

Letter to the Editor

Comments to the predecessor of human SARS coronavirus in 2003–2004 epidemic

Keywords: SARS; Coronavirus; Civet cats

Dear Editor-in-Chief,

Wang et al. (2006) analyzed the spike gene sequences of SARS coronavirus (SARS-CoV) from the recent sporadic (December 2003–January 2004) and 2002–2003 epidemic human cases and SARS-like-CoV from civet cats. The authors claimed that SARS-CoV strain GD03T0013 from the recent sporadic cases (WHO, 2004) was genetically closer to the human SARS-CoV from the early phase of the 2002–2003 epidemic than to the wild animal SARS-like-CoV from the 2002–2003 epidemic. Wang et al. (2006) therefore concluded that the recent sporadic human SARS-CoV was closer to an unknown SARS-CoV predecessor, which is remarkably different from the conclusions of previous studies (Kan et al., 2005; Song et al., 2005; Wang et al., 2005). A drawback of the study by Wang et al. (2006) is the exclusion of a number of civet cats SARS-like-CoV sequences, leading to the inability of their analyses to fully delineate the phylogenetic origin of strain GD03T0013.

To clarify the phylogenetic origin of strain GD03T0013 we analyzed the full length spike gene nucleotide sequences ($n = 60$) of human SARS-CoV from both the recent sporadic and 2002–2003 epidemic cases, as well as SARS-like-CoV isolated from wild animals (civet cats, raccoon dogs and bats). In particular, our dataset included the SARS-like-CoV

isolated from civet cats in the 2003–2004 epidemic, which were not in the dataset of Wang et al. (2006). The sequences were aligned with ClustalX 1.83 (Thompson et al., 1994) and gap columns were removed, generating an alignment of 3672 bp. Phylogenies were reconstructed from the alignment using three methods, including neighbor-joining (NJ) implemented in PAUP* 4.0b (Swofford, 2002), maximum likelihood (ML) implemented in PhyML 2.4.4 (Guindon and Gascuel, 2003) and Bayesian Markov Chain Monte Carlo (BMCMC) implemented in MrBayes (Ronquist and Huelsenbeck, 2003). The substitution model used was the best-fit model suggested by ModelTest 3.7 (Posada and Crandall, 1998). Five thousand bootstrap replications were performed in both the NJ and ML methods, whereas two sets of four tempered MCMC chains of 550,000 generations sampled every 100th generation with initial 10% burn-ins were used in the BMCMC method.

The topologies of NJ, ML and BMCMC phylogenies are essentially similar. Therefore, only the ML phylogeny is presented here and the confidences of its topology were summarized from ML and NJ bootstrapping, as well as the sampled trees in BMCMC chains. The summarized ML phylogeny (Fig. 1) shows that strain GD03T0013 shares a monophyletic relationship with the wild animal SARS-like-CoV cluster B (isolated in 2003–2004 epidemic) and this

