Available online at www.sciencedirect.com
SCIENCE DIRECT
Infection, Genetics and Evolution 5 (2005) 261–269
Infection, Genetics and Evolution
www.elsevier.com/locate/meegid

Phylogenetic analysis and sequence comparisons of structural and non-structural SARS coronavirus proteins in Taiwan

Yu-Ching Lan[a], Hsin-Fu Liu[b], Yi-Ping Shih[a], Jyh-Yuan Yang[c], Hour-Young Chen[c], Yi-Ming Arthur Chen[a,*]

[a]*AIDS Prevention and Research Center, Institute of Public Health, National Yang-Ming University, Taipei 112, Taiwan*
[b]*Department of Medical Research, Mackay Memorial Hospital, Taipei, Taiwan*
[c]*Center for Disease Control, Department of Health, Executive Yuan, Taipei 100, Taiwan*

Received 14 January 2004; accepted 7 August 2004
Available online 7 December 2004

Abstract

Taiwan experienced a large number of severe acute respiratory syndrome (SARS) viral infections between March and July 2003; by September of that year, 346 SARS cases were confirmed by RT-PCR or serological tests. In order to better understand evolutionary relationships among SARS coronaviruses (SCoVs) from different international regions, we performed phylogenetic comparisons of full-length genomic and protein sequences from 45 human SCoVs (including 12 from Taiwan) and two civet SCoVs. All the Taiwanese SARS-CoV strains which associated with nosocomial infection formed a monophyletic clade within the late phase of the SARS epidemic. This Taiwanese clade could be further divided into two epidemic waves. Taiwan SCoVs in the first wave clustered with three isolates from the Amoy Gardens housing complex in Hong Kong indicating their possible origin. Of the 45 human SCoVs, one isolate from Guangdong province, China, exhibited an extra 29-nucleotide fragment between Orf 10 and Orf 11—similar to the civet SCoV genome. Nucleotide and protein sequence comparisons suggested that all SCoVs of late epidemic came from human-to-human transmission, while certain SCoVs of early epidemic might have originated in animals.

Keywords: Severe acute respiratory syndrome; SARS coronavirus; SCoV; Civet; Phylogenetic analysis; Taiwan; Orf 3; Orf 10

1. Introduction

On 7 August 2003, the World Health Organization (WHO, 2003) reported that the 2003 SARS pandemic infection had spread to more than 30 countries, affecting 8422 people and killing 916. Later that year a novel coronavirus (SARS-CoV) was isolated from SARS patients (Drosten et al., 2003; Ksiazek et al., 2003; Peiris et al., 2003b; Poutanen et al., 2003); an animal inoculation experiment identified a causal relationship between SARS and SARS-CoV infection (Fouchier et al., 2003).

Zhong et al. (2003) identified the geographic origin of the epidemic as Guangdong province, China, and the originating month as November 2002. The first SARS case in Taiwan was diagnosed on 14 March 2003. Its history was traced to a trip by the index case to Guangdong in mid-February, when the SARS epidemic in that province reached its peak (CDC, 2003; ROC CDC, 2003). The index case transmitted the virus to his wife and son; the first SARS coronavirus in Taiwan—SCoV TW1—was isolated from the son (Hsueh et al., 2003).

On 26 March a male resident of the Amoy Gardens housing complex in Hong Kong (hereafter referred to as Mr. X) flew to Taiwan. On 27 March he took a train from Taipei to Taichung City to visit his younger brother. That night he experienced a high fever; most likely he also read a local news report of a major SARS outbreak in Amoy Gardens that same day (Peiris et al., 2003a). He returned to Hong Kong on 28 March. After he was admitted to a hospital, Mr.

* Corresponding author. Tel.: +886 2 28267193; fax: +886 2 28270576.
E-mail address: arthur@ym.edu.tw (Y.-M. Chen).

1567-1348/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2004.08.005

X made a phone call to warn his younger brother, but it was too late. The younger brother who developed symptoms on 31 March became the first SARS-related fatality (TC1) in Taiwan. A second index case (SCoV-TWC) was isolated from this patient (the younger brother) by the ROC CDC (Chen et al., 2003).

The third index case was Ms. A, a female adult who traveled on the same Taiwanese train as Mr. X on 27 March. Two days later she visited a hospital in Taipei complaining of general fatigue. In addition to the local hospital, she visited two private clinics before being referred to Taipei Municipal Hoping hospital on 9 April. She spent less than 6 h in that hospital's emergency room, but she probably transmitted the virus to two patients, an assistant nurse who escorted her to the X-ray room, and a laundry worker who handled her isolation gown. These individuals transmitted the SARS virus to other medical personnel and patients, resulting in the entire hospital being shut down for more than 2 months starting on 24 April. According to the ROC CDC, the Hoping hospital nosocomial infection resulted in 66 probable and 22 suspected SARS cases (Wu et al., 2003). Even though the Taiwanese government imposed a quarantine on 28 April on all air travelers arriving from China, Hong Kong, Singapore, Macau, or Toronto, the virus still spread to different parts of the main island of Taiwan and the adjacent Penghu Islands. By 1 September, 346 SARS cases in Taiwan had been confirmed by RT-PCR or serological tests (WHO, 2003).

The size of the SCoV genome is approximately 29.7 kb (Marra et al., 2003; Rota et al., 2003). The 5′ portion of the genome (21 kb, about two-thirds) contains the code for the replicase gene, including two large open reading frames (Orfs), referred to as Orfs 1a and 1b. The other one-third of the genome contains Orfs for four structural proteins (spike [S], envelope [E], membrane [M], and nucleocapsid [N]) and nine putative non-structural proteins (Orfs 3, 4, 7, 8, 9, 10, 11, 13 and 14). Recently, Guan et al. (2003) isolated SCoV-like viruses from Himalayan palm civets and raccoon dogs in southern China. According to a comparative analysis of human and animal SCoV genomes, the three animal SCoVs (SZ1, SZ13 and SZ16) all retain a 29-nucleotide sequence inserted between Orfs 10 and 11.

For this study, we used phylogenetic analysis to investigate relationship among 12 Taiwanese SARS-CoVs and between those SCoVs from other countries. One specific goal was to determine whether the SARS-CoV isolate from Mr. X's younger brother (TWC) clustered with the isolate from Ms. A (TWC2), and whether either one of those isolates clustered with isolates from other Amoy Gardens residents (Chim et al., 2003). We also compared the amino acid sequences of the S, E, M, and N structural proteins and three of the nine putative non-structural proteins (Orfs 3, 10, and 11) for 47 SARS-CoVs including 12 Taiwanese strains.

