHKtc42NPmar25
CUHKtc04NPmar17
CUHKtc55NSmay16
CUHKcs32STmay17
CUHKcs83TGmay29
CUHKcs80NSmay31
CUHKtc47NPapr9
CUHKtc31NPapr11
CUHKtc51Lapr10
CUHKtc50NPapr11
CUHKtc46NPapr8
CUHKtc32NPapr18
CUHKcs54RSapr22
CUHKcs60STmay7
CUHKcs57STapr22
CUHKcs13NPapr17
CUHKtc49NPapr11
CUHKtc43NPmar28
CUHKtc41SPmay12
CUHKtc37NPapr22
CUHKtc34NSapr17
CUHKtc28NPapr1
CUHKtc14NPmar8
CUHKtc11NPmar6
CUHKtc10NPmar10
CUHKtc05NPmar10
CUHKtc03NPmar9
CUHKcs65STmay31
CUHKcs52URapr28
CUHKcs04STmay15
CUHKcs03TGmay12
CUHKtc06NPmar22
CIV_007

Fig. 1. Maximum likelihood phylogram of the SARS-CoV S-gene from 56 Hong Kong patients. *Note*: Constructed under a Kimura three-parameters with unequal base frequencies (K81uf) model of evolution with invariable (I) sites and a gamma (G) distributed rate of substitution (i.e. K81uf + I + G), as selected by Modeltest (v3.7) under the Akaike Information Criteria (AIC), using a nearest-neighbor interchange (NNI) heuristic search strategy in PAUP*. The dates in the sample names are dates of fever onset for that patient. The scale for the branch lengths is indicated (number of nucleotide substitutions per site). There were no branches with bootstrap values >70. GenBank accession numbers for the 56 S-gene sequences: DQ412574–DQ412629. CUHK: Chinese University of Hong Kong; TC: tissue cultured isolate; CS: clinical sample; NP: nasopharyngeal aspirate; UR: urine; NS: nasal swab; SP: sputum; ST: stool; TS: throat swab; RS: rectal swab; TG: throat gargle; TI: terminal ileal biopsy; RL: right lung biopsy; L: lung biopsy.