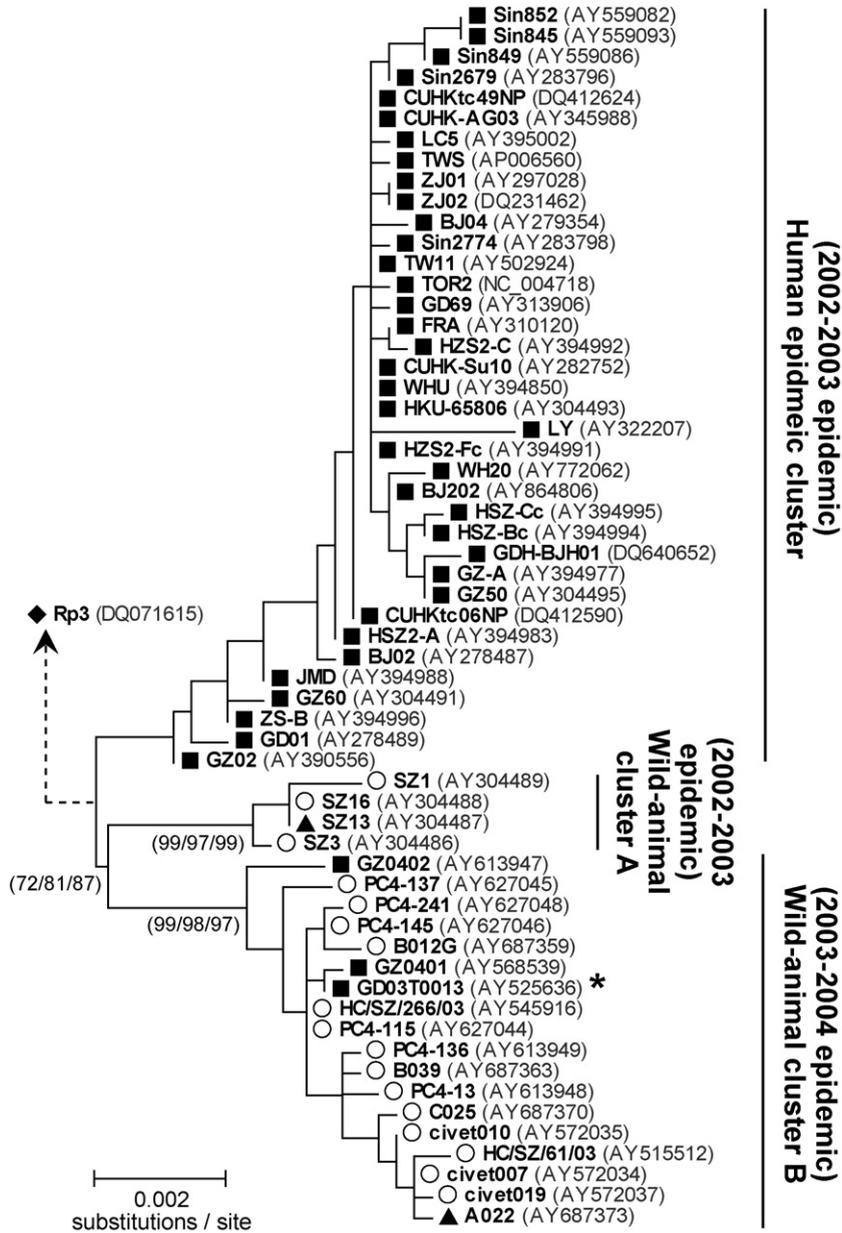


Fig. 1. Phylogenetic tree of spike gene nucleotide sequences from SARS-CoV and SARS-like-CoV isolated from humans, civet cats, raccoon dogs and bats. Sequences from humans, civet cats, raccoon dogs and bats were indicated with symbols (■), (○), (▲), and (◆), respectively. The tree was reconstructed using ML method, with confidences of topology summarized from 5000 sampled trees from ML and NJ bootstrapping and BMCMC chains. Only confidence values of major clusters were shown (ML/NJ/BMCMC, in the parenthesis). The recent human SARS-CoV isolate GD03T0013 (of the index patient) is indicated with an asterisk. GZ0402 was isolated from the second patient (a waitress) in the 2003–2004 epidemic. Accession numbers of the sequences are shown within round brackets after their strain names (in bold). The distance unit was substitutions/site. Rp3 isolated from bats was used as an out-group to root the tree, and the genetic distance of its branch is not shown.

Table 1

Comparison of pairwise nucleotide p-distances between the spike gene of strain GD03T0013 and other SARS-CoV and SARS-like-CoV isolated from human and wild animals

| Human epidemic cluster (2002–2003) | | Wild animal cluster A (2002–2003) | | Wild animal cluster B (2003–2004) | |
|------------------------------------|---------------------------------------|-----------------------------------|------------|-----------------------------------|----------|
| GZ02, HSZ-Bc | GZ60, GZ60, ZS-B, HSZ-Cc, JMD, HSZ2-A | SZ3 | SZ13, SZ16 | PC4-115, HC/SZ/266/03, GZ0401* | PC4-145 |
| 0.004630 | 0.004902 | 0.005719 | 0.005991 | 0.000272 | 0.000545 |

Isolates shared the shortest distances to GD03T0013 are shown for each cluster (human epidemic cluster, wild animal cluster A, wild animal cluster B). The p-distances were calculated based on the sequence alignment (length = 3672 bp) using MEGA 3.1 Kumar et al. (2004).

