

# Genome sequence variation analysis of two SARS coronavirus isolates after passage in Vero cell culture

JIN Weiwu<sup>1</sup>, HU Liangxiang<sup>2</sup>, DU Zhenglin<sup>1</sup>, GAO Qiang<sup>2</sup>, GAO Hong<sup>2</sup>, NING Ye<sup>2</sup>, FENG Jidong<sup>1</sup>, ZHANG Jiansan<sup>2</sup>, YIN Weidong<sup>2</sup> & LI Ning<sup>1</sup>

1. State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing 100094, China

2. Sino Biotech Co., Ltd., No. 39 Shangdi Xi Rd. Haidian District, Beijing 100085, China

Correspondence should be addressed to Li Ning (e-mail: ningbau@public3.bta.net.cn)

**Abstract** SARS coronavirus is an RNA virus whose replication is error-prone, which provides possibility for escape of host defenses, and even leads to evolution of new viral strains during the passage or the transmission. Lots of variations have been detected among different SARS-CoV strains. And a study on these variations is helpful for development of efficient vaccine. Moreover, the test of nucleic acid characterization and genetic stability of SARS-CoV is important in the research of inactivated vaccine. The whole genome sequences of two SARS coronavirus strains after passage in Vero cell culture were determined and were compared with those of early passages, respectively. Results showed that both SARS coronavirus strains have high genetic stability, although nearly 10 generations were passed. Four nucleotide variations were observed between the second passage and the 11th passage of Sino1 strain for identification of SARS inactivated vaccine. Moreover, only one nucleotide was different between the third passage and the 10th passage of Sino3 strain for SARS inactivated vaccine. Therefore, this study suggested it was possible to develop inactivated vaccine against SARS-CoV in the future.

**Keywords:** severe acute respiratory syndrome (SARS), SARS Coronavirus (SARS-CoV), virus, vaccine.

DOI: 10.1360/04wc0022

Severe acute respiratory syndrome (SARS), which first occurred in November, 2002, in Guangdong Province of China, is a life-threatening respiratory disease. By July 2, 2003, in the world, 8442 people were infected and 814 people died, with a mortality rate as high as 9.6%, according to WHO (<http://cmbi.bjmu.edu.cn/cmbidata/SARS/WHO/0702.mht>). The SARS causative pathogen is a member of the family coronaviridae, which has a single-stranded positive-sense RNA genome. The sequenced full-length genome of SARS-CoV is about 30 kilo-bases in length and has more than 10 open reading frames. And the genome organization is similar to that of other coronavirus. However, phylogenetic analyses and sequence

comparisons showed that SARS-CoV is not closely related to any of the previously characterized coronavirus<sup>[1-3]</sup>.

Although having a number of unique features, SARS-CoV is most closely related to group 2 coronavirus, and was an early split-off from group 2 branch<sup>[4]</sup>. Our previous study also indicated that SARS-CoV did not resemble to any other known coronavirus, including IBV Beijing isolate which was once epidemic in China<sup>[5]</sup>. However, the early infection of SARS-CoV may have originated from wild animals. SCoV-like viruses were isolated from Himalayan palm civets and genomes were identical to those of human isolates, but a 29-nt sequence found in the animal isolates was not detected in most human isolates. Other studies also showed that five strains from human contained this 29-nt segment, and 82-nt deletion or 415-nt deletion in the *orf8* region were observed in some human isolates and an animal isolate<sup>[6, 7]</sup>.

SARS-CoV, like HIV, is an RNA virus which is prone to high mutation rate for its error replication to escape host defenses. 127 base substitutions were identified among 14 SARS isolates, and 94 of them changed their amino acid sequences<sup>[8, 9]</sup>. In addition, 299 single nucleotide variations were detected among 63 sequences. The neutral mutation rate of SARS-CoV is similar to that of known RNA virus and is about one-third of that for the human immunodeficiency virus<sup>[7]</sup>. Molecular epidemiological approach and sequencing analysis of SARS-CoV viruses isolated from 10 Taiwanese patients suggested a recent introduction of the viruses into human populations. And the estimated genome mutation rate was about 0.1 per genome, demonstrating possibly one of the lowest rates among known RNA viruses<sup>[10]</sup>. A full-length cDNA of the SARS-CoV Urbani strain was assembled and molecularly cloned SARS viruses were rescued by using five contiguous cDNAs fragments that span the entire genome. Recombinant viruses replicated as efficiently as wild type virus. And the full-length cDNA of the SARS-CoV allowed for the development and testing of candidate vaccines, especially for the development of live attenuated vaccine<sup>[11]</sup>. In this work, the whole genome sequences of two SARS-CoV viruses after passage in Vero cell culture were determined to develop inactivated vaccine.

