



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.

## Articles

# Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome

Y Guan, J S M Peiris, B Zheng, L L M Poon, K H Chan, F Y Zeng, C W M Chan, M N Chan, J D Chen, K Y C Chow, C C Hon, K H Hui, J Li, V Y Y Li, Y Wang, S W Leung, K Y Yuen, F C Leung

## Summary

**Background** Severe acute respiratory syndrome (SARS) is a newly emerged disease caused by a novel coronavirus (SARS-CoV), which spread globally in early 2003, affecting over 30 countries. We have used molecular epidemiology to define the patterns of spread of the virus in Hong Kong and beyond.

**Methods** The case definition of SARS was based on that recommended by WHO. We genetically sequenced the gene for the S1 unit of the viral spike protein of viruses from patients with SARS in Hong Kong (138) and Guangdong (three) in February to April, 2003. We undertook phylogenetic comparisons with 27 other sequences available from public databases (Genbank).

**Findings** Most of the Hong Kong viruses (139/142), including those from a large outbreak in an apartment block, clustered closely together with the isolate from a single index case (HKU-33) who came from Guangdong to Hong Kong in late February. Three other isolates were genetically distinct from HKU-33 in Hong Kong during February, but none of these contributed substantially to the subsequent local outbreak. Viruses identified in Guangdong and Beijing were genetically more diverse.

**Interpretation** The molecular epidemiological evidence suggests that most SARS-CoV from the outbreak in Hong Kong, as well as the viruses from Canada, Vietnam, and Singapore, are genetically closely linked. Three viruses found in Hong Kong in February were phylogenetically distinct from the major cluster, which suggests that several introductions of the virus had occurred, but that only one was associated with the subsequent outbreak in Hong Kong, which in turn spread globally.

*Lancet* 2004; **363**: 99–104

**Departments of Microbiology** (Y Guan PhD, J S M Peiris DPhil, B Zheng PhD, L L M Poon DPhil, K H Chan PhD, S W Leung BSc, K Y Yuen MD) **and Zoology** (F Y Zeng PhD, C W M Chan BSc, M N Chan MPhil, J D Chen PhD, K Y C Chow BSc, C C Hon BSc, K H Hui MPhil, J Li PhD, V Y Y Li BSc, Y Wang BSc, F C Leung PhD), **University of Hong Kong, Hong Kong SAR, China**

**Correspondence to:** Dr F C Leung, Department of Zoology, University of Hong Kong, Pokfulam Road, Hong Kong SAR (e-mail: fcleung@hkucc.hku.hk)

## Introduction

Severe acute respiratory syndrome (SARS) was first reported in the Chinese province of Guangdong in November, 2002, and caused an outbreak there in January to April, 2003.<sup>1</sup> A novel coronavirus (SARS-CoV) has been identified as the causative agent of this emerging disease.<sup>2,3</sup> In Hong Kong Special Administrative Region, the index case of the SARS outbreak can be dated back to February, 2003. Since then the disease has become pandemic, with (at the time of writing) 8098 patients and 744 deaths in over 30 countries.<sup>4,5</sup> The complete genomes of four SARS-CoV isolates have been sequenced independently.<sup>6</sup> Comparison with sequences of other human and animal coronaviruses has shown that SARS-CoV is distinct from them and warrants classification as a new group within the Coronaviridae.<sup>7–9</sup>

SARS-CoV is an enveloped, positive-sense, single-stranded-RNA virus. The genome is about 27 000 bp, encoding the replicase and spike, envelope, membrane, and nucleocapsid proteins.<sup>7–9</sup> In most animal and human coronaviruses, the gene for the spike protein is divided into amino (S1) and carboxyl (S2) regions; the S1 subunit is associated with receptor-binding functions, and the S2 subunit is a conserved transmembrane protein mediating fusion of viral and cellular membranes.<sup>10</sup> The spike protein is also important in viral entry, pathogenesis,<sup>10</sup> antiviral immune response,<sup>11</sup> virulence,<sup>12</sup> and cellular<sup>13</sup> or even species tropism.<sup>14</sup> In general, the spike protein is the most variable region of the genome and has generally been used for genotyping of coronaviruses, as shown by studies on human coronaviruses<sup>15</sup> and on avian infectious bronchitis virus.<sup>16</sup> Therefore, phylogenetic analysis of SARS-CoV is primarily based on the S1 region of the spike protein sequence.