2. Materials and methods

2.1. SCoV strains and their origins

Twelve Taiwanese SCoV strains were included in this study: TW1 (Hsueh et al., 2003), TWC, TWC2, TWC3, TWH, TC1, TC2, TC3, TWJ, TWK, TWS and TWY. TW1 was isolated from a patient whose father spent time in Guangdong province in mid-February 2003. TWC was isolated from Taiwan's first SARS-related fatality. TWC2 and TWC3 were isolated from Taipei Municipal Hoping hospital patients, and TWC3 was from Ms. A, the third index case. An additional 33 full-length genomic sequences from human SCoV strains were selected from the GenBank: nine from Beijing (BJ01, BJ02, BJ03, BJ04, PUMC01, PUMC02, PUMC03, Sino3-11 and Sino1-11), six from Hong Kong (CUHK-W1, CUHK-AG03, CUHK-AG02, CUHK-AG01, CUHK-Su10 and HKU-39849), five from Singapore (Sin2679, Sin2677, Sin2500, Sin2774 and Sin2748), two from Guangzhou (GD01 and GZ50), two from Frankfurt (Frankfurt1 and FRA), two from Milan (AS and HSR1), two from Guangdong province (ZMY1 and GD69), and one each from Wuhan (WHU), Zhejiang province (ZJ01), Moscow (SoD), Toronto (TOR2), and Hanoi (Urbani).

2.2. Phylogenetic tree analysis

A BLAST search was performed to locate SARS CoV sequences in the GenBank database. A total of 47 full-length nucleotide sequences from SARS CoV isolates (including two civet isolates) were aligned and edited using the BioEdit program (Hall, 1999). Phylogenetic analyses were conducted with the Phylip 3.6b (Felsenstein, 1989) and MEGA2 programs (Kumar et al., 2001) using the neighbor-joining (NJ) and Fitch and Wagner parsimony (Pars) methods. Evolutionary distances were estimated with the Kimura two-parameter model (Kimura, 1980). NJ and Pars tree robustness were statistically evaluated by bootstrap analysis (100 samples).

2.3. Nucleotide sequence comparisons, sequence alignment, and amino acid sequence comparisons

SCoV nucleotide sequence variation was analyzed with the SIMPLOT program (Johns Hopkins University, Baltimore, MD). The 20 SCoVs used for this task were the Urbani, CUHK-W1, TOR2, HKU-39849, BJ01, BJ02, BJ03, BJ04, GD01, TW1, TWC, SIN2774, SIN2748, SIN2679, SIN2677, SIN2500, HSR1, CUHK-Su10, Frankfurt1, and GZ50. Two civet SCoVs (SZ3 and SZ16) were used as references for comparison. Sequence variation distance plots were generated with 1000 bp windows, 100 bp steps, and a Jukes-Cantor correction. Nucleotide sequences for the four structural genes, Orf 3, and Orf 10 were edited and translated into amino acid sequences using the BioEdit program prior to alignment for comparisons.

2.4. Nucleotide sequence accession numbers

The accession numbers for the SCoVs used in this study are Urbani: AY278741; CUHK-W1: AY278554; TOR2: AY274119; HKU-39849: AY278491; BJ01: AY278488; BJ02: AY278487; BJ03: AY278490; BJ04: AY279354; GD01-GZ01: AY278489; TW1: AY291451; TWC: AY321118; SIN2774: AY283798; SIN2748: AY283797; SIN2679: AY283796; SIN2677: AY283795; SIN2500: AY283794; HSR1: AY323977; CUHK-Su10: AY282752; Frankfurt1: AY291315; GZ50: AY304495; SZ3: AY304495; SZ16: AY304488; SoD: AY461660; FRA: AY310120; WHU: AY394850; ZJ01: AY297028; AS: AY427439; ZMY1: AY351680; GD69: AY313906; Sino3-11: AY485278; PUMC01: AY350750; PUMC02: AY357075; PUMC03: AY357076; Sino1-11: AY485277; CUHK-AG03: AY345988; CUHK-AG02: AY345987; CUHK-AG01: AY345986; TWC2: AY362698; TWC3: AY362699; TWH: AP006557; TC1: AY338174; TWJ: AP006558; TWK: AP006559; TWS: AP006560; TWY: AP006561; TC2: AY338175; TC3: AY348314. Two civet SCoVs: SZ-3: AY304495; SZ16: AY304488.

3. Results

To better understand evolutionary relationships between SCoVs isolated in Taiwan and those isolated in other parts of the world, we constructed phylogenetic trees with two different methods using full-length genomic sequences from 45 human (12 Taiwanese) and two civet SCoVs. Tree topologies were consistent for the NJ (Fig. 1a) and Pars (Fig. 1b) methods. Two human SCoV epidemics were identified. The late epidemic SCoVs formed a well-supported monophyletic clade with bootstrap values of 98 and 88 for the NJ and Pars trees, respectively. The early epidemic sequences did not cluster into a monophyletic clade, even though they did clearly differed from those of late epidemic.

