

ORF8-Related Genetic Evidence for Chinese Horseshoe Bats as the Source of Human Severe Acute Respiratory Syndrome Coronavirus

Zhiqiang Wu,^{1,a} Li Yang,^{1,a} Xianwen Ren,^{1,a} Junpeng Zhang,² Fan Yang,¹ Shuyi Zhang,^{3,b} and Qi Jin^{1,4,b}

¹MOH Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, ²Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou, ³State Key Laboratory of Estuarine and Coastal Research, Institute of Estuarine and Coastal Research, East China Normal University, Shanghai, and ⁴College of Animal Science and Veterinary Medicine, Shenyang Agricultural University, People's Republic of China

Several lineage B betacoronaviruses termed severe acute respiratory syndrome (SARS)-like CoVs (SL-CoVs) were identified from *Rhinolophus* bats in China. These viruses are characterized by a set of unique accessory open reading frames (ORFs) that are located between the M and N genes. Among unique accessory ORFs, ORF8 is most hypervariable. In this study, the ORF8s of all SL-CoVs were classified into 3 types, and, for the first time, it was found that very few SL-CoVs from *Rhinolophus sinicus* have ORF8s that are identical to that of human SARS-CoV. This finding provides new genetic evidence for Chinese horseshoe bats as the source of human SARS-CoV.

Keywords. severe acute respiratory syndrome coronavirus; origin; bats; ORF8.

The severe acute respiratory syndrome (SARS) pandemic in 2002–2003 spread to 29 countries, caused 8098 cases, and led to 774 deaths. A novel coronavirus (CoV), termed SARS-CoV, was identified as the etiological agent. SARS-CoV belongs to lineage B in the genus *Betacoronavirus* (beta-CoV) of the family *Coronaviridae* [1].

Although 12 years have passed without a recurrent SARS outbreak, the search for the original animal reservoir for human SARS-CoVs is ongoing. Researchers have discovered lineage B beta-CoVs related to SARS-CoVs in insectivorous *Rhinolophus* and *Chaerephon* bats in China. The nucleotide sequences in the ORF1ab, E, M, and N genes in these bat-borne lineage B beta-CoVs are 89%–93% similar to those in the SARS-CoVs from humans. The CoVs were thus named “SARS-like CoVs” (SL-CoVs). The finding of diverse SL-CoVs in bats led to the hypothesis that *Rhinolophus* bats were the natural reservoirs of SARS-CoVs [2–4]. However, this hypothesis is challenged by the significant differences in nucleotide and amino acid sequences in certain hypervariable regions. These regions include the S1 domain of the viral spike glycoprotein (S) and the open reading frames (ORFs) that encode a set of accessory proteins, particularly ORF8. Although a functionally similar bat-origin S gene was recently identified in the SL-CoVs of *Rhinolophus sinicus* (Chinese horseshoe bats) and *Rhinolophus affinis* (WIV1, Rs3367, and LYRa11), it had less sequence similarity to the human-origin S gene [2, 5]. Furthermore, genetic evidence for the identical ORF8 is needed to trace the origin of SARS-CoVs to bat SL-CoVs.

METHODS

Nucleotide Sequence Accession Numbers

All genome sequences were submitted to GenBank. The accession numbers for all viruses are KJ473811–KJ473822, JX993987, JX993988, and KF636752. The GA II sequence data were deposited into the National Center for Biotechnology Information Sequence Read Archive under accession number SRA051252.

Genomic and Phylogenetic Analysis

The nucleotide sequences of the genomes and the amino acid sequences of the ORFs were deduced by comparing them with the sequences in other CoVs. The conserved protein families and domains were predicted using Pfam and InterProScan 5 (available at: <http://www.ebi.ac.uk/services/proteins>). Routine sequence alignments were performed using Clustal Omega, Needle (available at: <http://www.ebi.ac.uk/Tools/>), MegAlign (Lasergene, DNASTAR, Madison, Wisconsin), and T-coffee with manual curation. MEGA5.0 (Phoenix, Arizona) was used to align the nucleotide sequences and the deduced amino acid sequences, using the MUSCLE package and default parameters. The best substitution model was then evaluated using the Model Selection package. Finally, we constructed a maximum-likelihood

Received 4 August 2015; accepted 24 September 2015.