We have used molecular biological techniques to investigate the phylogenetic relations among SARS-CoV isolates from Hong Kong and from three patients from Guangdong during the period February to April, 2003. By analysis of a 2149 bp fragment of the S1 gene of these viruses and data available in Genbank, the molecular epidemiological relations of these SARS-CoVs could be reconstructed.

## Methods

### Patients

The case definition of SARS we used was modified from that of WHO as previously described.<sup>3</sup> Samples (n=138) from 137 patients with SARS admitted to hospitals in Hong Kong Special Administrative Region were analysed: three from February, 64 from March, and 71 from April, 2003. We also studied three virus isolates obtained from three patients in Guangdong Chest Hospital during February. All these samples were collected as part of routine clinical management. The epidemiological relations and clinical features of the key patients associated with the Hong Kong and Guangdong isolates

Virus strain	Date of collection*	Clinical sample	Key epidemiological information
GZ-43	Feb 18	NPA	Patient in Guangdong Chest Hospital
GZ-50	Feb 18	NPA	Patient in Guangdong Chest Hospital
GZ-60	Feb 18	NPA	Patient in Guangdong Chest Hospital
HKU-33	Feb 24	NPA	Patient 1; index case in Hong Kong outbreak; onset of disease Feb 15; arrived in Hong Kong Feb 21; stayed in hotel M; admitted to hospital Feb 22
HKU-33867	Feb 24	NPA	Recent travel to Guangdong visiting his father who died of pneumonia Feb 22
HKU-36	Feb 28	NPA	Pneumonia 2 days after return from travel to Guangdong
HKU-39849	March 4	Lung biopsy	Patient 2A; social contact of patient 1; onset of disease Feb 24
HKU-65	March 9	Throat swab	Amoy Gardens outbreak; onset of disease March 24
HKU-66	March 9	Stool	Amoy Gardens outbreak; onset of disease March 24
HKU-55	March 27	Stool	Patient 3A: health-care worker
HKU-56	March 28	Stool	Patient 3B: health-care worker

NPA=nasopharyngeal aspirate. \*All dates are in 2003.

Table 1: **Origins and epidemiological data of the key virus isolates used in the study**

have been described previously.<sup>1,4,17</sup> Those relevant to this study are summarised in table 1. These include the presumed index case patient 1 (HKU-33) and patients directly related to him (figure 1).

Patient X, the index case of the large outbreak of over 300 cases in an apartment block, Amoy Gardens, was probably indirectly linked epidemiologically to patient 1.<sup>18-20</sup> To investigate whether apparent differences in transmission and clinical presentation of disease acquired in Amoy Gardens are related to viral genetic differences, the whole genomes of viral isolates from two of these patients (HKU-65 and HKU-66) were sequenced. The full genomes of four viruses from patients with no epidemiological link to patient 1 in Hong Kong (HKU-36) and Guangdong (GZ-43, GZ-50, GZ-60) were also studied.

### Procedures

Total RNA for investigation of viral genetic sequences of GZ-43, GZ-50, GZ-60, HKU-36, HKU-39849, HKU-65, and HKU-66 was obtained from virus isolates,

Name	Sequence (starting from 5')
<b>Forward primer</b>	
Set 1	
PF138	GGAAGTGTGTAAATGTCTCT
NF72	GGTAGGCTTATCATTAGAGA
SF30	CAACAGAGTTGTGGTTTCA
Set 2	
PF615	GTCCTGTGTACCTTCTGG
NF708	GCCTTTTACCTGCTCAA
SF1214	CCAGGACAAACTGGTGTAT
Set 3	
PF1194	CCAGGACAAACTGGTGTAT
NF1321	GGTACTTAGACATGGCAAGC
SF1610	GGACTACTGGTACTGGTGT
<b>Reverse primer</b>	
Set 1	
PR946	TCTCACAACTCTCTGA
NR868	CAGACGATTTGAGTTCAG
SF450	CATGGGTACACAGACACTACT
Set 2	
PR1729	CGCAAGGTGAAATGTCTAA
NR1476	TGGTAGCCAATGCCAGTAGT
SR1322	GCTTGCCATGTCTAAGATAC
Set 3	
PR2440	CAGCATCAGCGAGTGTCCAC
NR2300	CTTGAGCGAACACTTCAC
SR1710	CGCAAGGTGAAATGTCTAA

Table 2: **Primer sequences used for nested PCR and sequencing**

whereas all other sequences derived in this study were obtained by direct RT-PCR amplification of the S1 gene fragment from the clinical sample. The cDNA was generated by reverse transcription with random primers. The first 2149 bp of the spike protein gene were amplified from the viral cDNA by nested PCR and characterised by sequencing of the amplification product. RT-PCR and sequencing procedures were done as previously described.<sup>9</sup> Three sets of primers were used for nested PCR and sequencing reactions. These primers were numbered according to their nucleotide position relative to the start codon in the open reading frame of the spike protein gene (table 2). Viral RNA extraction, PCR amplification, and the analysis of PCR-amplified products were carried out in different areas of the laboratory. Measures to avoid cross-contamination by PCR-amplified DNA were adopted.