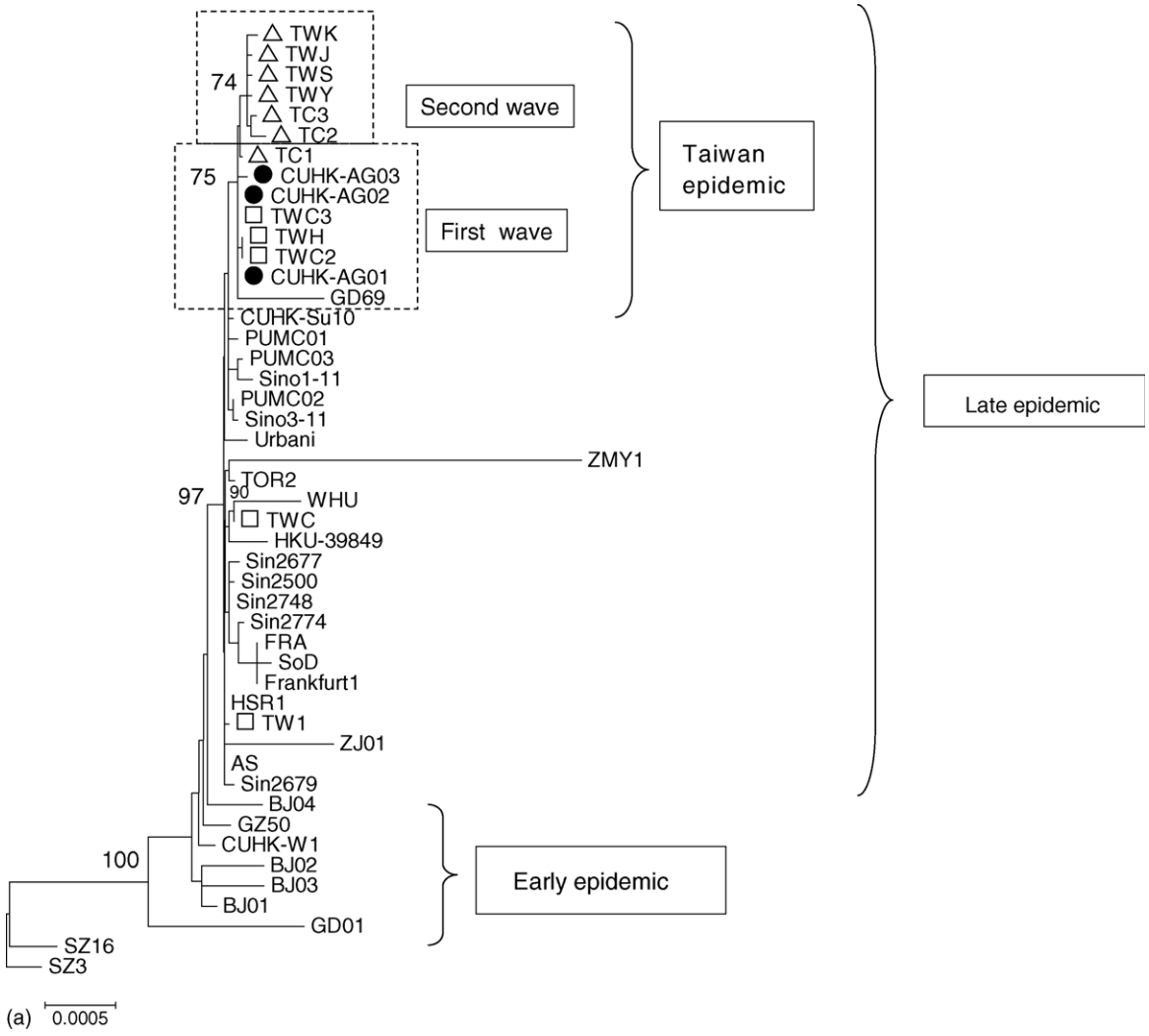


Fig. 1. Human and civet SCoV phylogenetic trees, produced with the neighbor-joining (NJ) method using full-length (29.7 kb) sequences. Branch bootstrap values from 100 reps: (a) using the SZ3 civet SCoV as a root; (b) a tree produced using the parsimony (Pars) method.

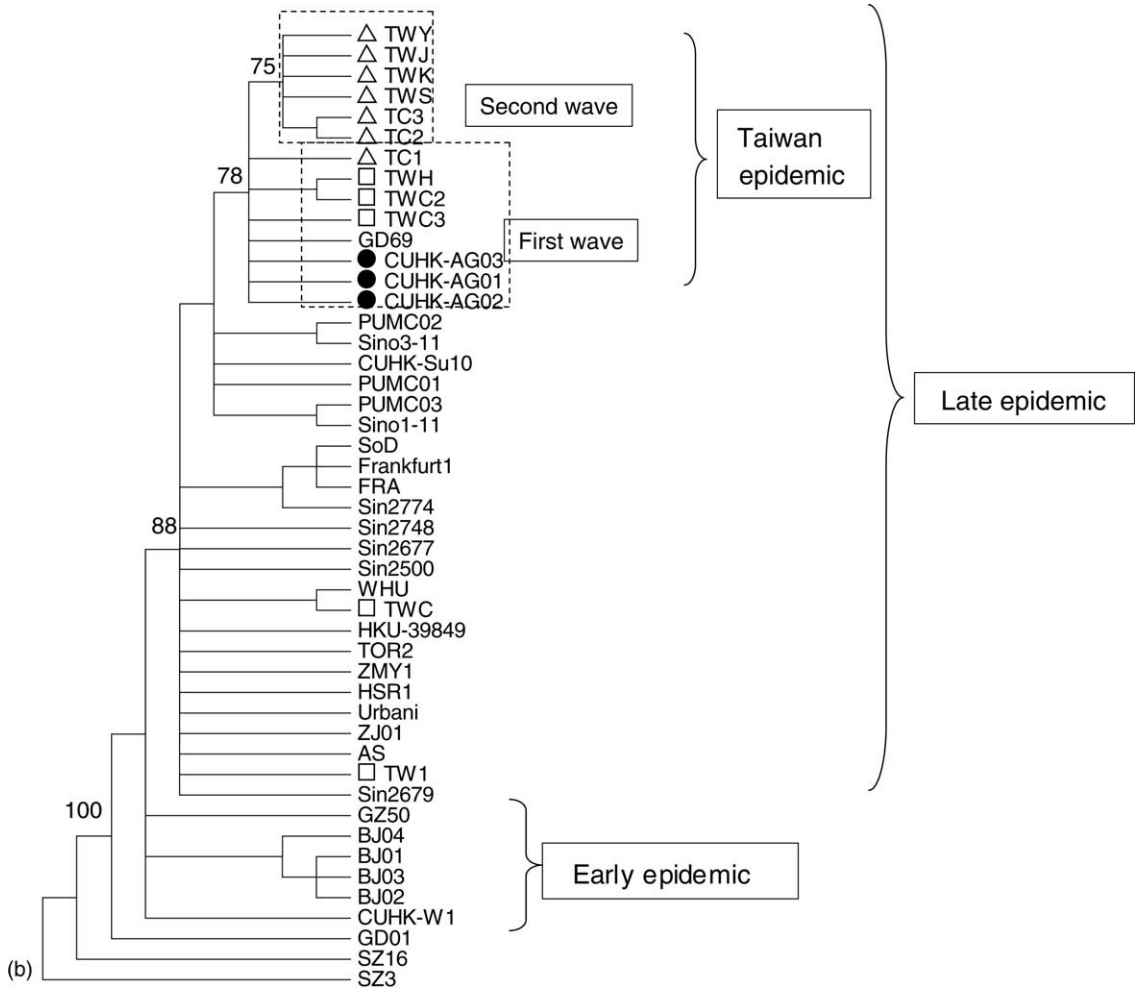


Fig. 1. (*Continued*).

All early epidemic SCoVs had Chinese origins: Beijing (BJ01, BJ02, BJ03 and BJ04), Guangzhou (GD01 and GZ50), and Hong Kong (CUHK-W1).