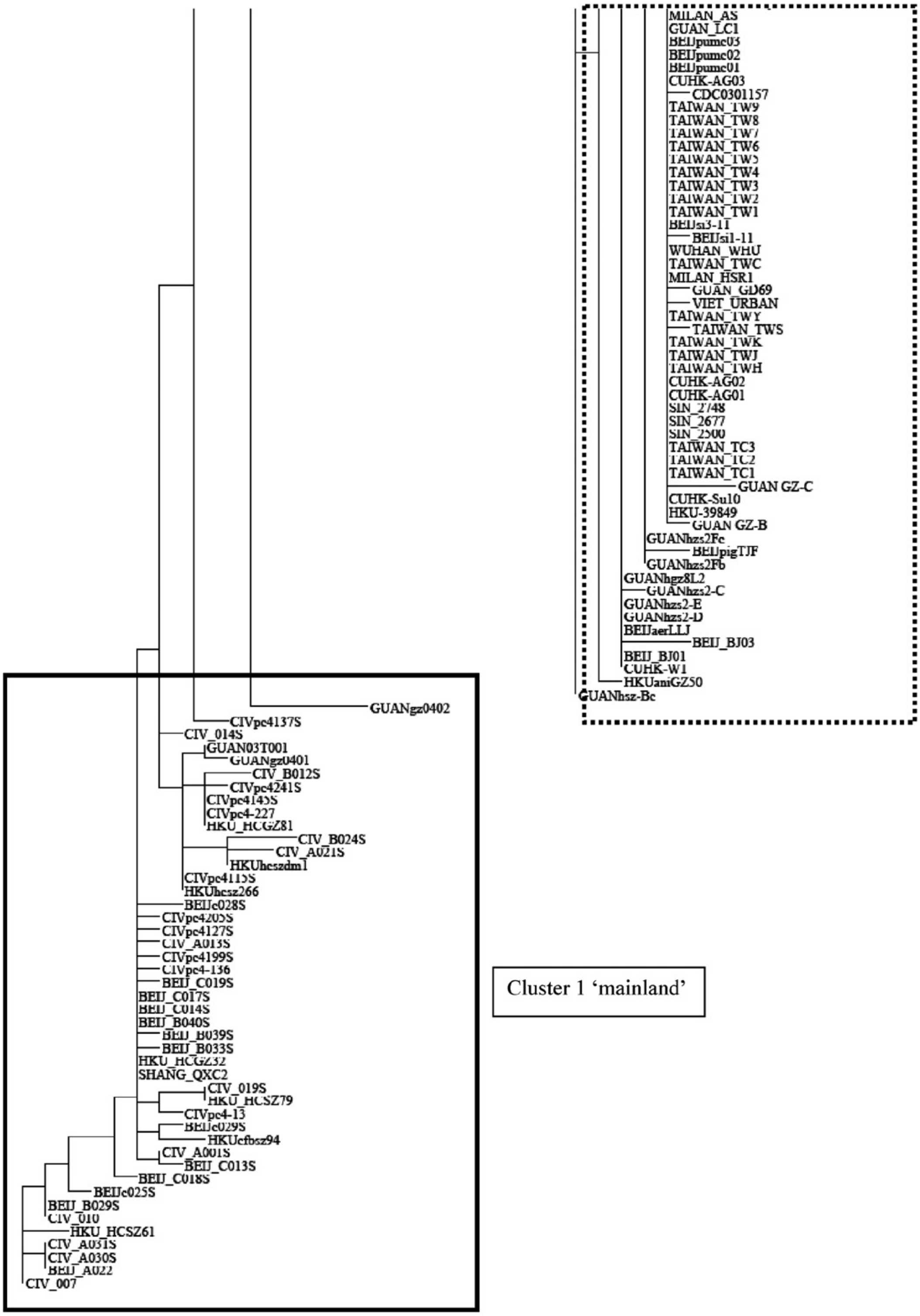
MILAN_AS
GUAN_LC1
BEIJpumc03
BEIJpumc02
BEIJpumc01
CUHK-AG03
CDC0301157
TAIWAN_TW9
TAIWAN_TW8
TAIWAN_TW7
TAIWAN_TW6
TAIWAN_TW5
TAIWAN_TW4
TAIWAN_TW3
TAIWAN_TW2
TAIWAN_TW1
BEIJsi3-11
BEIJsi1-11
WUHAN_WHU
TAIWAN_TWC
MILAN_HSR1
GUAN_GD69
VIET_URBAN
TAIWAN_TWY
TAIWAN_TWS
TAIWAN_TWK
TAIWAN_TWJ
TAIWAN_TWH
CUHK-AG02
CUHK-AG01
SIN_2748
SIN_2677
SIN_2500
TAIWAN_TC3
TAIWAN_TC2
TAIWAN_TC1
GUAN GZ-C
CUHK-Su10
HKU-39849
GUAN GZ-B
GUANhzs2Fc
BEIJpigTJF
GUANhzs2Fb
GUANhgz8L2
GUANhzs2-C
GUANhzs2-E
GUANhzs2-D
BEIJaerLLJ
BEIJ_BJ03
BEIJ_BJ01
CUHK-W1
HKUaniGZ50
GUANhsz-Bc
GUANgz0402
CIVpc4137S
CIV_014S
GUAN03T001
GUANgz0401
CIV_B012S
CIVpc4241S
CIVpc4145S
CIVpc4-227
HKU_HCGZ81
CIV_B024S
CIV_A021S
HKUhcszdm1
CIVpc4115S
HKUhcsz266
BEIJc028S
CIVpc4205S
CIVpc4127S
CIV_A013S
CIVpc4199S
CIVpc4-136
BEIJ_C019S
BEIJ_C017S
BEIJ_C014S
BEIJ_B040S
BEIJ_B039S
BEIJ_B033S
HKU_HCGZ32
SHANG_QXC2
CIV_019S
HKU_HCSZ79
CIVpc4-13
BEIJc029S
HKUcfbsz94
CIV_A001S
BEIJ_C013S
BEIJ_C018S
BEIJc025S
BEIJ_B029S
CIV_010
HKU_HCSZ61
CIV_A031S
CIV_A030S
BEIJ_A022
CIV_007
Cluster 1 ‘mainland’