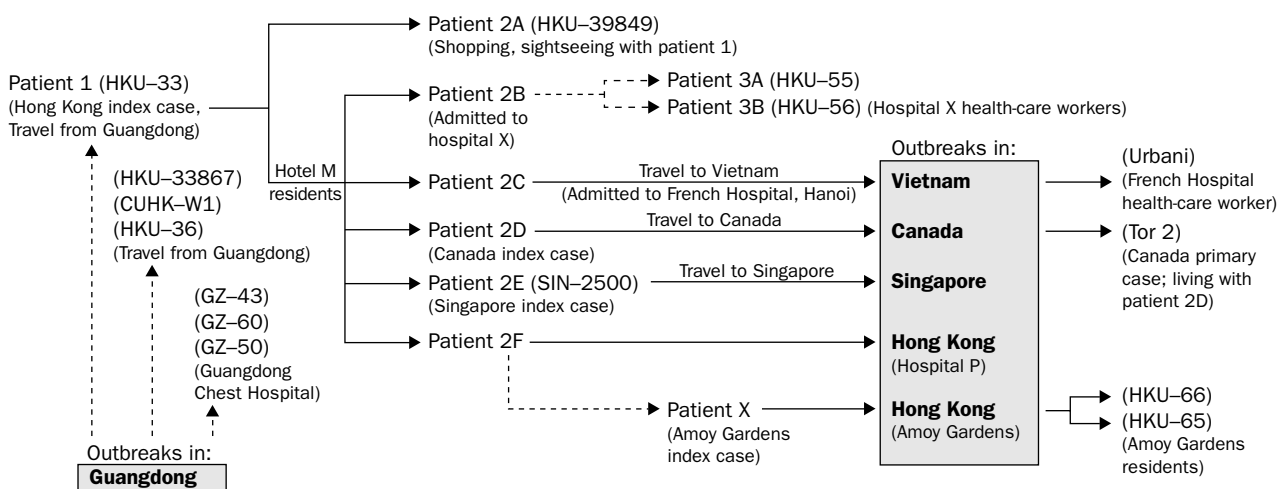


Figure 1: **Epidemiological data on selected key virus isolates**

Schematic representation of the relations between the key patients and the outbreaks in different regions of the world. Dotted arrows indicate uncertain transmission route, and grey boxes indicate regions of outbreaks. Designations of the viruses are shown in parentheses. The clinical and epidemiological relations of some of these patients have been described previously.<sup>17</sup> In reference 17 these patients were designated as: patient 1=patient 1; patient 2A=patient 2; patient 3A=patient 6; patient 3B=patient 7.