All the Taiwanese SCoVs sequences which associated with nosocomial infection clustered into a monophyletic clade (bootstrap values 75 and 78 for NJ and Pars trees, respectively) within the late epidemic and could be further classified into two epidemic waves. Second wave was a monophyletic clade supported by bootstrap values of 74 and 75 for NJ and Pars tree, respectively, while first wave was not a fully resolved cluster. TWC (from Mr. X's younger brother) did not cluster with three isolates from Amoy Gardens (CUHK-AG03, CUHK-AG02, CUHK-AG01), but did cluster with an isolate (WHU) from Wuhan, China (bootstrap value 90 for NJ tree) (Fig. 1a).

Pairwise comparison methods were used to analyze nucleotide sequence variation within the full-length genomes of 20 human SCoVs (7 from early epidemic and 13 from late epidemic) (Fig. 2). Two civet SCoV sequences (SZ3 and SZ16) were used as references for comparison.

Table 1
Amino acid sequence variation rates for spike (S1 and S2), envelope (E), membrane (M), nucleocapsid (N) and Orf 3 of 45 human SCoVs compared with civet SCoVs

Human SCoV	Variation rate	S1 (664) (%)	S2 (575) (%)	E (76) (%)	M (221) (%)	N (422) (%)	Orf 3 (274) (%)
Early epidemic ($n = 7$)	With civet SCoV	1.56	1.09	0	0.52	0.07	1.68
	Intra-group	0.24	0.20	0	0.13	0.14	0.38
Late epidemic ($n = 38$)	With civet SCoV	1.73	1.68	0.21	0.80	0.03	1.52
	Intra-group	0.07	0.05	0.41	0.45	0.06	0.10
Total	With civet SCoV	1.70	1.06	0.18	0.75	0.04	1.53

The S1 and S2 domains contain the amino acid residues 17–680 and 681–1255 of the SCoV spike protein. Numbers in parentheses indicate domain lengths.

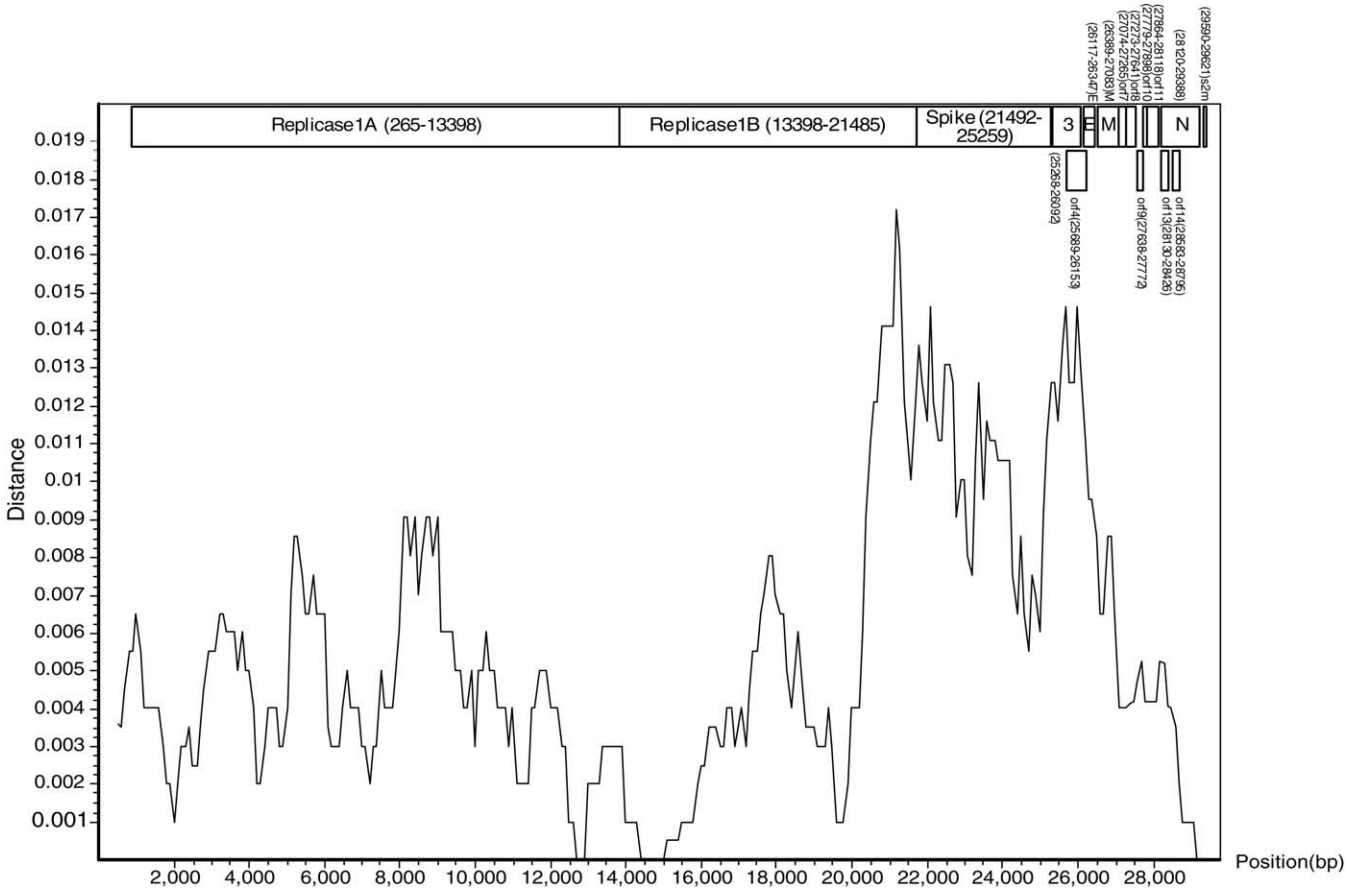


Fig. 2. Plot analyses were used to compare diversity distributions among genes from 20 human SCoVs. The average genetic distance from the reference genome of civet SCoVs of 20 human SCoVs are plotted over the entire genome of SCoV. Genomic sequences from the SZ3 and SZ16 civet SCoVs were used as references. The *X*-axis is the nucleotide location of the SARS-CoV genome. The *Y*-axis is the rate of nucleotide differences between 20 human SCoVs and civet SCoVs. Sequence variation distance plots were generated with a 1000 bp window, 100 bp steps by Simplot program.

Our results revealed that the highest variation rate was in the 3′ one-third of the viral genome, especially the nucleotide sequences near the junction between the replicase 1B and spike genes; Orf 3 also had a relatively high sequence variation rate.