Fig. 2. Maximum likelihood phylogram of the SARS-CoV S-gene from 56 Hong Kong samples and 138 downloaded S-gene sequences from GenBank. *Note*: Constructed under a Kimura three-parameters with unequal base frequencies (K81uf) model of evolution with invariable (*I*) sites and a gamma (*G*) distributed rate of substitution (i.e. $K81uf + I + G$), as selected by Modeltest (v3.7) under the Akaike Information Criteria (AIC), using a nearest-neighbor interchange (NNI) heuristic search strategy in PAUP*. The dates in the Hong Kong sample names are dates of fever onset for that patient, from whom the sample was taken. The scale for the branch lengths is indicated (number of nucleotide substitutions per site). The boxes highlight branches containing samples showing significant differences in their S-gene sequences from the rest of the group. Cluster 1: ‘mainland’ (solid-line box); cluster 2: ‘Guangdong’ (long-dashed line box); cluster 3: ‘worldwide’ (short-dotted line box). CUHK: Chinese University of Hong Kong; TC: tissue cultured isolate; CS: clinical sample; NP: nasopharyngeal aspirate; UR: urine; NS: nasal swab; SP: sputum; ST: stool; TS: throat swab; RS: rectal swab; TG: throat gargle; TI: terminal ileal biopsy; RL: right lung biopsy; L: lung biopsy. The names of the downloaded sequences from GenBank have been retained as far as possible, though some truncation has been applied in some cases.