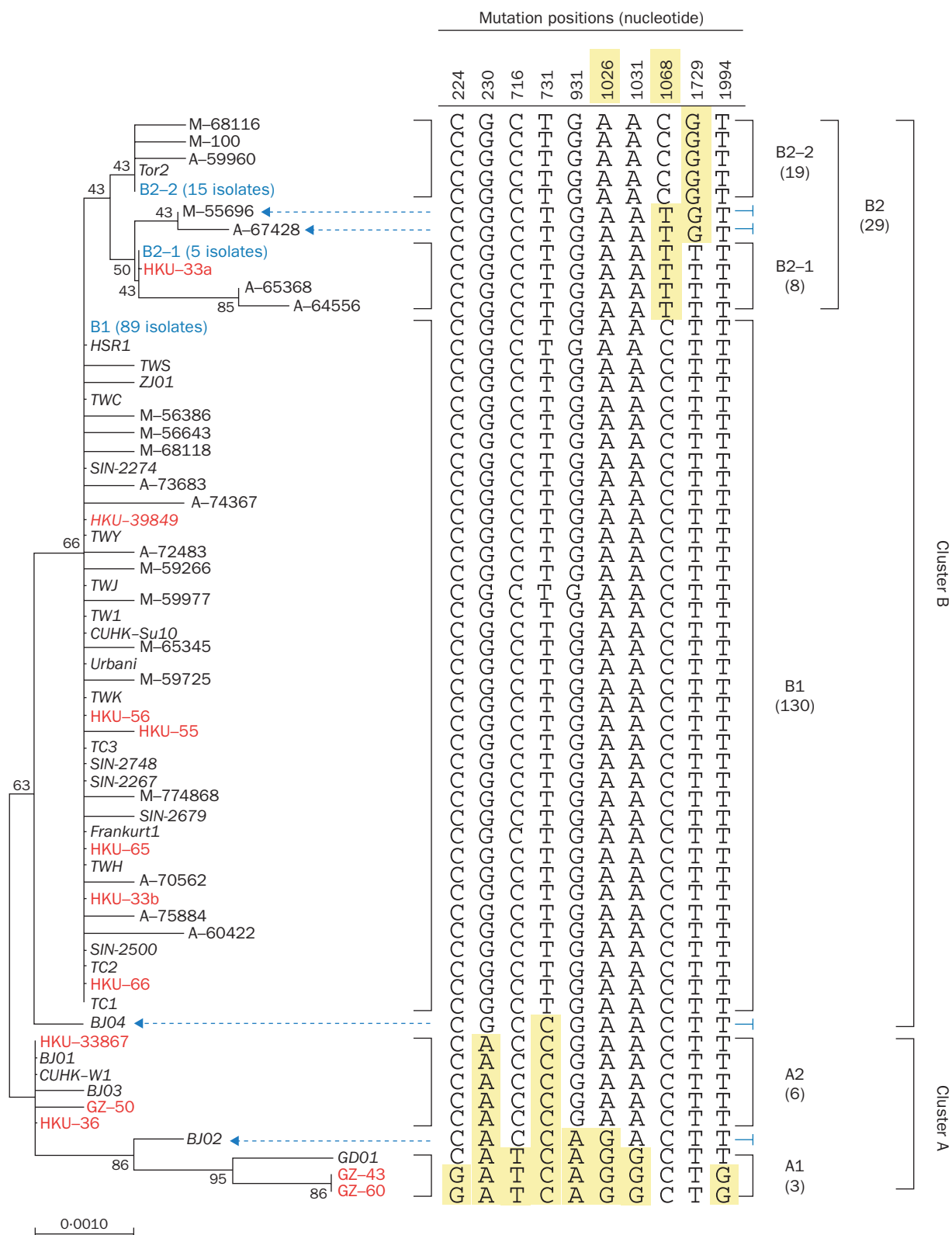


Figure 2: Phylogenetic analysis of 169 SARS-CoV isolates

Numbers at nodes indicate bootstrap values (%). Branch length shows the genetic distance with reference to the horizontal scale bar. For convenience of display of a large number of isolates, field isolates with zero genetic distance compared with their subcluster (B1, B2-1, and B2-2) common ancestors are replaced by the subcluster name with the number of isolates shown in parentheses (highlighted in blue). The full phylogenetic tree of the 169 isolates is available at <http://image.thelancet.com/extras/03art5344webfigure1.pdf>. Viral gene sequences highlighted in red represent key isolates described in figure 1. Sequences previously available in Genbank are shown in bold and italic. Subcluster names are shown on the right of the tree. Total numbers of isolates included in the cluster are shown in brackets. Cluster transition isolates are indicated by blue dotted arrows. Epidemiological details of key viruses are provided in table 1 and figure 1. Other viruses sequenced as part of this study are prefixed with F, M, or A, to represent the month of sample collection (February, March, or April). Sequence alignment of the recurrent critical point mutations that distinguish the groups and subgroups are shown on the right of the tree. All these mutations are non-silent except at 1026 and 1068 (highlighted in green).

The phylogenetic relations of these viruses, including the 27 sequences downloaded from Genbank (<http://image.thelancet.com/extras/03art5344webtable.pdf>), were reconstructed on the basis of the nucleotide sequence of the S1 region. Optimum alignments of the viral nucleotide sequences were made with MegAlign (version 4-03).<sup>21</sup> The phylogenetic trees based on the optimum alignment were constructed by the neighbour-joining method<sup>22</sup> with MEGA (version 2.1).<sup>23</sup> Robustness of the trees was assessed by bootstrap analysis on 1000 replicates of the dataset. Critical point mutations associated with different lineages are shown (figure 2).

The complete genome sequences of six selected virus isolates were obtained (primers designed on the basis of the genetic sequence of HKU-39849, data not shown).<sup>9</sup> The sequences were aligned and compared with the ten other full virus genomes available in Genbank (figure 3; webtable).