Amino acid sequences for the S, M, E, and N structural proteins of 45 human SCoVs were compared with those of the SZ-16 civet SCoV (Fig. 3). The S protein was divided into S1 and S2 domains according to the molecular model proposed by Spiga et al. (2003). The S1 domain (*N*-terminal 17–680 amino acid residues, responsible for receptor-binding) had 18 (2.7%) amino acid differences; the S2 domain (681–1255 amino acid residues) had 11 (1.9%)—a total of 29 (2.3%) differences in the S proteins of 43 SCoVs. The S genes of WHU and ZMY1 contained several nucleotide insertions that interrupted the open reading frames. The amino acid distances of S proteins were 1.3% (16.4/1239) for early epidemic SCoVs and 1.4% (17.2/1239) for late epidemic SCoVs in comparison with civet SCoVs. Intra-group sequence variation for early epidemic was 0.3% ($n = 7$) and for late epidemic 0.09% ($n = 38$) (Table 1). The numbers of amino acid differences were 4 for the E protein (5.3%), 7 for M (3.2%), 4 for N (0.9%), and 11 for Orf 3 (4.0%) (Fig. 3). Amino acid distances among the 45 human SCoVs were 0.18% (0.13/76) for the E protein, 0.75% (1.67/221) for M, 0.04% (0.16/422) for N, and 1.53% (4.20/274) for Orf 3 (Table 1 and Fig. 4).

Among the 45 human SCoVs that we analyzed, an isolate (GD01) from Guangdong province China, contained an extra 29-nucleotide fragment. Both WHU and TWC had di-nucleotide deletions at the 30th and 31st nucleotides of Orf 10, resulting in a frame shift and premature stop of the putative protein (Fig. 5). In addition, we observed a 5-nucleotide deletion at the 32nd nucleotide of Orf 10 in Sin2748; this also resulted in a frame shift and premature translation stop.

4. Discussion

Both the NJ and Pars trees separated the human SCoVs into two epidemics, even though early epidemic SCoVs failed to cluster into a well-supported monophyletic clade (Fig. 1a and b). The early epidemic sequences were more closely related than the late epidemic sequences to civet SCoVs; all seven early epidemic SCoVs were from either

Protein		Spike 0.023 (29/1255)				E 0.053 (4/76)	M 0.032 (7/221)	N 0.009 (4/422)
		S1 0.027 (18/664)		S2 0.019(11/575)				
Amino Acid Site					111			
		12222333	44455566	77	777888011		1	114
		4742346146	47805706	04	579669046	24	126795	5591
		9747941140	29718775	13	484014183	5645	5178394	0430
Civet Cat	SZ16-ref	SDMKLTKGRS	YKSFISPS	LA	VDPVSARLE	VSVN	SEFANFS	TNGN
Early epidemic	GD01	...N.TRF	.NTYF.SL	ST	A.S..T..K	....	G C....	...D
	GZ50	L..NS.T.KF	.NT.F.SL	ST	A....T..K	....	G.....	....
	BJ03	..LNS.T.KF	.NT.F.SL	ST	AY.LRT..K	....	G.....	.Y..
	BJ02	...NS.TRKF	.NT.F.SL	ST	AY...T..K	....	G.....	....
	[a]BJ01	...NS.T.KF	.NT.F.SL	ST	AY...T..K	....	G.....	....
	BJ04	.GNS.T.KF	.NT.F.SL	ST	AY...TMK	....	G.....	....
Late epidemic	[b]SoD	.GNSIT.KF	.NT.F.SL	ST	AY...T.FK	....	G.V...	I...
	[c]TWC	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G.V...	....
	WHU					....	G.V.L.	....
	ZJ01	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G.....	
	ZMY1					....	G.....	
	TOR2	.GNSIT.KF	.NT.FASL	ST	AY...T..K	....	G.....	....
	[d]TW1	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G.....	....
	Sin2774	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G.....	I...
	Sin2500	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	GK....	....
	Urbani	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G....P	....
	GD69	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	..M	G.....	....
	CUHK-Su10	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G C....	..C.
	[e]PUMC02	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	F...	G C....	....
	PUMC01	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	.L..	G C....	....
	PUMC03	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	...Y	G C....	....
	Sino1-11	.GNSIT.KF	SNT.F.SL	ST	AY...T..K	...Y	G C....	....
	[f]CUHK-AG01	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G C....	....
Taiwan 1st wave	[g]TWC3	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G C....	....
Taiwan 2nd wave	[h]TWJ	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G C....	....
	TC3	.GNSIT.KF	.NT.F.SL	ST	AY...T..K	....	G C.T..	....

a. CUHK-W1 also had the identical sequence variation.
b. other strains with identical sequences include Frankfurt and FRA.
c. HKU-39849 also had the identical sequence variation.
d. other strains with identical sequences include AS, HSR1, Sin2677, Sin2679 and Sin2748.
e. Sino3-11 also had the identical sequence variation.
f. other Hong Kong Amoy Garden strains with identical sequences include CUHK-AG02 and CUHK-AG03.
g. other Taiwan group1 strains with identical sequences include TWC2, TWH and TC1.
h. other Taiwan group2 strains with identical sequences include TWK, TWS, TWY and TC2.

Fig. 3. Amino acid comparisons: S, SCoV spike; E, envelope; M, membrane; N, nucleocapsid. The civet cat SZ16 SARS-CoV was used as a reference. A period (.) indicates concurrence with the top reference sequence (SZ16) in the alignment.

Guangdong province or Beijing. Among all the analyzed human SCoVs, GD01 was the only one having an extra 29-nucleotide fragment which was also found in the civet SCoVs (Guan et al., 2003). Furthermore, the average intra-group amino acid distance for the S gene in early epidemic was higher than for late epidemic (Table 1). We also identified a signature amino acid sequence pattern (amino acid residues 77 and 244; Fig. 6) shared by early epidemic isolates and civet SCoVs. These evidences suggested that late epidemic SCoVs were transmitted from human-to-human, while certain early epidemic SCoVs (e.g., GD01) might have been transmitted from animals to humans before spreading among various human populations.

Among the Taiwanese SCoVs, our phylogenetic analysis does not support the hypothesis of an epidemiological link between the first and third index cases (Mr. X and Ms. A). According to our NJ tree, TWC (a SCoV isolate from Mr. X's younger brother) clustered with the WHU SCoV from Wuhan, China (bootstrap = 92), while TWC-3 (Ms. A's isolate) clustered with CUHK-AG02 and CUHK-AG03, both of which originated in Hong Kong's Amoy Gardens housing complex. A sequence analysis demonstrated that TWC and WHU had di-nucleotide deletions in Orf 10, resulting in a shift in the open reading frame (Fig. 5). Therefore, even though Mr. X and Ms. A took the same train from Taipei to Taichung, the evidence indicates that Mr. X was not the source of Ms. A's infection; that source has yet to be identified.