## 1 Materials and methods

(i) Virus. Two SARS-CoV strains were obtained from the Chinese Academy of Medical Sciences and Peking Union Medical College and renamed Sino1 strain and Sino3 strain by Sino Biotech Co., Ltd, respectively. Both viral strains were originally isolated by Peking Union Hospital from nasal swabs of two SARS patients and were confirmed as SARS-CoV by PCR and electron microscopy. And the complete genomes of them were sequenced and submitted to the GenBank, named PUMC03

strain and PUMC02 strain respectively. In this study, PUMC03 strain (GenBank: AY357076) is the second passage of Sino1 strain (Sino1-2 isolate), while PUMC02 strain (GenBank: AY357075) is the third passage of Sino3 strain (Sino3-3 isolate).

Both SARS-CoV strains were propagated on Vero cells in a flask at 37°C until 75% cells showed cytopathic effect. After three freeze-thaw cycles, viruses were harvested and well characterized. The batch of virus was stored at -80°C as seeds or reference viruses for neutralization assay. Vero cells were obtained from American type culture collection (ATCC) and were grown at 37°C in minimum essential medium (MEM), containing 10% (*w/v*) fetal calf serum and 1% glutamine. Finally, the 11th passage of Sino1 strain (Sino1-11 isolate) and the 10th passage of Sino3 strain (Sino3-11 isolate) were obtained after subsequent passage in Vero cell culture. Virus was collected, and Trizol LS reagent (Invitrogen, USA) was added for viral RNA extraction. All work was finished at a biosafety level 3.

(ii) Viral RNA extraction and RT-PCR. Samples were removed from BSL-3 and viral genomic RNA was extracted under BSL-2 conditions. The random hexameric primer (Promega, USA) was used in cDNA first-strand synthesis. After cDNA first-strand was synthesized, 74 specific primer pairs were used in the reaction of PCR. Rapid amplification of cDNA ends (RACE) was performed to capture the 3'-end and 5'-end of the viral genome. Other three primers were designed for RACE (one for 3' RACE and the other two for 5' RACE).

(iii) Sequencing and analysis. RACE products and some PCR products were cut from 1% agarose gels and purified using the GeneClean III Kit (Qbiogene). Others were purified by MultiScreen-PCR Plates (Millipore). Each PCR fragment was directly sequenced from inward and outward directions. And automated sequencing with BigDye Terminator kit (Perkin Elmer) was run on an ABI 377 DNA sequencer. We used the Phred/ Phrap/ Consed package to process all the raw sequence reads for base calling, assembly, and editing. Sequence regions with poor quality were resequenced from purified PCR fragments. Alignments of genome sequences of SARS-CoV were performed with Clustal W or on the line ([http://](http://www.ncbi.nlm.nih.gov/blast/bl2seq)

[www.ncbi.nlm.nih.gov/blast/bl2seq](http://www.ncbi.nlm.nih.gov/blast/bl2seq)). Amino acid sequences of seven proteins also aligned with Clustal X1.81.

## 2 Results and discussion

In total, for each isolate, 76 fragments were obtained from RT-PCR and RACE, which span the entire genome. All component sequences were determined in duplicate and the average length of them was 420 bases. Therefore, for any region of the genome there were four-fold to six-fold coverage, and there were more than 20 bp for every overlap. Finally, the complete genome sequence was assembled and the genome of the 10th passage of Sino3 strain (Sino3-11 isolate, GenBank: AY485278) was 29740 nucleotides in length, while that of the 11th passage of Sino1 strain (Sino1-11 isolate, GenBank: AY485277) was 29741 bp. Compared with Tor2 isolate (GenBank: NC\_0004718), nucleotide deletions were kept in both viral strains during passage in cell culture (2-nt deletion in Sino3 strain, 3-nt deletion in Sino1 strain, respectively).