**Role of the funding source**

The sponsor of the study had no role in study design; collection, analysis, or interpretation of data; or in the writing of the report.

**Results**

We found that the viruses divided into two distinct clusters with five subclusters, designated A1, A2, B1, B2-1, and B2-2 (figure 2). Cluster A included ten isolates mainly derived from Beijing (BJ) and Guangdong (GZ), which were characterised by mutations at nucleotide positions 230 and 731, resulting in aminoacid changes. Only three viruses in cluster A were from patients from Hong Kong (HKU-33867, CUHK-W1, and HKU-36; figure 2; <http://image.thelancet.com/extras/03art5344webfigure1.pdf>). Since these three patients were recently returned from Guangdong (table 1) and the incubation period of the disease ranges from 1 day to 14 days, all three probably acquired their infections in

February directly from Guangdong, rather than from Hong Kong.<sup>1,18,24</sup>

The other viruses from Hong Kong grouped together in cluster B, together with the virus sequences derived from the presumed index case for the outbreak in Hong Kong (HKU-33a/b), as well as viruses from Canada (Tor2), Singapore (eg, SIN2500), Germany (Frankfurt 1), Italy (HSR1), Singapore, Taiwan, and Vietnam (Urbani). An isolate from Zhejiang province of China (ZJ01) was also in this group.

The initial nasopharyngeal aspirate collected from patient 1 (HKU-33) 7 days after onset of his illness had two distinct variants designated HKU-33a and HKU-33b, falling within subclusters B2-1 and B1, respectively, with a silent sequence variation at nucleotide position 1068 of the S1 region (C/T). Subsequent clinical samples from this patient contained one or other variant. HKU-33a was detected in a nasopharyngeal aspirate on day 9 and after death in liver and lung samples on day 25; HKU-33b was also detected in an endotracheal aspirate and bronchoalveolar lavage sample on day 10.

The viral sequences derived from cases directly linked to patient 1 (figure 1) were further subdivided into two subclusters B1 and B2 (figure 2). Viruses including HKU-39489, HKU-55, HKU-56, Urbani, and CUHK-Su10<sup>18</sup> were all grouped in cluster B1 in the same sublineage as HKU-33b. Tor2 was in cluster B2 in the same lineage as HKU-33a.

Four viruses, BJ04, BJ02, M-55696, and A-67428, carried the characteristic residues of two subcluster prototypes and could not be efficiently classed in any of the subclusters; they were defined as subcluster transition viruses (figure 2).

In total, 56 single nucleotide variations were identified within the S1 region of the 169 viruses analysed (from 168 patients, including 33a and b). 15 were silent (<http://image.thelancet.com/extras/03art5344webfigure2.pdf>). 41 of these sequence variations occurred in only one

