Journal of General Virology (2010), **91**, 1058–1062 **DOI** 10.1099/vir.0.016378-0

Short Communication

Intraspecies diversity of SARS-like coronaviruses in *Rhinolophus sinicus* and its implications for the origin of SARS coronaviruses in humans

Junfa Yuan,[1]† Chung-Chau Hon,[2]† Yan Li,[1] Dingming Wang,[3] Gelin Xu,[4] Huajun Zhang,[1] Peng Zhou,[1] Leo L. M. Poon,[5] Tommy Tsan-Yuk Lam,[2] Frederick Chi-Ching Leung[2] and Zhengli Shi[1]

Correspondence
Zhengli Shi
zlshi@wh.iov.cn

[1]State Key Laboratory of Virology, Wuhan Institute of Virology, Chinese Academy of Sciences (CAS), Wuhan, PR China

[2]School of Biological Sciences, The University of Hong Kong, Hong Kong SAR

[3]Disease Control and Prevention of Guizhou Province, Guiyang, PR China

[4]Wuhan Institute of Biological Products, Ministry of Health, Wuhan, PR China

[5]Department of Microbiology, The University of Hong Kong, Hong Kong SAR

Received 3 September 2009
Accepted 11 December 2009

The Chinese rufous horseshoe bat (*Rhinolophus sinicus*) has been suggested to carry the direct ancestor of severe acute respiratory syndrome (SARS) coronavirus (SCoV), and the diversity of SARS-like CoVs (SLCoV) within this *Rhinolophus* species is therefore worth investigating. Here, we demonstrate the remarkable diversity of SLCoVs in *R. sinicus* and identify a strain with the same pattern of phylogenetic incongruence (i.e. an indication of recombination) as reported previously in another SLCoV strain. Moreover, this strain possesses a distinctive 579 nt deletion in the *nsp3* region that was also found in a human SCoV from the late-phase epidemic. Phylogenetic analysis of the *Orf1* region suggested that the human SCoVs are phylogenetically closer to SLCoVs in *R. sinicus* than to SLCoVs in other *Rhinolophus* species. These findings reveal a closer evolutionary linkage between SCoV in humans and SLCoVs in *R. sinicus*, defining the scope of surveillance to search for the direct ancestor of human SCoVs.

Severe acute respiratory syndrome (SARS) is an infectious disease in humans caused by a coronavirus (CoV) named SARS CoV (SCoV) (Rota *et al.*, 2003). Small carnivores, including civet cats and raccoon dogs, were found to carry SCoVs and were therefore believed to be the immediate sources of the SARS epidemic in 2003 (Guan *et al.*, 2003; Kan *et al.*, 2005). Later, *Rhinolophus* species (horseshoe bats) were found to carry a diverse group of SARS CoV-like CoVs (SLCoV) and, thus, may be the natural reservoir of SLCoVs and SCoVs (Lau *et al.*, 2005; Li *et al.*, 2005). However, the currently sampled SLCoV in bats (Bt-SLCoV) may not be the direct ancestor of SCoVs in humans (Hu-SCoV) (and thus the civet and raccoon dog), because Bt-SLCoVs are phylogenetically distant from Hu-SCoVs (Ren *et al.*, 2006). Nonetheless, our previous study demonstrated the potential recombinant origin of a Bt-SLCoV strain (Rp3) (Hon *et al.*, 2008), and we speculated that an uncharacterized SLCoV lineage may exist in its host that may represent the direct ancestor of Hu-SCoV. It should be noted that we previously misidentified the host of Rp3 as *Rhinolophus pearsonii* (Li *et al.*, 2005); this is now corrected to *Rhinolophus sinicus* based on phylogenetic analysis of its *cytB* sequence in this study (Fig. 1a). Therefore, we anticipate that an investigation of the genetic diversity of SLCoVs in *R. sinicus* could provide insights into the possible direct ancestor of Hu-SCoVs.

Faecal samples of 24 *R. sinicus* individuals were collected from distinct locations within China during September 2006. RNA was extracted and used for first-strand cDNA synthesis as described previously (Li *et al.*, 2005). Viral detection using PCR amplification of a 440 nt conserved region of *nsp12* (RNA-dependent RNA polymerase) was performed as described previously (Lau *et al.*, 2005). For host species identification of positive samples, the mitochondrial cytochrome B gene (*cytB*) was sequenced as described previously (Cui *et al.*, 2007). The *cytB* genes of the hosts of the previously reported Bt-SLCoVs were also

The GenBank/EMBL/DDBJ accession numbers for the sequences obtained in this study are FJ588686 (complete genome of Rs672), FJ588687–FJ588692 (*nsp3*, *nsp13*, *S*, *Orf3a* and *N* sequences of Rs806) and FJ588681–FJ588685 and FJ561748 (*cytB* sequences of the hosts of Rp3, Rm1, Rf1, Rs672, Rs806 and HKU3).