As shown in the diversity plot, the S gene and Orf 3 at the junction between the replicase 1B and S genes had a higher number of sequence variations compared to other genomic regions (Fig. 2). This influenced our decision to perform additional sequence comparisons of the S, E, M and N structural genes and Orfs 3 and 10.

	Protein	orf3 0.040 (11/274)
	Amino Acid Site	11111 1 188902237 9 7115311961 3
Civet Cat	SZ16-ref	IGSLYMGLKER
Early epidemic	GD01	F. C. H. C. . .
	GZ50	F. CQH. C. . .
	[a]BJ02	FEC. H. C. . .
	BJ01	F. C. H. C. Q .
	[b]BJ04	F. C. H. C. . .
Late epidemic	[c]SoD	F. C. H. C. . .
	HKU-39849	F. C. HKC. . .
	TWC	F. C. H. C. . .
	TOR2	FRC. H. C. . .
	TW1	F. C. H. C. . .
	ZMY1	
Taiwan 1st wave	[d]TWC3	F. C. H. C. . .
Taiwan 2nd wave	[e]TWJ	F. C. H. C. . .
	TWK	F. C. H. CF. . .

a.BJ03 also had the identical sequence variation.
b.CUHK-W1 also had the identical sequence variation.
c. other strains with identical sequences include Frankfurt, FRA, WHU, ZJ01, AS, HSR1, Sin2677, Sin2679 and Sin2774, Sin2500, Sin2748, Sino3-11, Sino1-11, GD69, CUHK-Su10, PUMC01, PUMC02, PUMC03 and Amoy Garden group CUHK-AG01, CUHK-AG02, CUHK-AG03.
d. other strains with identical sequences include Taiwan group1 TWC2, TWH and TC1.
e. other strains with identical sequences include Taiwan group2 TWS, TWY, TC2 and TC3.

Fig. 4. Amino acid comparisons of the Orf 3 non-structural SCoV protein. The civet cat SZ16 SARS-CoV was used as a reference. A period (.) indicates concurrence with the top reference sequence (SZ16) in the alignment.

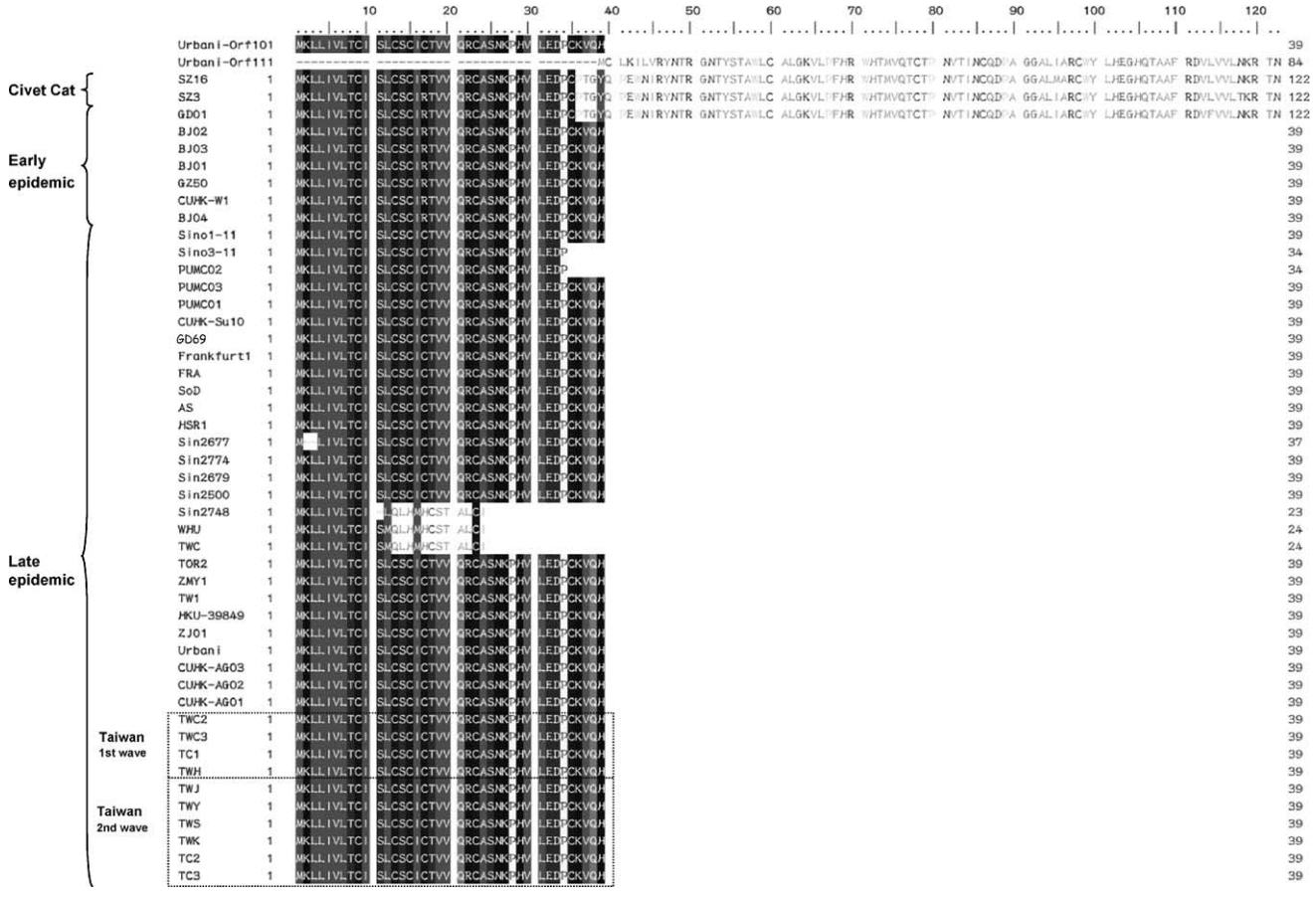
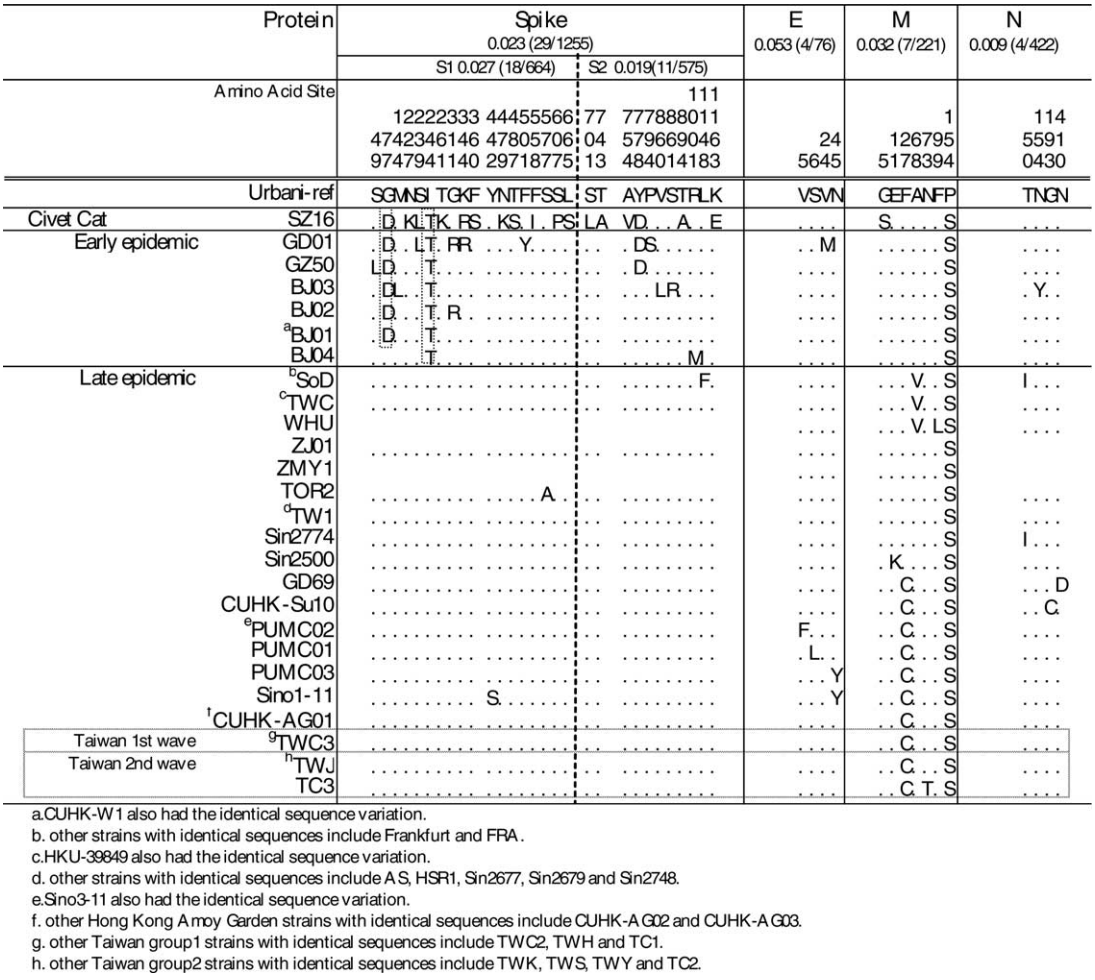