(i) Comparison of genome sequences. After alignments of the full-length genomes of these SARS-CoV isolates were finished, both Sino1-2 isolate and Sino3-3 isolate were found to have 14-nt deletion at the end of the 5'-end of genome, when compared to those of other three SARS-CoV (Sino3-11 isolate, Sino1-11 isolate and Tor2 isolate). Except these deletions, no other deletion or insert was detected between early passage and late passage of two virus strains. We think that the 5'-end of Sino1-2 and Sino3-3 isolate genomes could not be captured completely. In addition, four single nucleotide variations were detected between the second passage and the 11th passage of the Sino1 strain. For Sino3 strain, the genome of passage3 differs that of passage10 only by one nucleotide. The reversion at nucleotide position 6574 from G to A in Sino1-11 isolate might be the result of a back-mutational event occurring potentially during the passage of the virus. Other studies also showed that there existed few genetic variations between primary samples and viruses isolated after limited passages of cultures from two SARS patients. The two sequences from patient 1 were completely identical, but few single nucleotide variations were detected between the two sequences from

Table 1 Nucleotide sequence variations of the complete genome of five SARS-CoV isolates

Position*	6505	6574	10168	16006	16890	22816	23220	25298	26129	26156	26157	26158	26249	26477	27883	27884
orf*	<i>orf1a</i>	<i>orf1a</i>	<i>orf1a</i>	<i>orf1b</i>	<i>orf1b</i>	<i>s</i>	<i>s</i>	<i>3a</i>	<i>3b/e</i>	<i>e</i>	<i>e</i>	<i>e</i>	<i>e</i>	<i>m</i>	<i>orf8</i>	<i>orf8</i>
Tor2	A	A	T	A	C	A	G	A	G	G	T	T	A	T	T	A
Sino1-2	A	<b>G</b>	T	A	<b>T</b>	A	<b>T</b>	<b>G</b>	G	-	-	-	<b>T</b>	<b>G</b>	T	A
Sino1-11	A	<b>A</b>	<b>G</b>	<b>C</b>	<b>T</b>	<b>C</b>	<b>T</b>	<b>G</b>	G	-	-	-	<b>T</b>	<b>G</b>	T	A
Sino3-3	A	A	T	A	C	A	<b>T</b>	<b>G</b>	<b>T</b>	G	T	T	A	<b>G</b>	-	-
Sino3-11	<b>T</b>	A	T	A	C	A	<b>T</b>	<b>G</b>	<b>T</b>	G	T	T	A	<b>G</b>	-	-

Nucleotide positions and organization of open reading frames based on Tor2 sequence. Pane represent nucleotide mutation occurred during the passage of viruses in cell culture. The bold letters stand for nucleotide difference from the Tor2 isolate. Dashes indicate a nucleotide deletion.

# ARTICLES

patient 2<sup>[10]</sup>, showing a high genetic stability during the course of passage in Vero cell culture.

On the other hand, more or less variations were observed in the transmission of SARS. Genomic sequencing for viral isolates from five Amoy Gardens patients were finished and the virus sequence was identical in four of these five patients, but two single nucleotide variations existed in sequence from patient 2<sup>[12]</sup>. However, studies of five SARS isolates from an index patient, three primary contacts of the index patient and one secondary contact related to the index patient, suggested that more mutations occurred during the transmission to escape from host immune response. Two sequences from primary contacts were found to have 5-nt deletion and 6-nt deletion in a non-coding region, respectively. Besides the deletion, there were nucleotides substitutions among the five sequences<sup>[8]</sup>. Virus passage in cell culture has fewer mutations than the transmission. Because of a stable condition in cell culture and non-selective pressure from host, the mutations of each point in the viral genome are low and similar, which result in the most variations for *orf1ab*, for *s*, then for *n* and so on. In our study, four of five single nucleotide variations were found in the region encoding Pp1ab which contains 2/3 the length of the whole genome. And the remainder 20% was in the region encoding S glycoprotein.

