©2003, Acta Pharmacologica Sinica
Chinese Pharmacological Society
Shanghai Institute of Materia Medica
Chinese Academy of Sciences
http://www.ChinaPhar.com

Mutation analysis of 20 SARS virus genome sequences: evidence for negative selection in replicase ORF1b and spike gene[1]

HU Lan-Dian[2], ZHENG Guang-Yong[2], JIANG Hai-Song[2], XIA Yu[2], ZHANG Yi[2], KONG Xiang-Yin[2,3,4]

[2]Health Science Center, Shanghai Institutes for Biological Sciences,
Chinese Academy of Sciences and Shanghai Second Medical University, Shanghai 200025;
[3]State Key Lab for Medical Genomics, Rui Jin Hospital, Shanghai Second Medical University, Shanghai 200025, China

KEY WORDS severe acute respiratory syndrome; SARS-CoV virus; negative selection; replicase polyprotein; spike protein

ABSTRACT

AIM: Recently, more SARS-CoV virus genome sequences are released to the GenBank database. The aim of this study is to reveal the evolution forces of SARS-CoV virus by analyzing the nucleotide mutations in these sequences. **METHODS:** We obtained 20 SARS-CoV virus genome sequences from NCBI database, and calculated the ratio of non-synonymous nucleotide substitution per non-synonymous site (K_a) and synonymous nucleotide substitution per synonymous site (K_s) for SARS-CoV virus genes. **RESULTS:** The K_a/K_s ratios for replicase polyprotein ORF1a, ORF1b, and spike protein gene are 1.09 (P=0.6501), 0.38 (P=0.0074), 0.65 (P=0.0685) respectively. **CONCLUSION:** SARS-CoV virus replicase polyprotein ORF1b is undergoing negative selection; negative selection force is also probably operating on spike protein gene. These results provide basis for future developing a new drug and vaccine against SARS.

INTRODUCTION

Severe Acute Respiratory Syndrome (SARS) is a global outbreak disease, epidemic from November, 2002. The pathogen has been discovered as a novel coronavirus (SARS-CoV). SARS-CoV is a 30 kb ssRNA positive-strand virus[1-6]. Similar to other known coronaviruses, the viral RNA genome has five major open reading frames (ORFs) and additional nine potential ORFs. These ORFs-encoded proteins include the replicase polyprotein, the spike (S), envelope (E), and membrane (M) glycoproteins and the nucleocapsid protein (N)[2-5]. Sequence analysis reveals that the SARS-CoV virus is distinct from all known human viruses. Therefore, SARS-CoV virus is unlikely the mutant or recombinant of any known human coronaviruses, instead, probably jumps to human population from an unknown source[2-5]. Recently, a coronavirus resembling the SARS virus has been detected in palm civets (Paguma larvata) and a raccoon dog (Nyctereutes procyonoides)[7]. However, at present, it is uncertain whether these animals are the exact origin of human

[1] Project supported by National High Technology "863" Programs of China, National Natural Science Foundation of China, National Science Fund for Distinguished Young Scholars, State Key Technologies R&D programme (973).

[4] Correspondence to Prof KONG Xiang-Yin.
Phn 86-21-6467-8976. Fax 86-21-6467-8976.
E-mail xykong@sibs.ac.cn

Received 2003-06-22 Accepted 2003-07-07

SARS-CoV virus.

After jumping to human, to adapt itself to the new host and to avoid the host immune system, SARS-CoV virus should be under evolutionary selection. The decoding of the virus complete genome sequence and identification of its encoded proteins provide the basis for evolutionary analysis of the virus genome. Sequence comparison of 14 SARS-CoV virus genome sequences revealed the common origins of human SARS-CoV viruses[8]. Although the SARS-CoV virus is relatively stable[8], the mutations within virus genome are not even distributed. This indicates that some genes may mutate rapidly than others. Identifying the undergoing evolution process of these genes will be helpful in virus detection and therapy.

Worldwide SARS research accelerates the progress of SARS-CoV virus genome sequencing. So far, 20 SARS-CoV virus complete genome sequences have been released to GenBank. Thus, it is possible to obtain more mutations of SARS-CoV virus genome from these sequences. These mutations harbor information about the virus-host interaction during the past half year epidemic. In this study, we investigate the evolution of virus genes, particular the replicase polyprotein gene and spike protein gene.