^aZ. W., L. Y., and X. R. contributed equally to this work.

^bS. Z. and Q. J. contributed equally to this article and are co-senior authors.

Correspondence: Qi Jin, MD, PhD, MOH Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, People's Republic of China (zdsys@vip.sina.com).

The Journal of Infectious Diseases®

© The Author 2015. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com.

DOI: 10.1093/infdis/jiv476

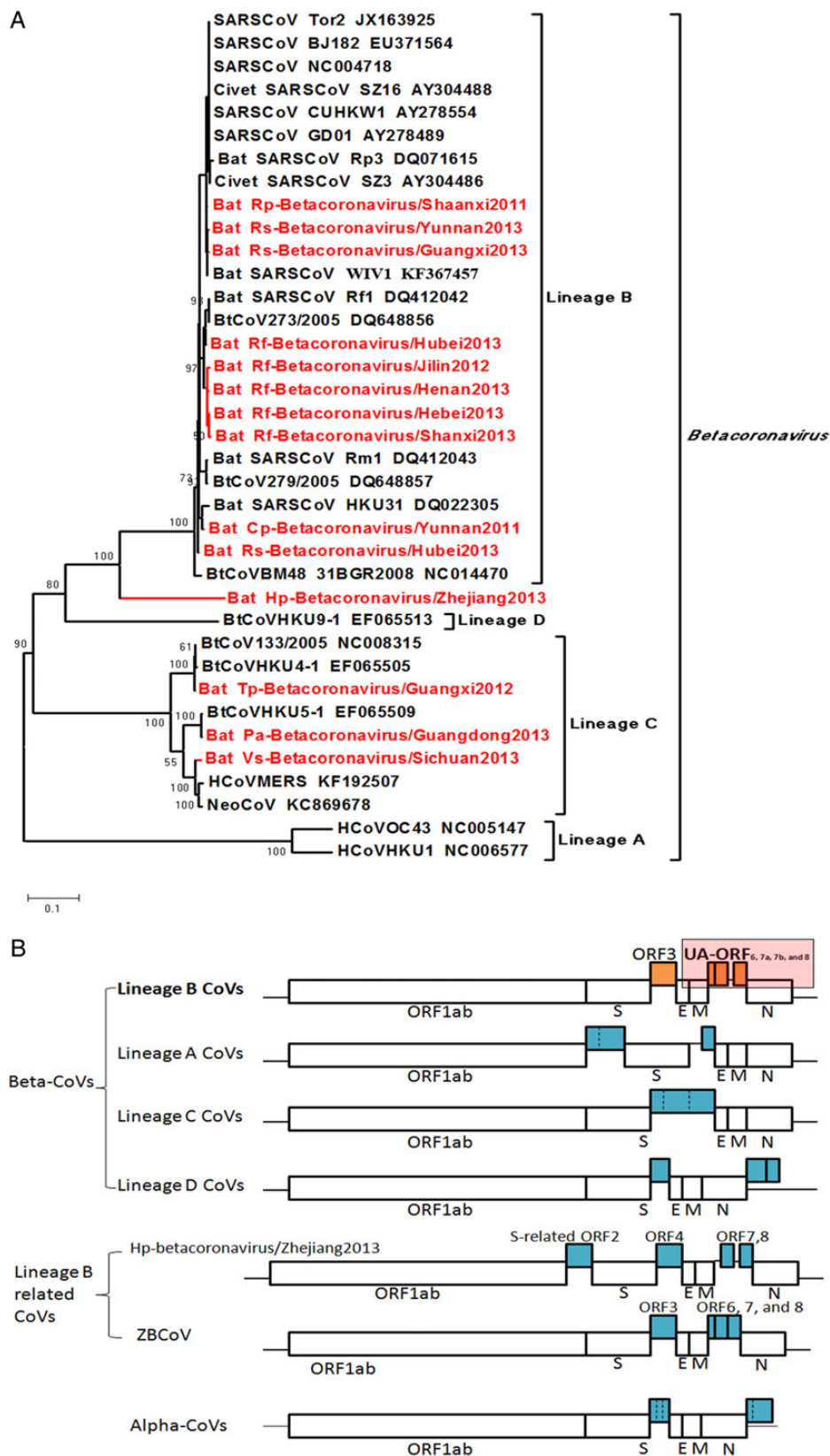


Figure 1. A, Phylogenetic tree based on the complete RNA-dependent RNA polymerase (RdRp; NSP12) proteins of betacoronaviruses. Scale bar represents 0.1 substitutions per nucleotide site. B, Differences in genome organization between coronaviruses of each lineage of betacoronaviruses and alphacoronaviruses. The unique accessory open reading frames (ORFs) of lineage B betacoronaviruses are labeled in orange. This figure is available in black and white in print and in color online.