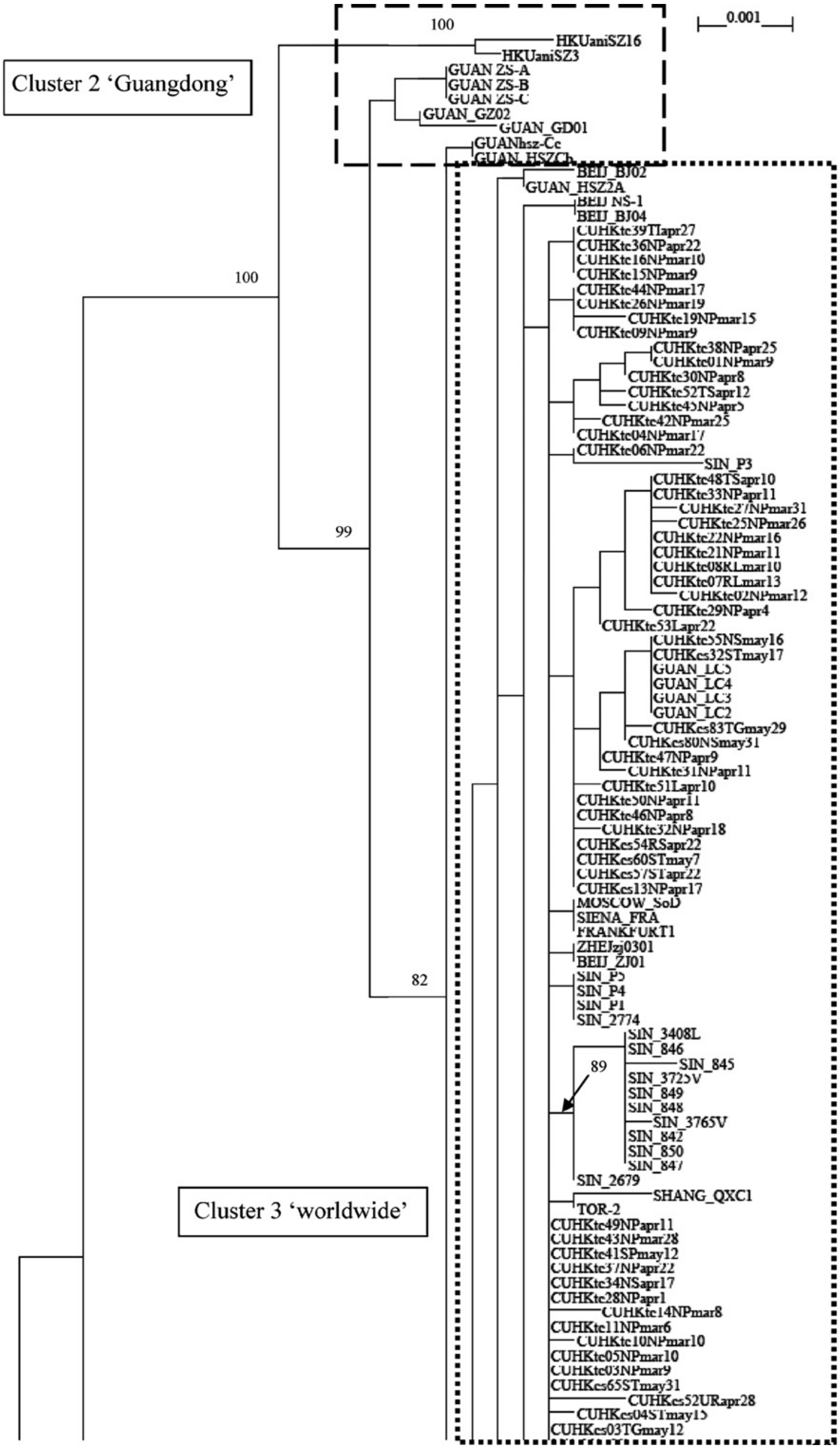
Cluster 2 ‘Guangdong’
100
0.001
HKUaniSZ16
HKUaniSZ3
GUAN ZS-A
GUAN ZS-B
GUAN ZS-C
GUAN_GZ02
GUAN_GD01
GUANhsz-Cc
GUAN_HSZCb
BEIJ_BJ02
GUAN_HSZ2A
BEIJ NS-1
BEIJ_BJ04
CUHKtc39TIapr27
CUHKtc36NPapr22
CUHKtc16NPmar10
CUHKtc15NPmar9
CUHKtc44NPmar17
CUHKtc26NPmar19
CUHKtc19NPmar15
CUHKtc09NPmar9
CUHKtc38NPapr25
CUHKtc01NPmar9
CUHKtc30NPapr8
CUHKtc52TSapr12
CUHKtc45NPapr5
CUHKtc42NPmar25
CUHKtc04NPmar17
CUHKtc06NPmar22
SIN_P3
100
CUHKtc48TSapr10
CUHKtc33NPapr11
CUHKtc27NPmar31
CUHKtc25NPmar26
CUHKtc22NPmar16
CUHKtc21NPmar11
CUHKtc08RLmar10
CUHKtc07RLmar13
CUHKtc02NPmar12
CUHKtc29NPapr4
99
CUHKtc53Lapr22
CUHKtc55NSmay16
CUHKes32STmay17
GUAN_LC5
GUAN_LC4
GUAN_LC3
GUAN_LC2
CUHKes83TGmay29
CUHKes80NSmay31
CUHKtc47NPapr9
CUHKtc31NPapr11
CUHKtc51Lapr10
CUHKtc50NPapr11
CUHKtc46NPapr8
CUHKtc32NPapr18
CUHKes54RSapr22
CUHKes60STmay7
CUHKes57STapr22
CUHKes13NPapr17
MOSCOW_SoD
SIENA_FRA
FRANKFURT1
ZHEJzj0301
BEIJ_ZJ01
82
SIN_P5
SIN_P4
SIN_P1
SIN_2774
SIN_3408L
SIN_846
89
SIN_845
SIN_3725V
SIN_849
SIN_848
SIN_3765V
SIN_842
SIN_850
SIN_847
SIN_2679
SHANG_QXC1
TOR-2
Cluster 3 ‘worldwide’
CUHKtc49NPapr11
CUHKtc43NPmar28
CUHKtc41SPmay12
CUHKtc37NPapr22
CUHKtc34NSapr17
CUHKtc28NPapr1
CUHKtc14NPmar8
CUHKtc11NPmar6
CUHKtc10NPmar10
CUHKtc05NPmar10
CUHKtc03NPmar9
CUHKes65STmay31
CUHKes52URapr28
CUHKes04STmay15
CUHKes03TGmay12

Fig. 2. (*Continued*).