Fig. 5. Amino acid comparison of Orf 10, the SCoV non-structural protein. The Orf 10 and Orf 11 proteins of the Urbani human SCoV were used as references.



a.CUHK-W1 also had the identical sequence variation.
b. other strains with identical sequences include Frankfurt and FRA.
c.HKU-39849 also had the identical sequence variation.
d. other strains with identical sequences include AS, HSR1, Sin2677, Sin2679 and Sin2748.
e.Sino3-11 also had the identical sequence variation.
f. other Hong Kong Amoy Garden strains with identical sequences include CUHK-AG02 and CUHK-AG03.
g. other Taiwan group1 strains with identical sequences include TWC2, TWH and TC1.
h. other Taiwan group2 strains with identical sequences include TWK, TWS, TWY and TC2.

Fig. 6. Amino acid comparisons of S proteins from 44 human and civet SCoVs. The Urbani SCoV was used as a reference. A period (.) indicates concurrence with the top reference sequence (Urbani) in the alignment.

The S proteins of coronaviruses have been described as large, type I membrane glycoproteins that are responsible for both the binding of receptors to host cells and membrane fusion (Li et al., 2003; Xiao et al., 2003). The type I glycoproteins of coronaviruses, whose trimers resemble typical viral spikes, is transformed into virions through non-covalent interactions with M proteins. Coronavirus S proteins contain two domains (or two subunits, depending on whether or not S is cleaved) (Spiga et al., 2003). The S1 domain contains virus-neutralizing epitopes and the receptor-binding domain (Leparc-Goffart et al., 1998; Sanchez et al., 1999). Xiao et al. (2003) recently localized the SCoV receptor-binding domain (RBD) to amino acid residues 303–537 of the S1 protein. As shown in Fig. 6, we observed seven amino acid differences in the RBD of the S protein, including amino acid residues 311, 344, 360, 442, 479, 487 and 501. If we assume that the RBD is (a) conserved among different SCoVs, including civet SCoVs (Bonavia et al., 2002), and (b) more than 30–50 amino acids in length (Lasky et al., 1987), then it is possible that the RBD can be mapped onto amino acid residues 360–442.

Acknowledgements

We thank Mr. Jon Lindemann for editing our manuscript. This work was supported in part by a grant from the National SARS Research Program of the Republic of China National Science Council (grant no. SVIR12).

References

Bonavia, A., Zelus, B.D., Wentworth, D.E., Talbot, P.J., Holmes, K.V., 2002. Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E. J. Virol. 77, 2530–2538.

Center for Disease Control, Department of Health, Taiwan (ROC CDC), 2003. SARS Epidemics in Taiwan. In: Memoir of Severe Acute Respiratory Syndrome Control in Taiwan, pp. 4–5.

Centers for Disease Control and Prevention (CDC), 2003. Severe acute respiratory syndrome, Taiwan, 2003. MMWR Morb. Mortal. Wkly. Rep. No. 52, pp. 461–466.

Chen, P.T., Kao, C.L., Yeh, H.H., Wang, H.Y., Yang, C.Y., Liu, H.W., Tsai, S.F., Tsai, C.Y., Chen, T.H., Su, Y.J., 2003. Analysis of the whole-length sequences of ten strains of SARS coronavirus in Taiwan and its epidemiological implications. Memoir of Severe Acute Respiratory Syndrome Control in Taiwan, 39-42.