†These authors contributed equally to this paper.

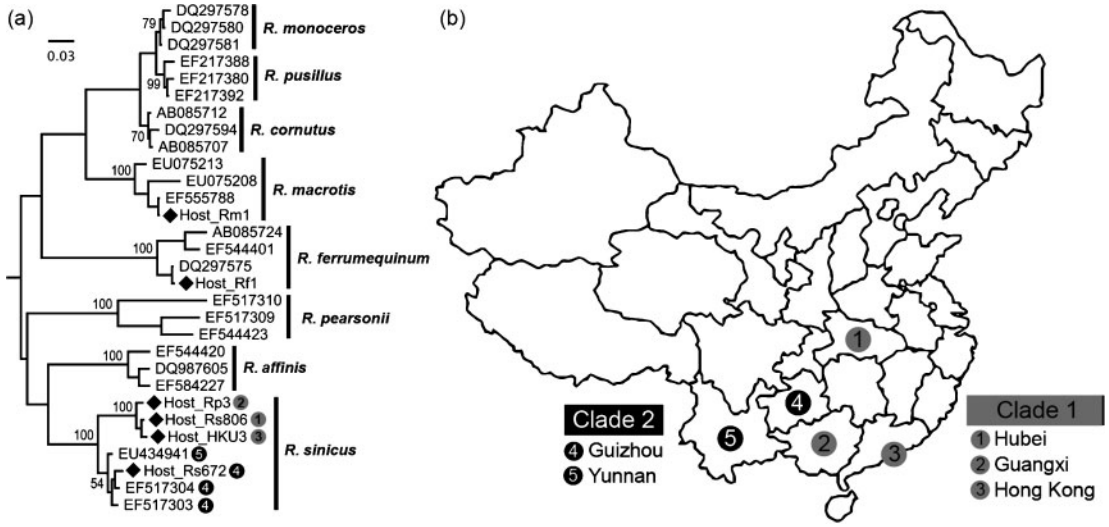

Fig. 1. Phylogeny of the hosts of Bt-SLCoVs and other *Rhinolophus* species based on their *cytB* sequences. (a) Bayesian phylogeny constructed using MRBAYES (Ronquist & Huelsenbeck, 2003) under the best-fit substitution model determined by MODELTEST 3.7 (Posada & Crandall, 1998). The taxa are represented by GenBank accession numbers; the hosts of Bt-SLCoVs are highlighted with diamonds. BtCoV279 and BtCoV273 (GenBank accession numbers DQ648857 and DQ648856) were excluded, since they were isolated from the same field samples as Rm1 and Rf1, respectively (S. Zhang, personal communication). Bayesian posterior probability support (BPS) of selected nodes is indicated as percentages. Species names are indicated to the right. Bar, 0.03 nucleotide substitutions per site. The geographical origins of *R. sinicus* taxa are indicated by circled numbers, referring to the legend in (b). (b) Geographical origins (at the level of province) of the *R. sinicus* taxa. Black lines on the map represent provincial borders.

sequenced. Two positive samples, designated Rs672 and Rs806, were collected from two relatively distant locations, in Guizhou and Hubei provinces, respectively (Fig. 1b). We obtained their genome sequences by using PCR amplification and sequencing strategies described previously (primer sequences available upon request) (Li *et al.*, 2005). For Rs672, the sequences for all putative coding regions were obtained. For Rs806, we failed to recover its full-length genome sequence, and only the regions covering *nsp3*, *nsp13*, *S*, *Orf3a* and *N* were sequenced (Fig. 2b). The *cytB* sequences of the hosts of Rs672, Rs806, Rp3, Rm1, Rf1 (Li *et al.*, 2005; Ren *et al.*, 2006) and HKU3 (Lau *et al.*, 2005) were analysed with those of other available *Rhinolophus* species. The phylogeny suggests that the hosts were correctly identified in the original publications with the exception of the host of Rp3 (Li *et al.*, 2005; Ren *et al.*, 2006), as mentioned above. Notably, the *R. sinicus* sequences appear as two topologically distinct clusters in the phylogeny, i.e. the hosts of Rp3, HKU3 and Rs806 in clade 1 and the host of Rs672, with several other taxa, in clade 2 (Fig. 1a). The geographical origins of clades 1 and 2 cover five provinces in southern China (Fig. 1b).