MATERIALS AND METHODS

SARS-CoV virus genome sequences The 20 SARS-CoV virus complete genome sequences were from GenBank (http://www.ncbi.nlm.nih.gov/).

Multiple sequence alignment and phylogenetic analysis We performed multiple sequence alignment and constructed consensus neighbor-joining tree of the 20 SARS-CoV virus genome sequences using the free online ClustW programm (http://www.ebi.ac.uk/clustalw/).

Prediction of transmembrane helices We used TMHMM Server v. 2.0 to predict the transmembrane helices of membrane glycoprotein (http://www.cbs.dtu.dk/services/TMHMM-2.0/).

Non-synonymous nucleotide substitution per nonsynonymous site (K_a) and synonymous nucleotide substitution per synonymous site (K_s) analysis We calculated the synonymous sites and non-synonymous sites, synonymous and nonsynonmous mutations using the DnaSP 3.51 programm[9]. We used Fisher' s exact test to calculate the P value under null hypothesis of equal rates of synonymous and non-synonymous changes[10].

RESULTS AND DISCUSSION

Mutation distribution SARS-CoV virus spread to human population half years ago, then quickly broke out in the world. In the new host, the viruses are subjected to either positive, negative or neutral evolution forces. Positive selection often operates on genes involved in evading the defensive systems or immunity, such as the human immunodeficiency virus-1 envelope gene (env). During genome replication, SARS-CoV viruses are apt to obtain mutations by its error-prone polymerase. Greater K_a/K_s ratio characterizes positive Darwinian selection; contrariwise, low K_a/K_s ratio implies negative selection[11]. In this study, we try to explore which gene is under the force of positive selection, and which gene is under negative selection. We calculated the number of non-synonymous nucleotide substitution and synonymous nucleotide substitution that occurred in each of the five major genes in the 20 SARS-CoV virus genome sequences (Tab 1). In the coding region of the five major genes, there are totally 129 nucleotide substitutions. Multiple sequence alignment shows that some mutations seem to be cluster in these genes (Fig 1). This result suggests that these regions are undergoing rapid adaptive evolution.

Negative selection of S protein The structure proteins of SARS-CoV virus include S protein, E protein, M protein and N protein[2-5]. The first three forms the surface of the SARS-CoV viral particles. These proteins were under selection pressure from the host immune response. Due to lack of enough nucleotide substitutions (Tab 1), we could not analyze the operating evolution forces for gene E, M and N. S protein is a large protein with 1255 amino acids. It recognizes specific receptors on the surface of host cells and mediates membrane fusion[3,12]. Furthermore, it is important in determining the species specificity, tissue tropism and virulence of virus infection. We observed 14 non-synonymous nucleotide mutations and 9 synonymous nucleotide substitutions in this gene (Tab 1). The K_a/K_s value is 0.65, P value is 0.0685 (Tab 1). This result suggests that the S protein is likely under negative selection in human host; S protein is stable during human passage. Thus, S protein is an ideal vaccine component. Meanwhile, it also hints high similarity of S protein binding receptors in human and its original host.

Evolution of replicase polyprotein gene The

Tab 1. Nucleotide substitutions and K_a/K_s values of the five SARS-CoV virus genes.

	ORF1a	ORF1b	S gene	M gene	E gene	N gene
Non-synonymous substitution	49	18	14	4	1	3
Synonymous substitution	16	14	9	0	0	1
Total substitution	65	32	23	4	1	4
Length (bp)	13149	8088	3768	666	231	1269
K_a/K_s	1.09	0.38	0.65	ND	ND	ND
P value	0.6501	0.0074	0.0685	ND	ND	ND

Note. Fisher' s exact test was used to test the null hypothesis of neutral evolution.