(ii) Comparison of amino acid sequences for major proteins. Alignments of the amino acid sequences for seven major proteins (Pp1ab, S, 3a, 3b, E, M, N) were performed and the variations were found (5 for Pp1ab, 2 for S, 3 for E, 1 for 3b, M, and none for 3b, N, respectively). The variations of S glycoprotein, E protein and M protein are shown in Fig. 1.

Among five amino acid variations in the polyprotein, three of them located in Orf1a, and the others lay in Orf1b. As far as S glycoprotein is concerned, the amino acid sequence of S1 subunit is more variable than that of S2 subunit and more variations were detected in S1. Alignments of 14 different sequences of S indicated that most of the amino acid mutations in S glycoprotein are located in the S1 subunit (1—680)<sup>[13]</sup>. In this study, we also found that both amino acid variations in S glycoprotein lay in S1 domain. In addition, the 3-nt deletion in the Sino1 strain genome caused an amino acid deletion in E protein. And the 2-nt deletion in the Sino3 strain genome was in the region encoding 8b protein and caused a stop of translation. Deletions occurred in the *orf8* region with no apparent effect on the survival of the virus, suggesting that this region is either noncoding or coding for a functionally unimportant putative protein<sup>[7]</sup>. On the other hand, five base substitutions during the passage of virus in cell culture also changed the amino acid sequence, also showing

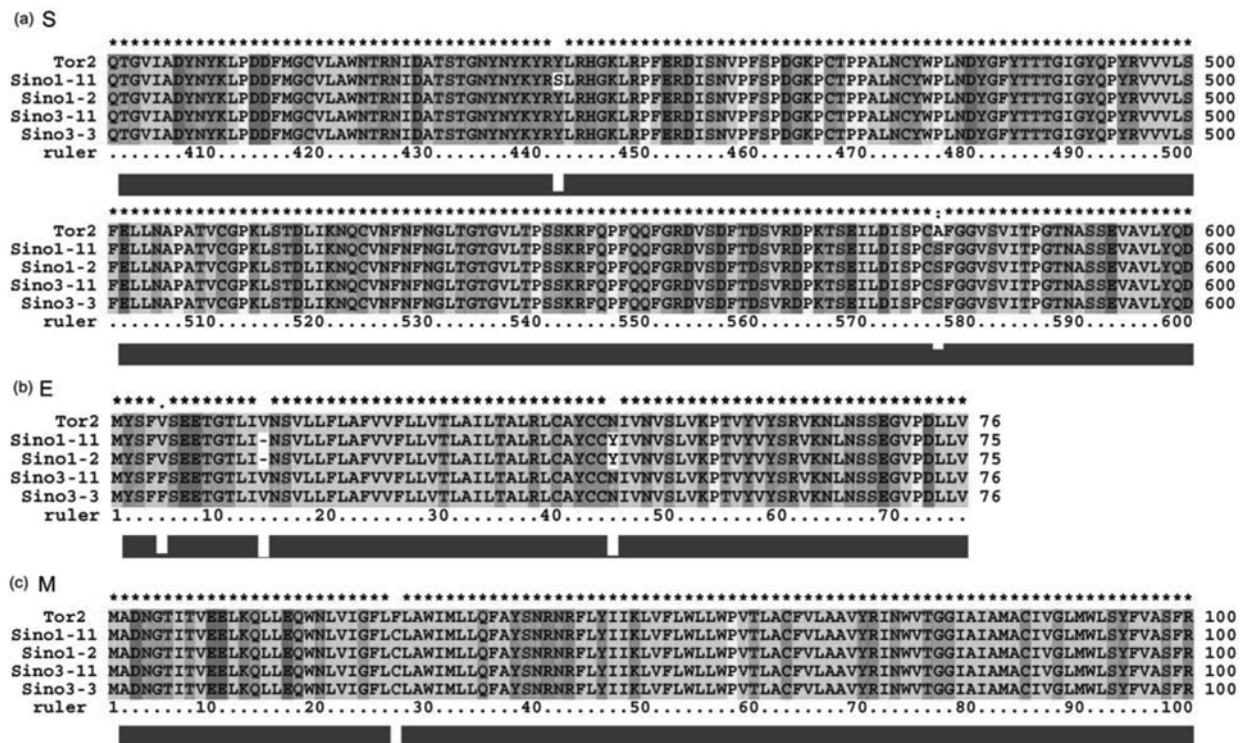


Fig. 1. Distribution of amino acid variations of S glycoprotein (a), E protein (b) and M protein (c). Only partial sequences of S protein and M protein with variations were shown here. Asterisks above the sequence indicate that all amino acids were identical. Ruler under the sequence was used to determine the position of amino acid.

that the rates of nonsynonymous changes were greater than those of synonymous changes for the *s* gene sequences and *orf1a* sequences<sup>[7, 10]</sup>.