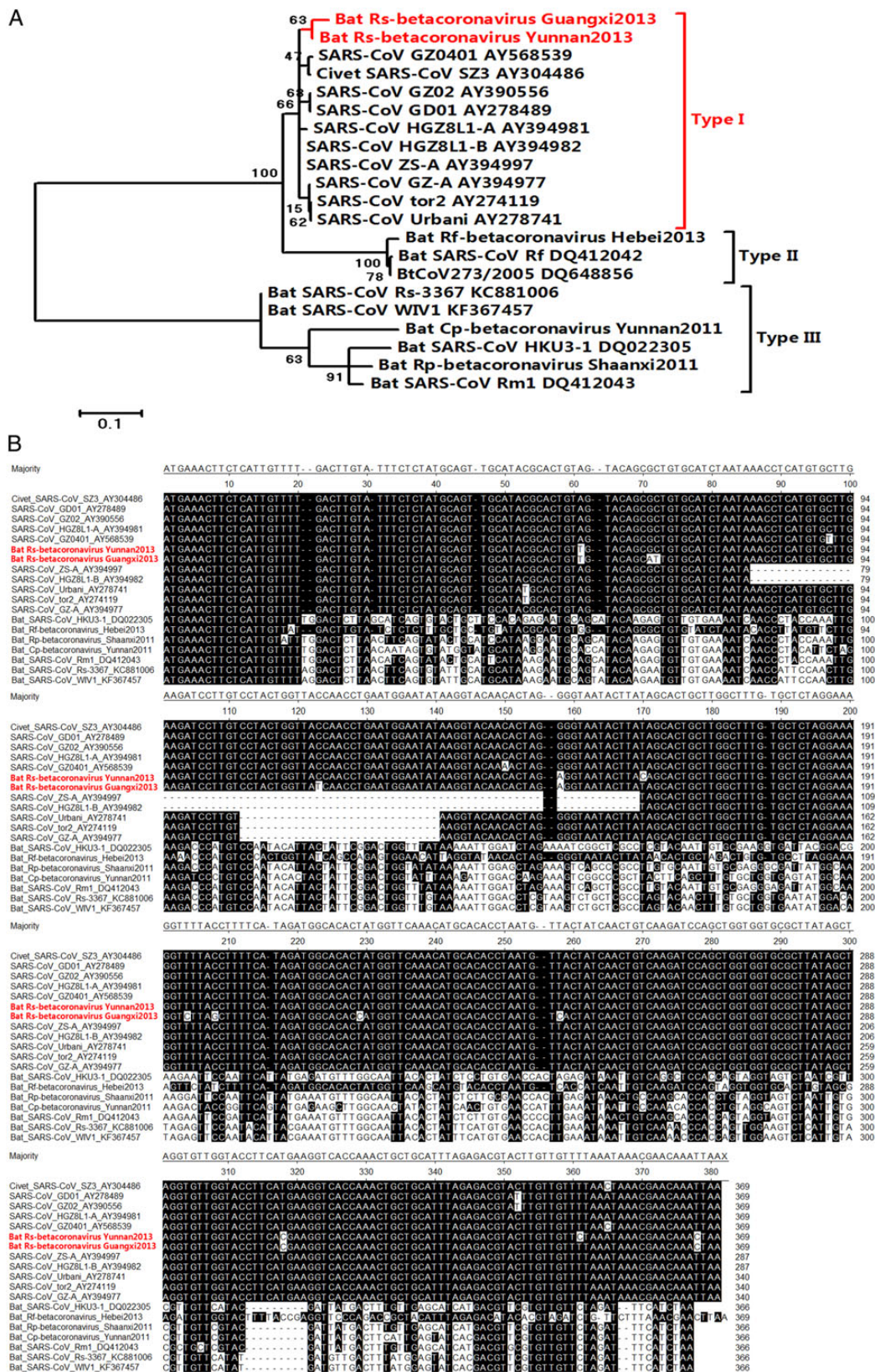


Figure 2. A, Phylogenetic analysis of lineage B betacoronaviruses, based on the complete nucleotide sequences of open reading frame (ORF8). Scale bar represents 0.1 substitutions per nucleotide site. B, Multiple alignments were conducted on the basis of the complete nucleotide sequences of ORF8.

method, using an appropriate model to process the phylogenetic analyses with 1000 bootstrap replicates.

RESULTS

In this study, a systematic survey of bat-borne CoVs was performed using bat virome data from throughout China, described in our previous report [6], to obtain genetic evidence indicating the source of SARS-CoVs. Fifteen SL-CoVs were identified from 9 bat species in 11 provinces (Figure 1A) [6]. Throat and anal swab specimens from 22 wild and 124 farmed palm civets from Guangxi, Hunan, and Fujian provinces were also used for virome analysis. However, we did not detect any CoV-related sequences in the civet samples.

Compared with other bat-borne alpha-CoVs and beta-CoVs, the bat lineage B beta-CoVs (or SL-CoVs) are characterized by a set of unique accessory ORFs (UA-ORFs; Figure 1B). These UA-ORFs encompassed approximately 1085–1095 base pairs and were located between the M and