coronavirus replicase polyprotein is a 7073-amino acid large protein, composed of two polyproteins ORF1a and ORF1b. They are translated from the virus genomic RNA sequence. The replicase polyprotein autocatalytically processes to produce a group of proteins including proteases PLPpro and 3CLpro, RNA-dependent polymerase (POL), RNA helicase (HEL) and other function unknown proteins[2-5]. These proteins are important targets for drug design[13,14]. Consistent with the observation that the product of ORF1 is relatively lower in conservation among different coronaviruses[3], the K_a/K_s ratio for ORF1a is 1.09 (P=0.6501) (Tab 1). Thus, according to the K_a/K_s ratio, ORF1a is likely evolving in a fashion of neutral evolution. However, in view of the uneven distribution of nucleotide substitutions (Fig 1), some parts of ORF1a might be under positive selection and other parts under negative selection. In fact, when analyzing the first 1667 codons, the K_a/K_s ratio reaches to 1.58 (p=0.3192). ORF1b encodes several important proteins including the RNA-dependent polymerase and the RNA helicase. Based on K_a/K_s ratio, this ORF is under negative selection (Tab 1), reflecting the functional conservation of its encoded proteins in new host.

CONCLUSION

In this study, we identified two genes, replicase polyprotein ORF1b and spike protein genes that are subject to negative selection. However, based on the current available SARS-CoV virus sequences, we could not detect positive selection effect. With the accumulation of new data, positive evolution force might be uncovered such as on M gene. In this study, we could not distinguish mutations caused by host pressure from mutations occurring in *in vitro* expansion or sequence errors[8], the current results need to be confirmed in future by analyzing SARS-COV sequences free of *in vitro* expansion mutations.

REFERENCES

1 Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, *et al.* SARS study group. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 2003; 361: 1319-25.

2 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.

3 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.

4 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.

5 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.

6 Fouchier RA, Kuiken T, Schutten M, van Amerongen G, van Doornum GJ, van den Hoogen BG, *et al.* Aetiology: Koch' s postulates fulfilled for SARS virus. Nature 2003; 423: 240.

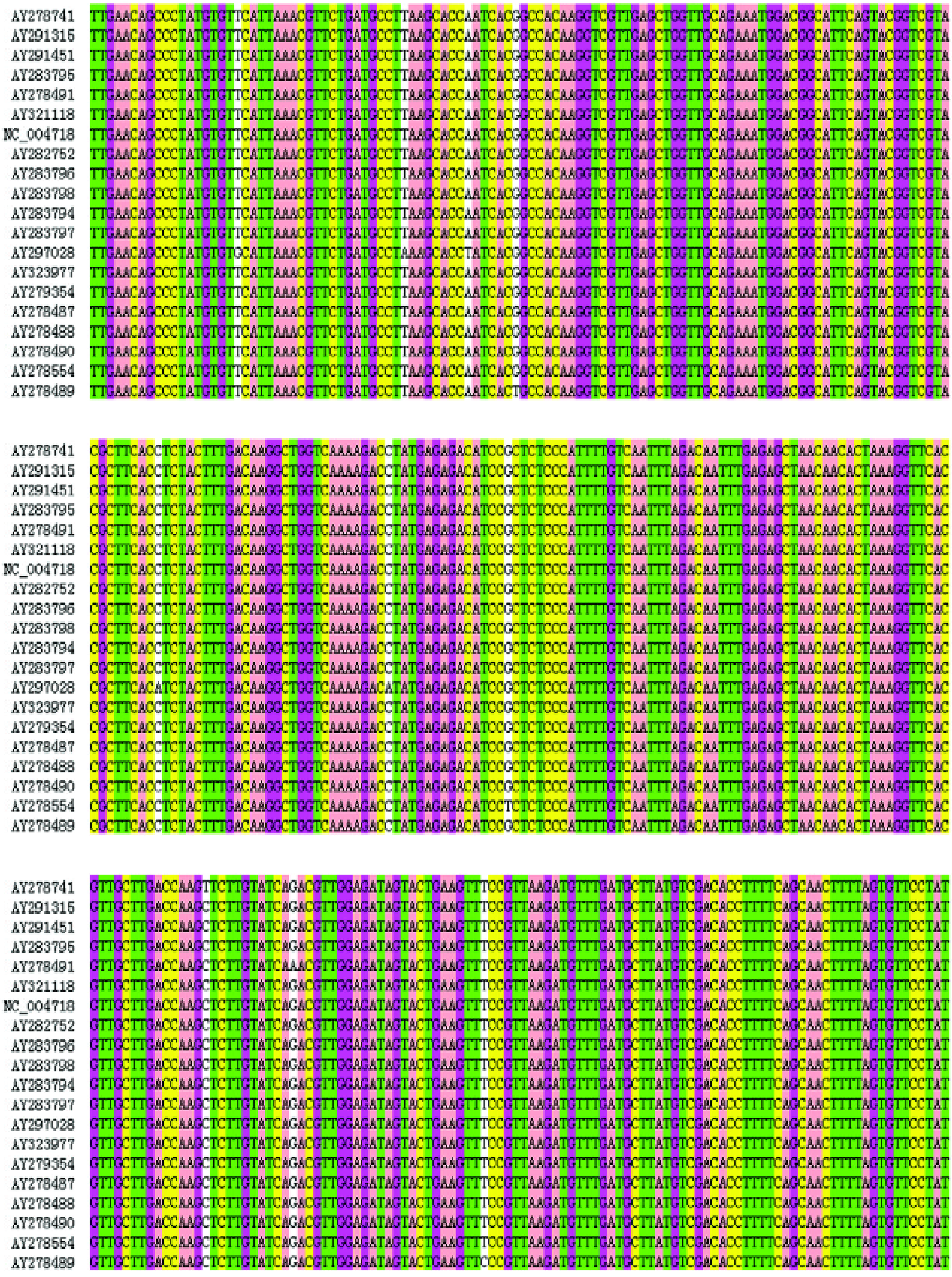
7 Enserink M. Infectious diseases. Clues to the animal origins of SARS. Science 2003; 300: 1351.

