Single nucleotide variations (SNVs) at 5 loci (17564, 21721, 22222, 23823, and 27827) were used to define the molecular epidemiologic characteristics of severe acute respiratory syndrome–associated coronavirus (SARS-CoV) from Beijing patients. Five fragments targeted at the SNV loci were amplified directly from clinical samples by using reverse transcription–polymerase chain reaction (RT-PCR), before sequencing the amplified products. Analyses of 45 sequences obtained from 29 patients showed that the GGCTC motif dominated among samples collected from March to early April 2003; the TGTTT motif predominated afterwards. The switch from GGCTC to TGTTT was observed among patients belonging to the same cluster, which ruled out the possibility of the coincidental superposition of 2 epidemics running in parallel in Beijing. The Beijing isolates underwent the same change pattern reported from Guangdong Province. The same series of mutations occurring in separate geographic locations and at different times suggests a dominant process of viral adaptation to the host.

Severe acute respiratory syndrome (SARS) is a new infectious disease that spread worldwide in early 2003, affecting >30 countries, with >8,098 cases and 774 deaths reported (1). Beijing, People’s Republic of China, experienced the largest SARS outbreak in the world, with 2,523 cases and 181 deaths by June 12, 2003 (2,3). The epidemic occurred in 2 phases. The first phase began on March 5, 2003, and was caused by a patient who had been infected in Guangzhou and was involved in a superspreader event (SSE) in Beijing hospitals. Most patients in this period proved to be directly or indirectly linked with the index patient by traditional epidemiologic investigations. Molecular epidemiology, based on genome sequencing of the early isolates, also provided evidence that Beijing infections were closely related to those from the Guangdong epidemic (4). The second phase was marked by widespread transmission in healthcare facilities and communities, with incidence peaking in late April, followed by a dramatic decline in occurrence during the first week of May. The last probable case was noted on May 29, 2003 (5). During this phase, many case-patients had no apparent contact with SARS patients.

After the sequencing of the whole genome (6–9) information on viral strains from different geographic and temporal origins became available in GenBank. Comparative sequence analyses identified 5 loci, sequence variants of which segregated together as specific genotypic patterns, which could be used to define epidemic phases (10). All or some of the 5 loci were included in previous molecular epidemiologic studies (4,11–13), making them important genetic signatures to differentiate lineage-specific and temporal-specific patterns. In this study, we investigated the genetic variations of SARS-CoV in Beijing based on the 5-locus signature. Also, by sequence comparison among patients from 1 case cluster and different samples from 1 patient, the adaptable mutation of the virus in the host was further explored.

Methods

Participants

Study participants were recruited from 2 hospitals designated for SARS patients in Beijing. All of them fit the World Health Organization (WHO) case definition for probable SARS, i.e., temperature ≥38°C, cough or shortness of breath, new pulmonary infiltrates on chest radiograph, and a history of exposure to a SARS patient or of
Results

A total of 160 samples (81 stools and 79 sputum samples) from 62 patients with positive results by RT-PCR were included in this study. Of these, 45 samples (36 sputum samples and 9 stools) from 29 patients (17 men and 12 women, with a median age of 32 years) yielded amplicons for the 5 targeted loci (Table 2). The patients came from 2 SARS-designated hospitals in Beijing, with disease onset ranging from March to May, 2003. Four patients had serious conditions during hospitalization, including pulmonary aggravation requiring oxygen ventilation or transfer to an intensive care unit. No patient died.

The sequences of the 45 positive specimens were compared with SARS-CoV genome sequences available from the public database (GenBank). The sequence variants in 5 loci (17564, 21721, 22222, 23823, and 27827) defined 3 kinds of motifs: GGCTC, TGGTTT, and GATTC (Table 2). In addition, 4 new SNVs were identified at nucleotides 17620, 22077, 22589, and 27749 in >1 patient. These variations appeared independently in several isolates, which indicates that they are not RT-PCR artifacts. None of them had been previously reported, with 3 nucleotide substitutions leading to amino acid changes (Table 3).

Twelve patients in this study belonged to a cluster. They derived from an SSE indirectly linked with the earliest SARS patients in Beijing. The first 2 patients of this cluster, who became ill on March 10 and 21, respectively, harbored the GGCTC motif. The remaining patients, who became ill from March 31 to May 4, showed the TGGTTT motif. Among patients outside of the cluster, 5 of 6 patients with onset date before April had the GGCTC motif, while the TGGTTT motif became predominant later (9 of 11 patients until May 12). A new motif, GATTC, was found in 2 patients outside the cluster. In addition, no intrapatient variation was observed in the 5 amplicons from specimens collected at different times or from different sources (sputum or stools).