A plot of similarity (Lole *et al.*, 1999) across the genome of Rs672 suggests that its 5′ region (i.e. *Orf1a* and *Orf1b*) is more similar to that of Hu-SCoV than to that of Bt-SLCoV. In contrast, the 3′ region of the genome of Rs672 is generally more similar to Bt-SLCoV than to Hu-SCoV (Fig. 2a). This pattern of incongruence is similar to that observed in the previously reported potential recombinant Rp3 (Hon *et al.*, 2008). We therefore applied the recombination detection methods implemented in GARD (Kosakovsky Pond *et al.*, 2006) and LARD (Holmes *et al.*, 1999) to locate the potential recombination breakpoint in Rs672. The parameters and strategies employed are the same as described previously (Hon *et al.*, 2008), except that Rs672 replaced Rp3 as the query. Both analyses suggested that the phylogenetic incongruence in the genome of Rs672 is statistically significant, and the potential breakpoint was estimated at the nucleotide immediately after the start codon of the spike gene (model average support in GARD >0.8; P value in likelihood ratio test in LARD <0.0001). This potential recombination breakpoint in the genome of Rs672 is at the same position as the one detected in the genome of Rp3 (Hon *et al.*, 2008). The genome regions upstream and downstream of this potential breakpoint were respectively designated the major and minor parental regions.

Two phylogenies were constructed based on the coding sequences of the parental regions of selected Hu-SCoV ($n=9$) and all available Bt-SLCoV ($n=6$) (Fig. 2c). To ensure the quality of the alignment of divergent regions, we used concatenated codon alignment instead of full genome nucleotide alignment (Fig. 2b). For Rs806, both phylogenies suggest its close phylogenetic relationship with another Bt-SLCoV in *R. sinicus*, HKU3. For Rs672, its phylogenetic positions in the two phylogenies are incongruent. Based on