Although amino acid sequences of two SARS-CoV isolates were changed after passage, no signs showed remarkable change in its immunogenicity and antigenicity. And the viral titer was at the same level before passage. In all, SARS-CoV passage in cell culture has high stability.

### 3 Conclusion

A number of variations were found among separate SARS-CoV viruses. A study of these mutations will help develop effective and safe vaccines<sup>[7, 8]</sup>. Moreover, the low mutation rate of SARS-CoV suggested that development of a vaccine against the SARS-CoV in future would be less difficult than expected before<sup>[7, 10]</sup>.

In addition, in this study, the Sino3 strain was used to develop inactivated vaccine, and the Sino1 strain was used for testing the candidate vaccine. It was very important to know the genetic stability of them. The complete genomes of these two SARS-CoV isolates, which repassaged in Vero cell culture, were sequenced and analyzed. And the result indicated a high genetic stability for SARS-CoV passage in Vero cell culture, which implies that there is high possibility to develop inactivated vaccine against the SARS-CoV.

**Acknowledgements** The author would like to thank Beijing Genomics Institute, Chinese Academy of Sciences for providing 74 specific primer pairs, and also to thank Chinese Academy of Medical Sciences and Peking Union Medical College for their isolated SARS-CoV virus and the complete genomes sequencing data. This work was partially supported by SARS Coronavirus Program of the Ministry of Education of China and the State High-Tech Research and Development Program (Grant No. 2003AA208201).

### References

1. Rota, P. A., Oberste, S. M., Monroe, S. S. et al., Characterization of a novel coronavirus associated with Severe Acute Respiratory Syndrome, *Science*, 2003, 300: 1394—1399.
2. Marra, M. A., Jones, S. J., Astell, C. R. et al., The genome sequence of the SARS-associated coronavirus, *Science*, 2003, 300: 1399—1404.
3. Qin, E. D., Zhu, Q. Y., Yu, M. et al., A complete sequence and comparative analysis of a SARS-associated virus (Isolate BJ01), *Chinese Science Bulletin*, 2003, 48 (10) : 941—948.
4. Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C. et al., Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage, *J. Mol. Biol.*, 2003, 331: 991—1004.
5. Jin, W. W., Chen, C., Zhang, Y. et al., Genome sequencing and characterization analysis of a Beijing isolate of chicken coronavirus infectious bronchitis virus, *Chinese Science Bulletin*, 2004, 49(6): 585—590.
6. Guan, Y., Zheng, B. J., He, Y. Q. et al., Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China, *Science*, 2003, 302: 276—278.
7. 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—1669.
8. Ruan, Y. J., Wei, C. L., Ling, A. E. 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—1785.
9. De Groot, A. S., How the SARS vaccine effort can learn from HIV—speeding towards the future, learning from the past, *Vaccine*, 2003, 21: 4095—4104.
10. Yeh, S. H., Wang, H. Y., Tsai, C. Y. et al., Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: Molecular epidemiology and genome evolution, *PNAS*, 2004, 101: 2542—2547.
11. Yount, B., Curtis, K. M., Fritz, E. A. et al., Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus, *PNAS*, 2003, 100: 12995—13000.
12. Chim, S. S. C., Tsui, S. K. W. Hung, E. C. W. et al., Genomic characterization of the severe acute respiratory syndrome coronavirus of Amoy Garden outbreak in Hongkong, *Lancet*, 2003, 362: 1807—1808.
13. Spiga, O., Bernini, A., Ciutti, A. et al., Molecular modeling of S1 and S2 subunits of SARS coronavirus spike glycoprotein, *Biochem. Biophys. Res. Co.*, 2003, 310: 78—83.

(Received May 15, 2004; accepted July 24, 2004)