Chim, S.S., Tsui, S.K., Chan, K.C., Au, T.C., Hung, E.C., Tong, Y.K., Chiu, R.W., Ng, E.K., Chan, P.K., Chu, C.M., Sung, J.J., Tam, J.S., Fung, K.P., Waye, M.M., Lee, C.Y., Yuen, K.Y., Lo, Y.M., 2003. Genomic characterisation of the severe acute respiratory syndrome coronavirus of Amoy Gardens outbreak in Hong Kong. Lancet 362, 1807–1808.

Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A.M., Berger, A., Burguiere, A.M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J.C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H.D., Osterhaus, A.D.M.E., Schmitz, H., Doerr, H.W., 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. 348, 1967.

Felsenstein, J., 1989. PHYLIP — Phylogeny Inference Package, Version 3.2. Cladistics 5, 164–166.

Fouchier, R.A., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G.J., van den Hoogen, B.G., Peiris, M., Lim, W., Stohr, K., Osterhaus, A.D., 2003. Aetiology: Koch's postulates fulfilled for SARS virus. Nature 423, 240.

Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., Butt, K.M., Wong, K.L., Chan, K.W., Lim, W., Shortridge, K.F., Yuen, K.Y., Peiris, J.S.M., Poon, L.L.M., 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302, 276–278.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95–98.

Hsueh, P.R., Hsiao, C.H., Yeh, S.H., Wang, W.K., Chen, P.J., Wang, J.T., Chang, S.C., Kao, C.L., Yang, P.C., 2003. Microbiologic characteristics, serologic responses, and clinical manifestations in severe acute respiratory syndrome, Taiwan. Emerg. Infect. Dis. 9, 1163–1167.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., Rollin, P.E., Dowell, S.F., Ling, A.E., Humphrey, C.D., Shieh, W.J., Guarner, J., Paddock, C.D., Rota, P., Fields, B., DeRisi, J., Yang, J.Y., Cox, N., Hughes, J.M., LeDuc, J.W., Bellini, W.J., Anderson, L.J., and the SARS Working Group, 2003. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. 348, 1953–1966.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

Lasky, L.A., Nakamura, G., Smith, D.H., Fennie, C., Shimasaki, C., Patzer, E., Berman, P., Gregory, T., Capon, D.J., 1987. Delineation of region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. Cell 50, 975–985.

Leparc-Goffart, I., Hingley, S.T., Chua, M.M., Phillips, J., Lavi, E., Weiss, S.R., 1998. Targeted recombination within the spike gene of murine coronavirus mouse hepatitis virus-A59: Q159 is a determinant of hepatotropism. J. Virol. 72, 9628–9636.

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., Choe, H., Farzan, M., 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426, 450–454.

Marra, M.A., Jones, S.J.M., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S.N., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., Cloutier, A., Coughlin, S.M., Freeman, D., Girn, N., Griffith, O.L., Leach, S.R., Mayo, M., McDonald, H., Montgomery, S.B., Pandoh, P.K., Petrescu, A.S., Robertson, A.G., Schein, J.E., Siddiqui, A., Smailus, D.E., Stott, J.M., Yang, G.S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T.F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G.A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R.C., Krajden, M., Petric, M., Skowronski, D.M., Upton, C., Roper, R.L., 2003. The genome sequence of the SARS-associated coronavirus. Science 300, 1399–1404.

Peiris, J.S., Chu, C.M., Cheng, V.C., Chan, K.S., Hung, I.F., Poon, L.L., Law, K.I., Tang, B.S., Hon, T.Y., Chan, C.S., Chan, K.H., Ng, J.S., Zheng, B.J., Ng, W.L., Lai, R.W., Guan, Y., Yuen, K.Y., 2003a. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. Lancet 361, 1767–1772.

Peiris, J.S., Lai, S.T., Poon, L.L., Guan, Y., Yam, L.Y., Lim, W., Nicholls, J., Yee, W.K., Yan, W.W., Cheung, M.T., Cheng, V.C., Chan, K.H., Tsang, D.N., Yung, R.W., Ng, T.K., Yuen, K.Y., 2003b. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 19 (361), 1319–1325.

Poutanen, S.M., Low, D.E., Henry, B., Finkelstein, S., Rose, D., Green, K., Tellier, R., Draker, R., Adachi, D., Ayers, M., Chan, A.K., Skowronski, D.M., Salit, I., Simor, A.E., Slutsky, A.S., Doyle, P.W., Krajden, M., Petric, M., Brunham, R.C., McGeer, A.J., and the National Microbiology Laboratory, C.a.t.C.S.A.R.S.S.T., 2003. Identification of severe acute respiratory syndrome in Canada. N. Engl. J. Med. 348, 1995–2005.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Penaranda, S., Bankamp, B., Maher, K., Chen, M.H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J.L., Chen, Q., Wang, D., Erdman, D.D., Peret, T.C.T., Burns, C., Ksiazek, T.G., Rollin, P.E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A.D.M.E., Drosten, C., Pallansch, M.A., Anderson, L.J., Bellini, W.J., 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300, 1394–1399.

Sanchez, C.M., Izeta, A., Sanchez-Morgado, J.M., Alonso, S., Sola, I., Balasch, M., Plana-Duran, J., Enjuanes, L., 1999. Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence. J. Virol. 73, 7607–7618.

Spiga, O., Bernini, A., Ciutti, A., Chiellini, S., Menciassi, N., Finetti, F., Causarono, V., Anselmi, F., Prischi, F., Niccolai, N., 2003. Molecular modelling of S1 and S2 subunits of SARS coronavirus spike glycoprotein. Biochem. Biophys. Res. Commun. 310, 78–83.

WHO, 2003. Summary table of SARS cases by country, 1 November 2002–31 July 2003. http://www.who.int/csr/sars/country/table2003_09_23/en/.

Wu, J.S., Ho, M.S., Huang, T.M., Chen, K.T., Hsu, K.H., Su, I.J., Chiang, D.H., King, C.C., Wang, Y.F., Lin, P.F., 2003. Epidemiological investigation of the SARS outbreak in the Taipei Municipal Hoping Hospital. Memoir of Severe Acute Respiratory Syndrome Control in Taiwan, pp. 45–48.

Xiao, X., Chakraborti, S., Dimitrov, A.S., Gramatikoff, K., Dimitrov, D.S., 2003. The SARS-CoVs glycoprotein: expression and functional characterization. Biochem. Biophys. Res. Commun. 312, 1159–1164.

Zhong, N.S., Zheng, B.J., Li, Y.M., Poon, L.L.M., Xie, Z.H., Chan, K.H., Li, P.H., Tan, S.Y., Chang, Q., Xie, J.P., Liu, X.Q., Xu, J., Li, D.X., Yuen, K.Y., Peiris, J.S.M., Guan, Y., 2003. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February. Lancet 362, 1353–1358.