Table 2
Single nucleotide variations (SNVs) present in the 56 CUHK SARS-CoV S-gene sequences

SNV position	Base change	Amino acid change	Total no. of samples with SNV (%)	No. of CS samples containing SNV (%)	No. of TC samples containing SNV (%)
21,615	T to A	Tyrosine to stop	21 (37.5)	2 (9.5)	19 (90.5)
21,901	T to A	Phenylalanine to isoleucine	18 (32.1)	0 (0)	18 (100)
23,220	G to T	Alanine to serine	56 (100)	11 (100)	45 (100)
25,114	C to T	Alanine to valine	17 (30.4)	7 (63.6)	10 (22.2)

a gamma-distributed site substitution rate (K81uf + I + G), as selected by Modeltest under Akaike Information Criteria, using a nearest-neighbor interchange heuristic search strategy in PAUP*.

Fig. 1 shows the S-gene sequences obtained from the 56 Hong Kong patients, rooted against CIV007. The dates in the sample names are dates of fever onset for that patient. Despite the apparent multiple branching topology, there are no significant branches (all bootstrap values < 70). The homology of these CUHK 56 S-gene sequences was 99–100%

Fig. 2 shows the position of these 56 sequences within the other 138 S-gene sequences downloaded from GenBank. Only bootstrap values > 70 are shown. Three distinct clusters can be seen, the sequences of which have been characterized and discussed elsewhere (CSMEC, 2004; Kan et al., 2005; Lan et al., 2005; Song et al., 2005; Wang et al., 2005; Yeh et al., 2004). All the CUHK S-gene sequence lie within cluster 3 'worldwide', together with other S-gene sequences from Hong Kong (CUHK-W1, AY278554; CUHK-AG01, AY345986; CUHK-AG02, AY345987; CUHK-AG03, AY345988; CUHK-Su10, AY282752; HKU-39849, AY278491).

When aligned and compared with the TOR-2 reference strain (AY274119), four single nucleotide variations (SNVs) were found in a large proportion of the CUHK S-gene sequences, as shown in Table 2.

4. Discussion

Since the end of the 2003 SARS epidemic, SARS-CoV has only reappeared a few times, sporadically, and only affecting a few people (WHO, 2003; Lim et al., 2004; WHO, 2004a,b). No further SARS epidemics have been identified. Therefore, the 56 samples analysed in this study, collected over a 3-month period during the SARS epidemic in Hong Kong (March–June 2003), represent one of the longest continuous periods over which SARS-CoV has been investigated for any viral mutation and evolution within a human population.

As shown in Fig. 1, these 56 CUHK sequences are all very similar, lying in Fig. 2's cluster 3 'worldwide' S-gene sequences. The pattern in Fig. 2 shows that SARS-CoV S-gene sequences isolated from some Guangdong patients and animals from Beijing and Guangdong cluster together (cluster 1 'mainland'). This correlates well with the early phase of the SARS epidemic, as has been defined previously (CSMEC, 2004; Song et al., 2005). Next, there is a transitional region seen, most likely to represent the spread of the SARS epidemic within Guangdong, and then out to Beijing and Hong Kong, then the rest of the world (cluster 2 'Guangdong'). This corresponds well with the late-early and middle phase of the SARS epidemic (CSMEC, 2004; Ruan et al., 2003; Song et al., 2005). Finally, the S-gene sequences from this study together with those obtained from outside Hong Kong and China (cluster 3 'worldwide'), are representative of the middle-to-late phase of the SARS epidemics (CSMEC, 2004).

It has been well established for almost 20 years that the ML approach is more accurate for reconstructing tree phylogenies, though the NJ approach is still widely used as it is much less demanding in terms of computational power and time (Hillis et al., 1994; Huelsenbeck and Hillis, 1993; Page and Holmes, 1998; Saitou, 1988; Saitou and Imanishi, 1989). Therefore, it is reassuring that this more complex and demanding ML approach, using optimal models of evolution as described here, with these complete S-gene sequences, independently supports previous NJ analysis of the epidemiology of the 2003 SARS outbreaks (e.g. the Kimura two-parameter distance method used by the CSMEC, 2004, Fig. 6). Due to the availability of more data, Fig. 2 contains more sequences spread over a longer time-frame, and the clusters appear more clearly defined, than the similar figure presented by the CSMEC (2004, Fig. 6).