The possible role of genetic mutations in patients’ prognosis was also investigated. The presence of nucleotide substitution was compared between 2 groups of patients: 1

| Table 1. Primers used for nested polymerase chain reaction and sequencing |
|-----------------------------|-----------------------------|-----------------------------|
| **Position** | **Amplification region** | **Primer sets (starting from 5')** |
| 17564 | 17440–18281 | Forward: ACGTCTATATCGCAGCTTCTGTGCAGACTTTAATGAAAAACATAA Revert: GTTTTCGATTTAAACTCTTGTTGTTAGGACACAGCATCCTAGT |
| 21721 | 21585–22304 | Forward: GATGATGTTCAAGGCTCTAAAATCAGAGACTTTGATTGGTTCAG Revert: CAACATACTTCTATCTAGGGGTACACCATTCTTATCATCCTTTATGAG |
| 22222 | 22177–22874 | Forward: AGATGATGTTCTGTTGTCAGTGTCAGCTTCTCAATCCGCTAAGTACACCATTAAGAAGAGAGGAGCGATTTGGAATGTCG |
| 23823 | 23455–24263 | Forward: CAACATTAGCCATTTCTACGGAAGGAGAAGAGAGGAGCGATTTGGAATGTCG Revert: GATCAGCTTCTTACTGAGAGAGAGAGAGAGAGCGATTTGGAATGTCG |
| 27827 | 27449–28270 | Forward: TAGAACGACTTCTTCTGGCTCGGCATATTGGAAGGAGAAGAGAGGAGCGATTTGGAATGTCG Revert: GCATCGCGATTACGTGAGAGAGAGAGAGAGGAGCGATTTGGAATGTCG |

*The nucleotide position was given with TOR2 as the reference strain (accession no. NC004718).
with good prognosis (absence of pulmonary aggravation; n = 25) and 1 with adverse outcome (pulmonary aggravation 8–12 days after onset of symptoms requiring oxygen ventilation or transfer to ICU; n = 4). No mutation was found associated with disease severity (Table 2).

Discussion

During the 2003 SARS epidemic, conventional epidemiologic investigation, aided by viral sequencing analysis, identified viral genetic signatures that are linked to geographic and temporal clusters of infection (4,10–12, 15–18). Findings of these studies are summarized in the Figure, connecting the worldwide epidemic to a transmission event in hotel M in Hong Kong in late February 2003. Beijing had experienced the SARS epidemic from March to June; however, only a few Beijing strains from the early epidemic have been analyzed in previous studies. Our study is the first to provide phylogenetic information on Beijing strains from the early and middle epidemic, as well as the late epidemic, by using the 5-locus motif of previous studies. The series of mutations in the 5-locus motif observed in Beijing followed the same path as isolates in Guangdong Province and the worldwide epidemic, i.e., the early introduction of GACTC motif was followed by transition to a GGCTC motif, before switching to a stable TGTTT motif. The observation of the same series of mutations occurring in 2 separate locations at different times suggests a dominant process of viral adaptation to the host.