Gene/region		1a	1b	1b	1b	1b	1b	1b	1b	S	S	S	S	S	S	X1	X1	X1	X1	X1	M	M	M	X3	N	3'UTR										
Strain/position		9404	17564	17846	18065	19064	19084	20840	20992	21706	21721	22207	22222	22422	22517	22522	23485	23823	24566	25027	25286	25298	25779	25844	26032	26477	26586	26600	27827	29176	29725					
Substitutions	Nucleotides	HKU-39849	T	T	C	A	C	A	G	T	G	C	T	G	A	A	T	T	T	T	T	G	A	A	T	T	T	T	T	T	T	C	G			
		HKU-65	T	T	C	G	A	C	A	G	T	G	C	T	G	A	A	T	T	T	T	T	G	A	A	T	T	T	T	T	T	T	C	C		
		HKU-66	T	T	C	G	A	C	A	G	T	G	C	T	G	A	A	T	T	T	T	T	G	A	A	T	G	T	T	T	T	T	T	C	C	
		HKU-36	/	G	C	A	A	C	A	G	T	A	C	C	G	A	A	T	T	T	T	T	T	G	A	A	T	T	T	T	T	T	C	C		
		CUHK-W1	C	G	T	G	G	C	A	G	T	A	C	C	G	A	A	T	T	T	T	T	T	G	A	A	T	T	T	T	T	T	C	C		
		CUHK-Su10	T	T	T	G	A	C	A	G	T	G	C	T	G	A	A	T	T	T	T	T	T	G	A	A	T	G	T	C	T	T	T	C		
		GZ-43	/	G	C	G	A	C	G	A	A	A	T	C	A	G	G	C	G	C	C	A	A	C	T	A	T	C	C	C	C	H	C	C		
		GZ-50	/	G	C	G	A	C	A	G	T	A	C	C	G	A	A	T	G	T	T	T	G	A	A	T	T	T	T	T	T	T	C	C		
		GZ-60	/	G	C	G	A	C	G	A	A	A	T	T	A	G	G	C	G	C	C	A	A	C	T	A	T	C	C	C	C	T	C	C		
		BJ01	C	G	C	G	A	C	A	G	T	A	C	C	G	A	A	T	T	T	T	T	T	G	A	A	T	T	T	T	T	T	T	C	C	
		TW1	T	T	C	G	A	C	A	G	T	G	C	T	G	A	A	T	T	T	T	T	T	G	A	A	T	T	T	T	T	T	T	C	C	
		Sin-2500	T	T	C	G	A	T	A	G	T	G	C	T	G	A	A	T	T	T	T	T	T	G	A	A	T	T	T	T	T	T	T	T	C	
		Tor2	T	T	C	G	A	C	A	G	T	G	C	T	G	A	A	T	T	T	T	T	T	A	A	A	T	T	T	T	T	T	T	T	C	
		Urbani	T	T	C	G	G	C	A	G	T	G	C	T	G	A	A	T	T	T	T	T	T	G	A	A	T	T	T	T	T	T	T	T	C	
		Frankfurt	T	T	C	G	A	T	A	G	T	G	C	T	G	A	A	T	T	T	T	T	T	T	G	A	A	T	T	T	T	T	T	T	T	C
		HSR1	T	C	G	A	C	A	G	T	G	C	T	G	A	A	T	T	T	T	T	T	T	G	A	A	T	T	T	T	T	T	T	T	C	
Aminoacids		V	E	Silent	Silent	Silent	T	Silent	R	I	G	S	I	G	Silent	K	L	Y	Silent	L	F	G	E	R	Silent	F	Silent	Silent	C	Silent	N/A					
		A	D				I		K	N	D	L	T	R		R	S	D		P	I	R	A	W		C			R							

Figure 3: Nucleotide and aminoacid substitutions of SARS-CoV isolated in Amoy Gardens

Comparisons were made with other virus isolates by full genome sequence alignment. Except for the Hong Kong and Guangdong samples, the index case or first uploaded sequence from other cities and countries was chosen. Nucleotides were numbered on the basis of the HKU-39849 full sequence.<sup>9</sup> Slash indicates that the sequence data are not available, and nucleotides highlighted in yellow are those that differ from the majority at the corresponding location. Only substitutions occurring in more than one virus are shown.

isolate. A 3 bp deletion was identified at position 215–217 in the S1 gene in viruses GZ-43, GZ-60, and M-61576. GZ-43 and GZ-60 were isolated from nasopharyngeal aspirates of two health-care workers in Guangdong Chest Hospital on Feb 18, 2003, and are probably epidemiologically related. The M-61576 sequence was derived by direct RT-PCR and sequencing from the urine sample of a patient from the Amoy Gardens outbreak in April, 2003, which is geographically and epidemiologically unrelated to the two Guangdong viruses. This 3 bp deletion leads to deletion of an aminoacid residue and a change of the residue next to it.

Isolates HKU-65 and HKU-66, derived from the throat and stool samples, respectively, of two patients from the Amoy Gardens outbreak, were completely genetically sequenced (figure 3). With HKU-39849 as the reference strain, there were seven sequence variations. Four were recurrent variations that were found in more than one virus (figure 3), and the other three were non-recurrent (data not shown) and located at positions 20485 (ORF1ab) and 27384 (X2) in the genome of HKU-65 and at position 16325 (ORF1ab) in HKU-66. Another characteristic sequence variation was located in the 3' untranslated region (nucleotide 29725) of both genomes (G to C). This mutation occurred in only three Guangdong-related virus isolates, GZ-43, GZ-50, and HKU-36. However, these three viruses did not cluster together in the phylogenetic analysis of the S1 gene region, and the patients do not share a known epidemiological linkage (table 1).

## Discussion

The molecular epidemiological analysis suggests that there were introductions of several virus strains into Hong Kong during February, 2003—HKU-39849, HKU-33867, HKU-36, and CUHK-W1. However, it also supports the conclusion derived from the conventional epidemiology<sup>4</sup> that the SARS outbreak in Hong Kong during March and April was largely, if not completely, derived from a common source.