It is possible that using tissue cultured (TC) isolates for S-gene sequence analysis may have introduced some new mutations, not otherwise present in the original clinical samples (CS). Poon et al. (2005a,b) reported non-synonymous mutations occurring in the S, membrane and ORF 8a protein coding genes, during the passaging of SARS-CoV through non-human primate cell lines (FRhK4 and Vero E6). Within the S-gene region they found mutations at positions (with respect to the TOR-2): C23412T, A23473G, C23518T, C23632T, C24864T, and A24978G. Tong et al. (2004) also reported two tissue-culture (Vero E6)-associated mutations (with reference to the SARS-CoV Urbani strain): T21938C and T24872C. In this study, we did not find any of these previously reported mutations, however, there were mutations found mainly in the TC samples (T21615A and T21901A) that have not been previously reported elsewhere. These are most likely to be tissue-culture associated mutations. It is interesting that the T21615A mutation results in a stop codon, yet still manages to produce a viable virus in tissue culture.

Several studies have documented various SNVs in the SARS-CoV genome, with some even proposing them as the basis for a genotypic classification (Poon et al., 2005a,b; Ruan et al., 2003; Tsui et al., 2003; Wang et al., 2005; Yeh et al., 2004). In this study, the SNV G23220T that was found in all of the patients (56/56, 100%) has been reported elsewhere (Yeh et al., 2004), though the functional significance of this is presently uncertain. The 100% prevalence of this SNV in these samples strongly suggests that a non-TOR-2-like strain of SARS-CoV has infected these Hong Kong patients, rather than this being an example of natural selection or of SARS-CoV behaving as a quasispecies.

The second, less frequently found SNV (C25114T) present in 17/56 (30.4%) of these S-gene sequences has also been found in four sequences from Guangdong (GUAN_LC2, AY394999; GUAN_LC3, AY395000; GUAN_LC4, AY395001; GUAN_LC5, AY395002, but not in GUAN_LC1, AY394998), of which GUAN_LC3 and GUAN_LC5 have been previously reported by Wang et al. (2005). The 17 CUHK S-gene sequences containing this polymorphism all occur in samples collected towards the end of the SARS outbreaks in Hong Kong (from 4 April to 31 May 2003). Unfortunately, we do not know on what date the GUAN_LC2/LC3/LC4/LC5 samples were taken (they are unpublished sequences that were submitted to GenBank on 19 September 2003), and it is therefore difficult to ascertain whether this SNV represents ongoing host-adaptive evolution. All of these 56 CUHK samples came from patients admitted to Hong Kong's New Territories East hospital cluster during this period. The 17/56 patients carrying this C25114T mutation may have acquired their virus via a chain of contacts infected with this particular virus, or this mutation occurred in all of these viruses independently, perhaps in response to some common, unknown selection pressure. Therefore, without a more accurate epidemic history for this particular virus, it is hard to say whether this is a founder effect or natural selection acting upon the virus towards the end of the Hong Kong 2003 SARS outbreaks.

In summary, the analysis of 56 complete SARS-CoV S-gene sequences from Hong Kong in this study using ML analysis confirms the previous epidemiology of SARS-CoV obtained using mostly NJ analysis. It also describes two new SNVs that are most likely to be associated with the tissue-culture passage of SARS-CoV.

Acknowledgement

The work described in the paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU7542/03M).

References

Chan PK, To WK, Ng KC, et al. Laboratory diagnosis of SARS. Emerg Infect Dis 2004;10(5):825–31.

Chim SS, Tong YK, Hung EC, Chiu RW, Lo YM. Genomic sequencing of a SARS coronavirus isolate that predated the Metropole Hotel case cluster in Hong Kong. Clin Chem 2004;50(1):231–3.

Chinese, SARS Molecular Epidemiology Consortium. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 2004;303(5664):1666–9.

Guan Y, Peiris JS, Zheng B, et al. Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome. Lancet 2004;363(9403):99–104.