| Patient no. (sex, age [y]) | Onset date† | Sampling date,† clinical sample | 5-loci genotype | Other variant loci |
|--------------------------|-------------|--------------------------------|----------------|-------------------|
| 13 (M, 25)               | 3/10        | 4/28, Sp                        | GGCTC          | 22589             |
| 24 (F, 68)               | 3/21        | 4/28, Sp                        | GGCTC          | 22589, 27749      |
| 3 (M, 19)                | 3/31        | 4/28, Sp; 5/5, Sp               | GGCTC          | 22589             |
| 4T (F, 34)               | 3/31        | 5/5, St; 4/28, Sp               | TGTTT          | 17620, 22589      |
| 5H (M, 21)               | 3/31        | 4/28, Sp; 5/5, Sp               | TGTTT          | 22589             |
| 6T (F, 34)               | 4/2         | 4/28, Sp                        | TGTTT          | 22077             |
| 7 (F, 27)                | 4/2         | 4/28, Sp                        | GGCTC          | 22589             |
| 8T (M, 31)               | 4/3         | 4/28, Sp                        | TGTTT          | 22077, 22589, 27749 |
| 9H (M, 20)               | 4/5         | 5/5, Sp; 4/28, Sp               | TGTTT          | 17620, 22077, 22589 |
| 10 (F, 23)               | 4/8         | 5/22, St; 5/15, Sp              | GGCTC          | 22589             |
| 11H (M, 47)              | 4/8         | 5/5-Sp; 5/5, St                 | GGCTC          | 22589             |
| 12 (M, 73)               | 4/9         | 4/28, Sp; 4/28, St              | TGTTT          | 22589, 27749      |
| 13 (M, 54)               | 4/9         | 4/28, Sp; 4/28, St              | GGCTC          | 22589, 27749      |
| 14T (F, 21)              | 4/11        | 4/28, Sp; 4/28, St; 5/5, Sp    | TGTTT          |                  |
| 15H (M, 61)              | 4/12        | 5/22, Sp                        | GATTC          | 22589             |
| 16T (F, 25)              | 4/15        | 5/5, Sp                         | TGTTT          | 17620, 22589, 27749 |
| 17 (M, 25)               | 4/17        | 5/5, Sp                         | TGTTT          | 22077, 22589, 27749 |
| 18 (F, 20)               | 4/18        | 4/28, Sp; 4/28, St              | TGTTT          | 22589, 27749      |
| 19T (F, 25)              | 4/20        | 5/5, Sp; 5/5, St                | TGTTT          | 22077, 22589      |
| 20 (F, 34)               | 4/21        | 4/28, Sp; 5/5, Sp               | TGTTT          | 22077             |
| 21 (M, 33)               | 4/23        | 5/5, Sp; 5/5 St; 4/28-Sp        | TGTTT          | 17620, 22589      |
| 22 (M, 28)               | 4/24        | 5/5, Sp                         | TGTTT          | 22589, 27749      |
| 23 (M, 61)               | 5/1         | 5/15, Sp                        | GATTC          | 22589             |
| 24T (F, 31)              | 5/1         | 5/5, Sp                         | TGTTT          | 22077, 2589, 7749 |
| 25 (M, 25)               | 5/2         | 5/5, Sp                         | TGTTT          | 22077             |
| 26 (M, 25)               | 5/4         | 5/22, Sp                        | TGTTT          | 22077             |
| 27 (M, 19)               | 5/6         | 5/22, Sp                        | TGTTT          | 22589, 27749      |
| 28 (F, 28)               | 5/7         | 5/22, Sp                        | TGTTT          | 17620, 22589, 27749 |
| 29 (M, 22)               | 5/12        | 4/28, Sp; 5/5, Sp              | TGTTT          | 22589, 27749      |

*F, female; M, male; Sp, sputum; St, stool.
†All dates are in 2003.
‡Patients were from the same cluster.
§Patients with adverse clinical outcome.

Table 3. Characterization of nucleotide (nt) substitutions in 29 severe acute respiratory syndrome patients, Beijing, China*

| ORF or protein | Position† | nt substitution | aa change | No. patients |
|---------------|-----------|-----------------|-----------|--------------|
| ORF 1b        | 17620     | C→T             | Leu→Ser   | 5            |
| S protein     | 22077     | G→T             | Phe→Tyr   | 9            |
| S protein     | 22589     | C→T             | Noncoding region | 24          |
| ORF 9         | 27749     | G→A             | Lys→Glu   | 12           |

*SARS, severe acute respiratory syndrome; ORF, open reading frame; aa, amino acid.
†The nt positions are numbered with TOR2 as reference strain (accession no. NC_004718).
Moreover, this finding can expand our understanding of SARS-CoV response to selection pressures in humans, since early Beijing isolates (BJ01, BJ02, and BJ03), which are traceable to Guangdong, underwent an independent selection process and would not be subject to the same sampling bias caused by superspreading events in Hong Kong isolates. The GGCTC→TGTTT switch was observed among patients belonging to the same cluster in this study, which rules out the possibility of the coincidental superposition of 2 epidemics (GGCTC and TGTTT) coexisting in Beijing.

The mutations involved in the GGCTC→TGTTT switch are responsible for amino acid changes in a non-structural protein (17564, region Orf1b) in S protein (21721 and 22222) and in a noncoding region (27827, X3). We were not able to identify a correlation between these changes and the clinical status of patients. We did not find sequence variations in specimens obtained from the same patients either collected at different times or among different specimen types, which suggests that within-individual variations are rare in the partial genome of this study, although the phenomenon was described in a previous study (15). A new motif, GATTC, which represents a new transitional motif between GACTC and TGTTT, was described on 2 occasions in patients who were not part of the cluster. Similarly, 4 new SNVs were identified at nucleotides 17620, 22077, 22589, and 27749.