Although only one of the four patients diagnosed with SARS in Hong Kong in February had viruses within cluster A, all 65 patients diagnosed in March (including HKU-39849 and CUHK-Su10 viruses previously sequenced) and 71 patients diagnosed in April had viruses of cluster B, together with the viruses (HKU-33a/b) derived from the presumed index case patient 1. Although molecular epidemiology cannot pinpoint the initiator of this outbreak with precision, the first virus from cluster B identified in Hong Kong was that isolated from patient 1 (HKU-33). Thus, taken together with the epidemiological data, the molecular epidemiology is compatible with the suggestion<sup>4</sup> that patient 1 was the index case of the outbreak in Hong Kong. The viruses in Canada, Singapore, Taiwan, and Vietnam also belong to cluster B, so the evidence supports the contention that these outbreaks derived from patient 1.

Our findings also reveal greater genetic diversity among strains in mainland China. Viruses from Guangdong and Beijing viruses mainly grouped in cluster A, whereas a Zhejiang (ZJ01) isolate grouped in cluster B. Further analysis of strains from mainland China is needed to consolidate the overall patterns of viral spreading but the greater genetic diversity of viruses in mainland China implies that SARS-CoV has been circulating there for a while. We propose that a single virus strain (HKU-33a/b) from this diverse pool infected patient 1, who then travelled to Hong Kong and initiated

the outbreak there, which then rapidly spread to other countries around the world.

The three patients diagnosed during February with viruses in cluster A (HKU-33867, CUHK-W1, and HKU-36) all had epidemiological evidence that suggested infection had been acquired directly from Guangdong, and they represent evidence for several independent entries of SARS-CoV into Hong Kong from Guangdong during February. Why these viruses did not spread in the Hong Kong population when HKU-33a/b did so is unclear. Epidemiological data suggest that a few index cases caused a disproportionate number of secondary cases, the so-called “super-spreading incidents”.<sup>24</sup> The initiation by HKU-33a/b in patient 1 of the first super-spreading incident in Hong Kong may have been simply a matter of chance. Once this had occurred, second super-spreading events (as happened in Prince of Wales Hospital and in the Amoy Gardens apartment block) were stochastically more likely to follow from this cluster. An alternative hypothesis is that the viruses of the HKU-33a/b lineage (cluster B) are biologically more predisposed to initiating super-spreading events. This hypothesis needs more detailed investigation. Whether the presence of more than one genetic variant in patient 1 is relevant to the rapid spreading of this virus strain also requires further investigation.

There was an indication that the epidemiology and the clinical range of illness associated with the large outbreak in the high-rise apartment block in Amoy Gardens differed from that reported previously.<sup>19,20</sup> Sequence variations in the S1 gene are associated with significant changes in tissue tropisms of animal coronaviruses.<sup>25,26</sup> Therefore, investigation of whether the isolates from Amoy Gardens patients had significant variation in the gene for the spike protein was crucial, in case it could explain the presumed difference in disease phenotype. However, sequence comparison of the two Amoy Gardens isolates did not reveal any significant variations within the S1 gene (figure 2) nor across the rest of the genome (figure 3). This virus is therefore unlikely to have arisen from a different source from that of the major Hong Kong outbreak. The explosive transmission of the virus during the Amoy Gardens outbreak may be explained by an unusual route of transmission.<sup>20</sup> Aerosolisation of contaminated sewage<sup>20</sup> and a role for animal vectors such as rodents<sup>27</sup> have been proposed.

The use of molecular epidemiology to complement conventional epidemiology provides additional understanding of the transmission of the disease and will have a major effect on controlling the spread of disease. The findings of this study lay the foundation for understanding of the evolution and the genetic diversity of SARS-CoV.

## Contributors

F C Leung, Y Guan, and J S M Peiris are the coprincipal investigators who planned the study and wrote the report. F Y Zeng is the principal coordinator of the sequencing team; C W M Chan, M N Chan, J D Chen, K Y C Chow, C C Hon, K H Hui, J Li, V Y Y Li, Y Wang, and S W Leung did the genetic sequencing of the virus isolates and phylogenetic analysis. B Zheng coordinated the virology studies in Guangdong, and L L M Poon, K H Chan, and K Y Yuen supervised the virology and provided clinical information for this project. All the investigators contributed to the revision of the report and agreed to its final content.

## Conflict of interest statement

None declared.

## Acknowledgments

This research was supported by the US National Institute of Allergy and Infectious Diseases (public health research grant A195357) and the Research Grants Council of Hong Kong (HKU7542/03M and HKU7553/03M).