Guan Y, Zheng BJ, He YQ, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 2003;302(5643):276–8.

Hillis DM, Huelsenbeck JP, Cunningham CW. Application and accuracy of molecular phylogenies. Science 1994;264:671–7.

Huelsenbeck JP, Hillis DM. Success of phylogenetic methods in the four-taxon case. Syst Biol 1993;42:247–64.

Kan B, Wang M, Jing H, et al. Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. J Virol 2005;79(18):11892–900.

Lan YC, Liu TT, Yang JY, et al. Molecular epidemiology of severe acute respiratory syndrome-associated coronavirus infections in Taiwan. J Infect Dis 2005;191(9):1478–89.

Lau SK, Woo PC, Li KS, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci USA 2005;102(39):14040–5.

Li W, Shi Z, Yu M, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science 2005;310(5748):676–9.

Lim PL, Kurup A, Gopalakrishna G, et al. Laboratory acquired severe acute respiratory syndrome. N Engl J Med 2004;350:1740–5.

Marra MA, Jones SJ, Astell CR, et al. The genome sequence of the SARS-associated coronavirus. Science 2003;300(5624):1399–404.

Page RDM, Holmes EC. Inferring molecular phylogeny. In: Molecular evolution—a phylogenetic approach. Malden, MA, USA: Blackwells Science; 1998. p. 172–227.

Perrière G, Gouy M. WWW-query: an on-line retrieval system for biological sequence banks. Biochimie 1996;78:364–9, http://pbil.univ-lyon1.fr/software/njplot.html.

Poon LL, Chu DK, Chan KH, et al. Identification of a novel coronavirus in bats. J Virol 2005a;79(4):2001–9.

Poon LL, Leung CS, Chan KH, Yuen KY, Guan Y, Peiris JS. Recurrent mutations associated with isolation and passage of SARS coronavirus in cells from non-human primates. J Med Virol 2005b;76(4):435–40.

Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics 1998;14:817–8, http://darwin.uvigo.es/software/modeltest.html.

Ruan YJ, Wei CL, Ee AL, et al. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet 2003;361(9371):1779–85.

Saitou N, Imanishi T. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. Mol Biol Evol 1989;6:514–25.

Saitou N. Property and efficiency of the maximum likelihood method for molecular phylogeny. J Mol Evol 1988;27:261–73.

Song HD, Tu CC, Zhang GW, et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc Natl Acad Sci USA 2005;102(7):2430–5.

Swofford DL. *PAUP** phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, Massachusetts: Sinauer Associates; 2001 [http://paup.csit.fsu.edu/win.html].

Tippmann HF. Analysis for free: comparing programs for sequence analysis. Brief Bioinform 2004;5(1):82–7, http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

Tong S, Lingappa JR, Chen Q, et al. Direct sequencing of SARS-coronavirus S and N genes from clinical specimens shows limited variation. J Infect Dis 2004;190(6):1127–31.

Tsui SK, Chim SS, Lo YM. Chinese University of Hong Kong Molecular SARS Research Group. Coronavirus genomic-sequence variations and the epidemiology of the severe acute respiratory syndrome. N Engl J Med 2003;349(2):187–8.

Wang ZG, Zheng ZH, Shang L, et al. Molecular evolution and multilocus sequence typing of 145 strains of SARS-CoV. FEBS Lett 2005;579(22):4928–36.

WHO. Severe Acute Respiratory Syndrome (SARS) in Taiwan, China. 17 December 2003. Available from: http://www.who.int/csr/don/2003_12_17/en/.

WHO. Review of probable and laboratory-confirmed SARS cases in southern China. 27 January 2004a. Available from: http://www.who.int/csr/don/2004_01_27/en/.

WHO. SARS: one suspected case reported in China. 22 April 2004b. Available from: http://www.who.int/csr/don/2004_04_22/en/.

Yeh SH, Wang HY, Tsai CY, et al. Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: molecular epidemiology and genome evolution. Proc Natl Acad Sci USA 2004;101(8):2542–7.

Zhao Z, Li H, Wu X, et al. Moderate mutation rate in the SARS coronavirus genome and its implications. BMC Evol Biol 2004;4:21.