8 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.

9 Rozas J, Rozas R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 1999; 15: 174-5.

10 Zhang J, Rosenberg HF, Nei M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci US A 1998; 95: 3708-13.

11 Kimura M. The neutral theory of molecular evolution. London: Cambridge University Press; 1983.



```
 AY278741  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY291315  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY291451  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY283795  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY278491  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY321118  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
NC_004718  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY282752  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY283796  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY283798  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY283794  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY283797  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY297028  TTGAACAGCCCTATGTGTGCATTAAACGTTCTGATGCCTAAAGCACCTATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY323977  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY279354  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY278487  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY278488  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY278490  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY278554  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA
 AY278489  TTGAACAGCCCTATGTGTTCATTAAACGTTCTGATGCCTTAAGCACCAATCACTGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACGGCATTCAGTACGGTCGTA

 AY278741  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY291315  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY291451  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY283795  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY278491  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY321118  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
NC_004718  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY282752  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY283796  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY283798  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY283794  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY283797  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY297028  CGCTTCACATCTACTTTGACAAGGCTGGTCAAAAGACATATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY323977  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY279354  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY278487  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY278488  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY278490  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY278554  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCTCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC
 AY278489  CGCTTCACCTCTACTTTGACAAGGCTGGTCAAAAGACCTATGAGAGACATCCGCTCTCCCATTTTGTCAATTTAGACAATTTGAGAGCTAACAACACTAAAGGTTCAC

 AY278741  GTTGCTTGACCAAGTTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY291315  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY291451  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY283795  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY278491  GTTGCTTGACCAAGCTCTTGTATCAAACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY321118  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
NC_004718  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY282752  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY283796  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY283798  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY283794  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY283797  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY297028  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY323977  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY279354  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY278487  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY278488  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY278490  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY278554  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTTCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
 AY278489  GTTGCTTGACCAAGCTCTTGTATCAGACGTTGGAGATAGTACTGAAGTTCCCGTTAAGATGTTTGATGCTTATGTCGACACCTTTTCAGCAACTTTTAGTGTTCCTAT
```

Fig 1. Multiple SARS-CoV viruses sequence alignment showing clustered mutations in different regions of the virus genome.

12 Yu XJ, Luo C, Lin JC, Hao P, He YY, Guo ZM, *et al*. Putative hAPN receptor binding sites in SARS-CoV spike protein. Acta Pharmacol Sin 2003; 24: 481-8.

13 Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science 2003; 300: 1763-7.

14 Xiong B, Gui CS, Xu XY, Luo C, Chen J, Luo HB, *et al*. A 3D model of SARS-CoV 3CL proteinase and its inhibitors design by virtual screening. Acta Pharmacol Sin 2003; 24: 497-504.