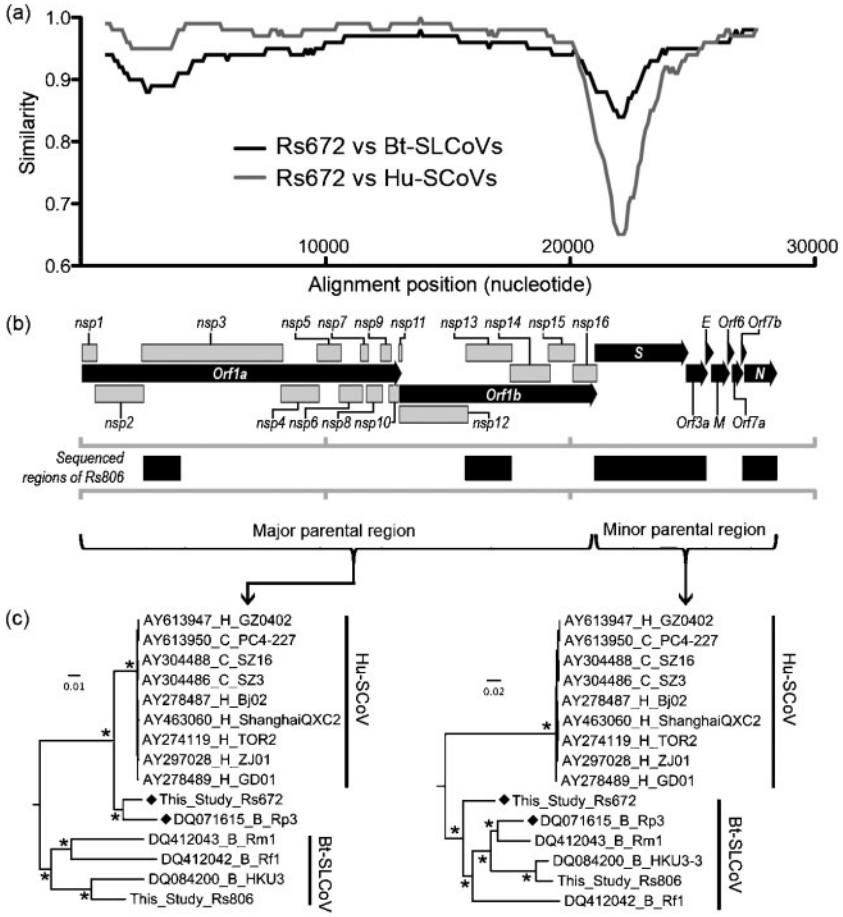

Fig. 2. Phylogenetic incongruence within the genome of Rs672. (a) Similarity plot based on concatenated codon alignment as indicated in (b). 'Hu-SCoVs' refers to all Hu-SCoVs in (c); 'Bt-SLCoVs' refers to HKU3, Rm1 and Rf1. The plot was generated using Simplot (Lole *et al.*, 1999), with window size of 2000 nt and a step size of 200 nt, under the F84 distance model. (b) Schematic diagram of the concatenated coding sequences used in (a) and (c). ORFs are displayed as filled arrows and the non-structural protein products of *Orf1a* and *Orf1b* are displayed as shaded boxes. For annotation of ORFs and non-structural proteins, refer to the sequence of TOR2 (GenBank accession no. AY274119). Sequenced regions of Rs806 are also indicated. (c) Bayesian phylogenies constructed using concatenated codon alignments of the indicated regions. Unavailable regions of the sequence of Rs806 were treated as missing data in the Bayesian estimations. The phylogenies were constructed as described in the legend of Fig. 1. Hu-SCoV and Bt-SLCoV lineages are indicated. Accession numbers, host species [civets (C), humans (H) or bats (B)] and strain names are given. Bar, 0.01 (left) and 0.02 (right) nucleotide substitutions per site. Asterisks represent nodes with BPS >90 %. Phylogenetically incongruent taxa in both phylogenies are highlighted by diamonds.

Fig. 2(c) (left panel), the major parental regions of Rs672 and Rp3 (designated the HB-SLCoV lineage) shared a monophyly that is phylogenetically closer to Hu-SCoVs than to Bt-SLCoVs. In contrast, the minor parental region of Rs672 is clustered with the Bt-SLCoVs lineage (Fig. 2c, right panel). The current data suggest that Rs672 and Rp3 are likely to have evolved from a common ancestor with the same pattern of phylogenetic incongruence. Interestingly, the host of Rs672 is phylogenetically distinct from those of the other three Bt-SLCoVs in *R. sinicus* (Fig. 1), implying that Rs672 and Rp3 (as well as their hosts) may have diverged a relatively long time ago, which may also explain the observation that the minor parental region of Rs672 did not share a monophyly with that of Rp3.

Despite the statistical significance, such incongruence could be explained as laboratory artefacts or asymmetrical genome evolution and not necessarily as homologous recombination. Nonetheless, the identical pattern of phylogenetic incongruence identified in Rp3 and Rs672 may exclude the possibility that it represents a laboratory artefact. Alternatively, if the mode of evolution (e.g. substitution rate) of *Orf1* in Rp3/Rs672 is significantly different from that of other Bt-SLCoVs, it could also lead to an apparently closer phylogenetic relationship between the *Orf1* of Rp3/Rs672 and HuSCoVs. However, from a biological perspective, it may not be sensible to assume that the same ORFs in viruses of the same host species would evolve in a significantly different manner, as Rs806, HKU3, Rp3 and Rs672 all reside in the same host species, *R. sinicus.* In other words, unless differential selection pressure was operating on the viruses residing in specific individuals of *R. sinicus*, the homologous genome regions of Rs806, HKU3, Rp3 and Rs672 would be expected to evolve in a similar manner. Therefore, based on the current data, homologous recombination seems to be the most plausible explanation for the observed phylogenetic incongruence. If such recombination did happen, it is unlikely to have occurred recently, since Rp3 and Rs672 showed considerable divergence, and their major parent is expected to be a Bt-SLCoV that shares a close phylogenetic relationship with Hu-SCoVs.

We previously used a Bayesian relaxed molecular clock to estimate the time of divergence between Rp3 and Hu-SCoVs (tDiv-Hu/HB), and the results suggested that the interspecies transfer of SLCoVs to the amplifying host (e.g. civets) might have happened a median of 4.08 years before the SARS outbreak (Hon *et al.*, 2008). Here, we have reanalysed the dataset, using the same method and model parameters as described previously (Hon *et al.*, 2008), with the addition of two sequences derived from this study