N genes. The UA-ORFs encoded putative ORF6, ORF7a, ORF7b, and ORF8 proteins from the 5' terminus to the 3' terminus. These UA-ORFs were absent or significantly different from those of all other alpha-CoVs, beta-CoVs, gamma-CoVs, and delta-CoVs [7–8].

The sequence analysis of the identified lineage B beta-CoVs revealed that, in UA-ORFs from all human SARS-CoVs and SL-CoVs, ORF6 and ORF7s are highly conserved (90.8%–96.9% nucleotide sequence identities for ORF6 and ORF7a, and 85.2%–99.3% nucleotide sequence identity for ORF7b). However, ORF8 is hypervariable (approximately 48.6% nucleotide sequence identity; Supplementary Tables) [6]. The phylogenetic analysis of the sequences of ORF8 from all available bat-borne lineage B beta-CoVs suggested that ORF8s can be divided into 3 types (Figure 2A). Type I ORF8s shared high intra-type sequence similarity (97.6% nucleotide sequence identity) and low inter-type sequence similarities to the type II ORF8s (approximately 80.3% nucleotide sequence identity) and type III ORF8s (<52% nucleotide sequence identity). The type II ORF8s showed >97% intra-type identity, whereas the type III ORF8s showed 80%–95% intra-type identity and <50% nucleotide sequence identity with the type II ORF8s.