* GZ0401 (AY568539) was the complete genome sequence of the SARS-CoV isolated from the same patient (the index patient in 2003–2004 epidemic) from which the GD03T0013 (AY525636) spike gene sequence was obtained.

conclusion is supported by high confidence values. Table 1 shows the comparison of the pairwise genetic distances (nucleotide p-distances) between the spike gene of strain GD03T0013 and other strains of our dataset. The genetic distance between civet cat SARS-like-CoV strain PC4-115 (and HC/SZ/266/03) from wild animal cluster B and strain GD03T0013 is smaller than that between any strains in the human epidemic cluster and the wild animal cluster A isolated during the 2002–2003 epidemic. Remarkably, only one nucleotide difference was identified between the spike genes of strains GD03T0013 and PC4-115 (and HC/SZ/266/03). In the spike gene phylogeny (Fig. 1), PC4-115 (and HC/SZ/266/03) could be also interpreted as the direct phylogenetic predecessor of GD03T0013. Although the first patient (GD03T0013) reported no contact with civet cats and other animals in the 2 months preceding the disease onset (Wang et al., 2005), these phylogenetic and genetic evidences, as well as those presented by previous studies (Kan et al., 2005; Song et al., 2005; Wang et al., 2005) suggested that the recent human case was a sporadic infection originating from the SARS-like-CoV of a wild animal.

Our results, based on the genetic analyses of the spike gene, showed that the closest strains to the recent sporadic human SARS-CoV are the civet cat SARS-like-CoV in 2003–2004 epidemic rather than the civet cat SARS-like-CoV in 2002–2003 epidemic suggested by Zhao et al. (2004), or the human SARS-CoV from the earlier phase of 2002–2003 epidemic, or an unknown predecessor suggested by Wang et al. (2006). The major difference between Wang et al. (2006) and our analysis is the sample size and diversity of sequences used. They used fewer isolates and, more importantly, did not include the civet cat SARS-like-

CoV sequences isolated in 2003–2004 epidemic (shown in Fig. 1 under the wild animal cluster B grouping) in their analysis. Consequently, they could not fully delineate the phylogenetic origin of the SARS-CoV in the recent sporadic human cases. We want to emphasize the importance of sample size and diversity in phylogenetic analysis, especially in any search for the possible origins of a disease causing agent.

References

- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Kan, B., Wang, M., Jing, H., et al., 2005. Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. *J. Virol.* 79, 11892–11900.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* 5, 150–163.
- Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Song, H.D., Tu, C.C., Zhang, G.W., et al., 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2430–2435.
- Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Wang, M., Yan, M., Xu, H., et al., 2005. SARS-CoV infection in a restaurant from palm civet. *Emerg. Infect. Dis.* 11, 1860–1865.

- Wang, Z.G., Xu, S.P., Zhang, Y.J., et al., 2006. Genetic distance of SARS coronavirus from the recent natural case. *Vet. Microbiol.*, doi:10.1016/j.vetmic.2006.10.022.
- WHO. Review of probable and laboratory-confirmed SARS cases in southern China. http://www.who.int/csr/don/2004_01_27/en/.
- Zhao, G.P., He, J.F., Peng, G.W., et al., 2004. Molecular evolution of the coronavirus during the course of the SARS epidemic in China. *Science* 303, 1666–1669.

Tommy Tsan-Yuk Lam
Chung-Chau Hon
Pui-Yi Lam
Chi-Wai Yip
Fanya Zeng

Frederick Chi-Ching Leung*
*Department of Zoology, The University of
Hong Kong, HKSAR, China*

*Corresponding author at: 5N-01, 5/F,
North Wing, Kadoorie Biological Science Building,
The University of Hong Kong, HKSAR, China.
Tel.: +852 2299 0817; fax: +852 2857 4672
E-mail address: fcleung@hkucc.hku.hk
(F.C.-C. Leung)

15 January 2007