In summary, this study confirms the evolution of SARS-CoV strains towards a TGTTT motif in positions 17564, 21721, 22222, 23823, and 27827 in Beijing, as was observed in Guangdong province before the hotel M outbreak in Hong Kong. Whether this motif is associated with higher transmission or virulence remains to be elucidated.

Acknowledgments

We thank Guo-Ping Zhao, Huai-Dong Song, and Guo-Wei Zhang for their assistance with this study.

This work was partly supported by the EC grant EPISARS (511063), the Programme de Recherche en Réseaux Franco-Chinois (P2R), the National Institutes of Health CIPRA Project (NIH U19 AI51915), and the National 863 Program of China (2003AA208406, 2003AA208412C).

Dr Liu is an epidemiologist in the Department of Epidemiology, Beijing Institute of Microbiology and Epidemiology. Her primary research interests are molecular epidemiology and emerging infectious disease.

References

1. World Health Organization. SARS epidemiology to date [monograph on the Internet]. 2003. [cited 2003 Apr 11]. Available from: http://www.who.int/csr/sars/country/en/
2. World Health Organization. Multicentre Collaborative Network for Severe Acute Respiratory Syndrome (SARS) Diagnosis. A multicentre collaboration to investigate the cause of severe acute respiratory syndrome. Lancet. 2003;361:1730–3.
3. World Health Organization. Cumulative number of reported probable cases of severe acute respiratory syndrome (SARS) [monograph on the Internet]. [cited 2003 Jul 11]. Available from: http://www.who.int/csr/sars/country/en/
4. Guan Y, Peiris JS, Zheng B, Poon LL, Chan KH, Zeng FY, et al. Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome. Lancet. 2004;363:99–104.
5. Pang X, Zhu Z, Xu F, Guo J, Gong X, Liu D, et al. Evaluation of control measures implemented in the severe acute respiratory syndrome outbreak in Beijing, 2003. JAMA. 2003;290:3215–21.
6. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med. 2003;348:1967–76.
7. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science. 2003;300:1394–9.
8. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 2003;348:1953–66.
9. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, et al. The genome sequence of the SARS-associated coronavirus. Science. 2003;300:1399–404.
10. The Chinese SARS Molecular Epidemiology Consortium. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science. 2004;303:1666–9.
11. Zhong NS, Zheng BJ, Li YM, Poon LLM, Xie ZH, Chan KH, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People’s Republic of China, in February, 2003. Lancet. 2003;362:1533–8.
12. Ruan YJ, Wei CL, Ee AL, Vega VB, Thoreau H, Su ST, et al. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet. 2003;361:1779–85.

Figure. Epidemiologic and phylogenetic links between patients of different worldwide SARS outbreaks (4,10,11,12). New information that concerns the Beijing epidemic is represented in boldface. Epidemiologic links that are still speculative are in dotted lines.
13. 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;349:187–8.
14. World Health Organization. Case definitions for surveillance of severe acute respiratory syndrome (SARS). [cited 2003 Apr 29]. Available at: http://www.who.int/csr/sars/casedefinition/en.
15. Xu DP, Zhang Z, Chu F, Li Y, Jin L, Zhang L, et al. Genetic variation of SARS coronavirus in Beijing hospital. Emerg Infect Dis. 2004;10:789–94.
16. Yeh SH, Wang HY, Tsai CY, Kao CL, Yang JY, Liu HW, et al. Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: molecular epidemiology and genome evolution. Proc Natl Acad Sci U S A. 2004;101:2542–7.
17. Tsang KW, Ho PL, Ooi GC, Yee WK, Wang T, Chan-Yeung M, et al. A cluster of cases of severe acute respiratory syndrome in Hong Kong. N Engl J Med. 2003;348:1977–85.
18. Wang Z, Li L, Luo Y, Zhang J, Wang M, Cheng S, et al. Molecular biological analysis of genotyping and phylogeny of severe acute respiratory syndrome associated coronavirus. Chin Med J. 2004;117:42–8.

Address for correspondence: Wu-Chun Cao, Beijing Institute of Microbiology and Epidemiology, State Key Laboratory of Pathogen and Biosecurity, Beijing, People’s Republic of China; fax: 86-10-63812060; email: caowc@nic.bmi.ac.cn