## References

- 1 Zhong NS, Zheng BJ, Li YM, et al. Epidemiological and aetiological studies of patients with severe acute respiratory syndrome (SARS) from Guangdong in February 2003. *Lancet* 2003; **362**: 1353–58.
- 2 Peiris JS, Lai ST, Poon LL, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 2003; **361**: 1319–25.
- 3 Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 2003; **348**: 1953–56.
- 4 Centers for Disease Control and Prevention. Update: outbreak of severe acute respiratory syndrome—worldwide. *MMWR Morb Mortal Weekly Rep* 2003; **52**: 241–45.
- 5 WHO. Cumulative number of reported probable cases of severe acute respiratory syndrome (SARS). <http://www.who.int/csr/sars/country/en> (accessed Aug 25, 2003).
- 6 Enserink M, Vogel G. Hungry for details, scientists zoom in on SARS genomes. *Science* 2003; **300**: 715–16.
- 7 Marra MA, Jones SJ, Astell CR, et al. The genome sequence of the SARS-associated coronavirus. *Science* 2003; **300**: 1399–404.
- 8 Rota PA, Oberste MS, Monroe SS, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003; **300**: 1394–99.
- 9 Zeng FY, Chan CW, Chan MN, et al. The complete genome sequence of severe acute respiratory syndrome coronavirus (SARS-CoV) strain HKU-39498 (HK-39). *Exp Biol Med* 2003; **228**: 866–73.
- 10 Gallagher TM, Buchmeier MJ. Coronavirus spike proteins in viral entry and pathogenesis. *Virology* 2001; **279**: 371–74.
- 11 Krokhn O, Li Y, Andonov A, et al. Mass spectrometric characterization of proteins from the SARS virus: a preliminary report. *Mol Cell Proteomics* 2003; **2**: 346–56.
- 12 Phillips JJ, Chua MM, Rall GF, Weiss SR. Murine coronavirus spike glycoprotein mediates degree of viral spread, inflammation, and virus-induced immunopathology in the central nervous system. *Virology* 2002; **301**: 109–20.
- 13 Casais R, Dove B, Cavanagh D, Britton P. Recombinant avian infectious bronchitis virus expressing a heterologous spike gene demonstrates that the spike protein is a determinant of cell tropism. *J Virol* 2003; **77**: 9084–99.
- 14 Haijema BJ, Volders H, Rottier PJ. Switching species tropism: an effective way to manipulate the feline coronavirus genome. *J Virol* 2003; **77**: 4528–38.
- 15 Hays JP, Myint SH. PCR sequencing of the spike genes of geographically and chronologically distinct human coronaviruses 229E. *J Viral Methods* 1998; **75**: 179–93.
- 16 Lee CW, Jackwood MW. Origin and evolution of Georgia 98 (GA98), a new serotype of avian infectious bronchitis virus. *Virus Res* 2001; **80**: 33–9.
- 17 Tsang KW, Ho PL, Ooi GC, et al. A cluster of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 2003; **348**: 1977–85.
- 18 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**: 187–88.
- 19 Peiris JS, Chu CM, Cheng VC, et al. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* 2003; **361**: 1767–72.
- 20 Department of Health, Hong Kong Special Administrative Region, China. Outbreak of severe acute respiratory syndrome (SARS) at Amoy Gardens, Kowloon Bay, Hong Kong: main findings of the investigation. 2003 Apr. [http://www.info.gov.hk/info/ap/pdf/amoy\\_e.pdf](http://www.info.gov.hk/info/ap/pdf/amoy_e.pdf) (accessed Aug 25, 2003).
- 21 Clewley JP, Arnold C. MEGALIGN. The multiple alignment module of LASERGENE. *Methods Mol Biol* 1997; **70**: 119–29.
- 22 Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**: 406–25.
- 23 Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 2001; **17**: 1244–45.
- 24 Donnelly CA, Ghani AC, Leung GM, et al. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* 2003; **361**: 1761–66.
- 25 Britton P, Mawditt KL, Page KW. The cloning and sequencing of the virion protein genes from a British isolate of porcine respiratory coronavirus: comparison with transmissible gastroenteritis virus genes. *Virus Res* 1991; **21**: 181–98.
- 26 Hasoksuz M, Sreevatsan S, Cho KO, Hoet AE, Saif LJ. Molecular analysis of the S1 subunit of the spike glycoprotein of respiratory and enteric bovine coronavirus isolates. *Virus Res* 2002; **84**: 101–09.
- 27 Ng SK. Possible role of an animal vector in the SARS outbreak in Amoy Gardens. *Lancet* 2003; **362**: 570–72.