(Rs672 and Rs807), to investigate whether the addition of these highly relevant sequences would alter our previous estimate significantly. Fig. 3 shows a time-scaled phylogeny constructed using BEAST version 1.4.8 (Drummond & Rambaut, 2007). The time of divergence between Rs672/Rp3 and Hu-SCoVs (i.e. tDiv-Hu/HB) was estimated to be at a median of 1999.04, with 95 % highest posterior density (HPD) from 2002.06 to 1992.87, which is consistent with out previous estimate (median 1998.51, 95 % HPD 1993.55–2001.32) (Hon *et al.*, 2008). The length of branch A in Fig. 3, representing the window period between the hypothetical interspecies transmission event and the onset of the SARS epidemic, was estimated at a median of 4.29 years (95 % HPD 0.70–9.57 years). The estimate in this study is fairly consistent with our previous estimate (median 4.08 years, 95 % 1.45–8.84 years) (Hon *et al.*, 2008), although its HPD range is slightly wider. This analysis was repeated with exponential and logistic coalescent tree prior assumptions and the results were generally consistent (not shown), suggesting that our estimation is robust under different coalescent tree prior assumptions. The robustness of the estimation of this relatively short window period further supports our previous speculation that this uncharacterized HB-SLCoV lineage may contain the direct ancestor of SCoV.

The *nsp3* region of *Orf1a* is relatively variable in the genome of Bt-SLCoVs, as a number of large deletions were observed within this region (Ren *et al.*, 2006). Interestingly, in PCR amplifications of the *nsp3* region of Rs672, we identified a size variant in one of the amplicons, which was later confirmed (using sequencing) to contain a deletion of 579 nt (Del-579nt). In an attempt to characterize the potential viral subpopulations in sample Rs672, we tried to culture and isolate viruses using methods described previously (Li *et al.*, 2005), but were unsuccessful. Nonetheless, our sequencing results suggest that the flanking sequences of Del-579nt in both size variants of the amplicon are identical (GenBank accession no. GQ870287), and these results were confirmed using another set of primers for PCR amplification and sequencing (not shown). Moreover, the sequences of cloned amplicons (n=10) in the *S* region of Rs672 were almost identical, except for some non-recurrent substitutions which are probably due to PCR artefacts or the presence of a homogeneous viral quasispecies population. Therefore, there is no evidence to support the presence of multiple distinct viral subpopulations in sample Rs672, except for Del-579nt in the *nsp3* region. We therefore believe that the consensus genome of Rs672 is a robust representation of the homogeneous viral population in the sample. Surprisingly, Del-579nt was also found in an Hu-SCoV strain in the late phase of the 2003 epidemic, ShanghaiQXC (GenBank accession no. AY463060), but not in other Hu-SCoVs or Bt-SLCoVs available in GenBank. Considering the observation that the viral genomes in Rs672 are homogeneous except for Del-579nt in the *nsp3* region, this deletion was likely to have occurred recently