Type II and type III ORF8s but not type I ORF8s were previously detected in bat lineage B beta-CoVs. In this study, we observed that most of the ORF8s in bat lineage B beta-CoVs are either type II or type III. All type II ORF8s were found in *Rhinolophus ferrumequinum*. However, type III ORF8s were detected in multiple bat species, including *R. sinicus*. We are the first to have found a type I ORF8 in bat lineage B beta-CoVs. Type I ORF8 is rare. In the country-wide screening, we found that only 2 CoVs collected from *R. sinicus* in Kunming City in Yunnan Province (Rs-betacoronavirus/Yunnan2013) and Hezhou City in Guangxi Province (Rs-betacoronavirus/Guangxi2013) contained type I ORF8s. Thus, according to the currently available

data, *R. sinicus* is the only bat species that harbors SL-CoVs with type I ORF8. Additionally, *R. sinicus* is also the only bat species with SL-CoVs containing 2 different types of ORF8s. The type III ORF8 is the dominant type; type I is the minor type.

The ORF8s of SARS-CoVs are also type I. The ORF8s found in SARS-CoVs from human patients in the early phase of the first epidemic of SARS in 2003 (represented by the GZ02 and GD01 isolates) and the 4 patients during the 2003–2004 outbreak (represented by the GZ0401 isolate) are nearly identical to those of the 2 newly identified CoVs, Rs-betacoronavirus/Yunnan2013 and Rs-betacoronavirus/Guangxi2013, with a few single-nucleotide mutations (98% and 99% nucleotide sequence identities, respectively; Figure 2B). This region of SARS-CoV experiences ongoing adaptive evolution in humans with gradual deletions (29-nucleotide, 82-nucleotide, or 415-nucleotide deletions) after transmission to humans [9–10]. The undeleted region of viruses with the 29-nucleotide deletion (the 29-nucleotide deletion splits ORF8 into ORF8a and ORF8b, represented by the GZ-A, Urbani, and TOR2 isolates) and viruses with the 82-nucleotide deletion (represented by the ZS-A and HGZ8L1-B isolates) also showed approximately 98% nucleotide sequence identities with the type I ORF8, with a few single-nucleotide mutations. The nearly identical ORF8s between SARS-CoVs and SL-CoVs from Chinese horseshoe bats identified in this study suggests a critical role for Chinese horseshoe bats in the maintenance of SARS-CoVs.

DISCUSSION

The discovery of SL-CoVs in several bat species (including *R. ferrumequinum*, *R. sinicus*, *Rhinolophus pusillus*, *Rhinolophus macrotis*, *R. affinis*, and *Chaerephon plicata*) and the characterization of UA-ORFs shared only by SARS-CoVs and SL-CoVs in the family *Coronaviridae* established a genetic relationship between bats and human SARS-CoVs. The ORF8 nearly identical to that in SARS-CoV was found only in SL-CoVs from *R. sinicus* and traces the source of SARS-CoVs to Chinese horseshoe bats. Functional studies for the proteins of ORF8, ORF8a, and ORF8b have been reported. The 8a protein enhances SARS-CoV replication and induces caspase-dependent apoptosis [11]. The expression of the 8b protein is related to DNA synthesis and the degradation of E protein [12–13]. The ORF8 protein may be functional in SL-CoVs from *R. ferrumequinum* [14]. However, the ORF8 region may code for a functionally unimportant protein for human SARS-CoVs, because gradual deletions in this region found in the early phase, the middle phase, and the late phase of the epidemic of SARS did not apparently affect the survival of the virus [9]. Thus, changes in this region can act as fingerprints to trace the genesis of SARS-CoVs. The identical ORF8s found in SL-CoVs from *R. sinicus* and SARS-CoVs from patients in the early phase of the SARS

epidemic provides a link indicating that the first human infection with SARS-CoV may have originated from bats.

Although an S gene identical to that of SARS-CoV has not been found in any bat species, the distinctive diversity of the S region in SL-CoVs from *R. sinicus* may imply that SL-CoVs in *R. sinicus* are prone to recombine within *R. sinicus* or with CoVs from other hosts [6]. The identification of the S protein from SL-WIV1 from *R. sinicus* greatly increases the possibility of recombination of different SL-CoVs to generate SARS-CoV in *R. sinicus*.