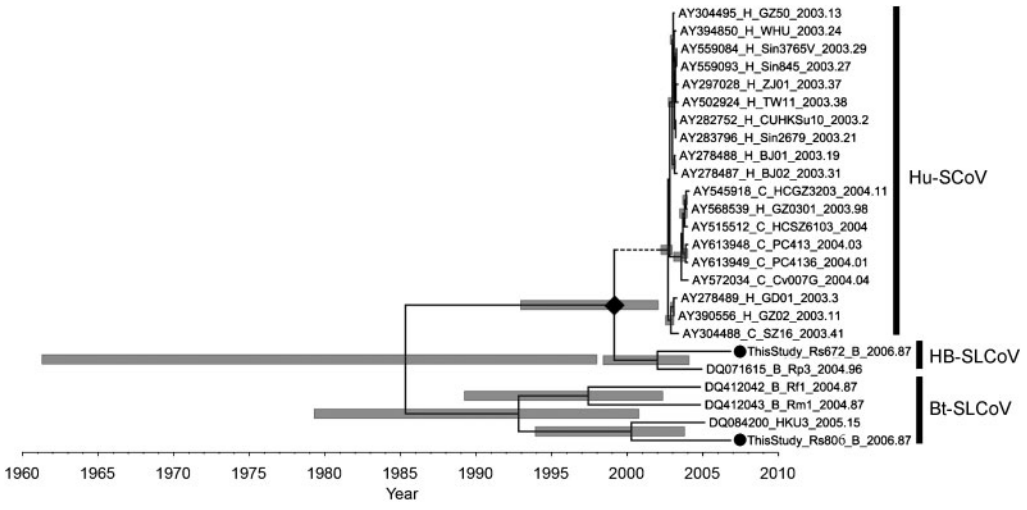

Fig. 3. Estimation of the time of divergence between the major parental region of Rs672/Rp3 and Hu-SCoVs. The time-scaled phylogeny was constructed as described previously (Hon *et al.*, 2008), using BEAST version 1.4.8 (Drummond & Rambaut, 2007) under the assumption of a constant coalescent tree prior and an exponentially relaxed molecular clock (UCED) (Drummond *et al.*, 2002, 2006). The dataset was the same as described previously (Hon *et al.*, 2008) except for the addition of the full sequence of Rs672 and the partial sequence of Rs806 (indicated by filled circles); the unsequenced region of Rs806 was treated as missing data in the Bayesian analysis. tDiv-Hu/HB is indicated by a filled diamond. The dashed line indicates branch A, the window between the hypothetical interspecies transmission event and the onset of the SARS epidemic. Shaded bars represent 95 % HPD of the estimated divergence time of the corresponding nodes (with Bayesian posterior probability >70 %). The taxa are annotated as in Fig. 2, except that the sampling time is given at the end.

within the host of Rs672; in other words, Del-579nt in Rs672 and ShanghaiQXC might have arisen independently and was probably not acquired through homologous recombination. We cannot establish a plausible hypothesis for the origin of Del-579nt; a wider survey of Del-579nt in Bt-SLCoVs may shed light on its origin.

In conclusion, this is the first study to report the intraspecies diversity of Bt-SLCoVs. Our data suggest the presence of at least two distinct genotypes of Bt-SLCoV in *R. sinicus* (i.e. Rp3/Rs672 and HKU3/Rs806); the diversity of Bt-SLCoV in *Rhinolophus macrotis* and *Rhinolophus ferrumequinum* is currently unknown. Our findings also provide evidence for the potential recombinant origin of Rp3 and Rs672. Their major parent has a relatively closer phylogenetic relationship with Hu-SCoVs, suggesting the possible presence of a Bt-SCoV lineage in *R. sinicus* that may contain the direct ancestor of Hu-SCoVs, as we proposed previously (Hon *et al.*, 2008). However, we note that these speculations are based on a few strains only, and extensive surveys on the prevalence of this genotype would indicate their credibility. *R. sinicus* is an extremely common *Rhinolophus* species in China (Lau *et al.*, 2005), and our data imply that Hu-SCoVs are phylogenetically closer to some Bt-SLCoVs in *R. sinicus* than to those in *R. macrotis* and *R. ferrumequinum*. Therefore, we suggest more-focused surveillance of Bt-SLCoVs in *R. sinicus*, which may provide insights into the possible direct ancestor of Hu-SCoVs.

Acknowledgements

This work was jointly funded by a State Key Program for Basic Research grant (2005CB523004) and the International Cooperation Centre Project (2008GR1409) from the Chinese Ministry of Science, the National Natural Science Foundation (30970137) and the Technology and Knowledge Innovation Program Key Project administered by the Chinese Academy of Sciences (KSCX1-YW-R-07).

References

Cui, J., Han, N., Streicker, D., Li, G., Tang, X., Shi, Z., Hu, Z., Zhao, G., Fontanet, A. & other authors (2007). Evolutionary relationships between bat coronaviruses and their hosts. *Emerg Infect Dis* **13**, 1526–1532.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. & Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307–1320.

Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., Luo, S. W., Li, P. H., Zhang, L. J. & other authors (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276–278.

Holmes, E. C., Worobey, M. & Rambaut, A. (1999). Phylogenetic evidence for recombination in dengue virus. *Mol Biol Evol* **16**, 405–409.

Hon, C. C., Lam, T. Y., Shi, Z. L., Drummond, A. J., Yip, C. W., Zeng, F., Lam, P. Y. & Leung, F. C. (2008). Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J Virol* **82**, 1819–1826.

Kan, B., Wang, M., Jing, H., Xu, H., Jiang, X., Yan, M., Liang, W., Zheng, H., Wan, K. & other authors (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.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096–3098.

Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Tsoi, H. W., Wong, B. H., Wong, S. S., Leung, S. Y., Chan, K. H. & Yuen, K. Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* **102**, 14040–14045.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Crameri, G., Hu, Z. & other authors (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152–160.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Ren, W., Li, W., Yu, M., Hao, P., Zhang, Y., Zhou, P., Zhang, S., Zhao, G., Zhong, Y. & other authors (2006). Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. *J Gen Virol* **87**, 3355–3359.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K. & other authors (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394–1399.