As the original site of the SARS pandemic, Guangdong Province is the primary region in China in which wildlife (including bats) is consumed. The supply of wildlife in Guangdong comes from surrounding provinces, such as Guangxi, Yunnan, Hunan, and Fujian. The SL-CoVs from *R. sinicus* have nearly identical ORF8s and similar backbone genes to those in SARS-CoVs in Yunnan and Guangxi provinces. Furthermore, the observation that SL-CoVs from *R. sinicus* are prone to recombine with CoVs from other hosts may suggest that the wildlife markets in Guangdong may provide an ideal incubator for the genesis of SARS-CoVs. Moreover, human consumption of wildlife increases the possibility of human exposure to viruses carried by wildlife. SL-CoVs closely related to human SARS-CoVs are still present in nature, and the custom of wildlife consumption is ongoing. Thus, there is an ongoing risk of SARS reemergence or the emergence of a similar zoonotic infectious disease in humans.

Although palm civets were once suspected to be the natural reservoirs of human SARS-CoV, the isolation and genome sequencing of SARS-CoVs in civets was limited to those present in the marketplace of the epidemic area of the SARS outbreak and only during the outbreak period [10, 15]. In our study in 2012, we used the same metagenomic methods to detect SL-CoV genomes in civets from Guangxi, Hunan, and Fujian provinces that supply palm civets to the markets in Guangdong. However, we did not detect any SARS-CoV or SL-CoV sequences in any samples. This finding is consistent with a non-civet-origin CoV reported prior to SARS and after the pandemic.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (<http://jid.oxfordjournals.org>). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Financial support. This work was supported by the Program for Changjiang Scholars and Innovative Research Team in University of China (IRT13007); the National S&T Major Project “China Mega-Project for Infectious Disease,” People’s Republic of China (grants 2011ZX10004-001 and 2014ZX10004001); and the National Natural Science Foundation of China (grants 81501773 and 31570382).

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Donnelly CA, Fisher MC, Fraser C, et al. Epidemiological and genetic analysis of severe acute respiratory syndrome. *Lancet Infect Dis* **2004**; 4:672–83.
2. Ge XY, Li JL, Yang XL, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **2013**; 503:535–8.
3. Li W, Shi Z, Yu M, et al. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **2005**; 310:676–9.
4. Rota PA, Oberste MS, Monroe SS, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **2003**; 300:1394–9.
5. He B, Zhang Y, Xu L, et al. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. *J Virol* **2014**; 88:7070–82.
6. Wu Z, Yang L, Ren X, et al. Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *ISME J* **2015**; doi:10.1038/ismej.2015.138.
7. Holmes KV, Enjuanes L. Virology. The SARS coronavirus: a postgenomic era. *Science* **2003**; 300:1377–8.
8. Woo PC, Lau SK, Huang Y, Yuen KY. Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med (Maywood)* **2009**; 234:1117–27.
9. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **2004**; 303:1666–9.
10. Song HD, Tu CC, Zhang GW, et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci U S A* **2005**; 102:2430–5.
11. Chen CY, Ping YH, Lee HC, et al. Open reading frame 8a of the human severe acute respiratory syndrome coronavirus not only promotes viral replication but also induces apoptosis. *J Infect Dis* **2007**; 196:405–15.
12. Law PY, Liu YM, Geng H, Kwan KH, Waye MM, Ho YY. Expression and functional characterization of the putative protein 8b of the severe acute respiratory syndrome-associated coronavirus. *FEBS Lett* **2006**; 580:3643–8.
13. Keng CT, Choi YW, Welkers MR, et al. The human severe acute respiratory syndrome coronavirus (SARS-CoV) 8b protein is distinct from its counterpart in animal SARS-CoV and down-regulates the expression of the envelope protein in infected cells. *Virology* **2006**; 354:132–42.
14. Lau SK, Feng Y, Chen H, et al. SARS coronavirus ORF8 protein is acquired from SARS-related coronavirus from greater horseshoe bats through recombination. *J Virol* **2015**.
15. Guan Y, Zheng BJ, He YQ, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **2003**; 302:276–8.