Bats Are Natural Reservoirs of SARS-Like Coronaviruses

Wendong Li,^{1,2} Zhengli Shi,^{2*} Meng Yu,³ Wuze Ren,² Craig Smith,⁴ Jonathan H. Epstein,⁵ Hanzhong Wang,² Gary Crameri,³ Zhihong Hu,² Huajun Zhang,² Jianhong Zhang,² Jennifer McEachern,³ Hume Field,⁴ Peter Daszak,⁵ Bryan T. Eaton,³ Shuyi Zhang,^{1,6*} Lin-Fa Wang^{3*}

¹Institute of Zoology of Chinese Academy of Sciences (CAS), Beijing, China. ²State Key Laboratory of Virology, Wuhan Institute of Virology, CAS, Wuhan, China. ³CSIRO Livestock Industries, Australian Animal Health Laboratory, Geelong, Australia. ⁴Department of Primary Industries and Fisheries, Queensland, Australia. ⁵The Consortium for Conservation Medicine, New York, USA. ⁶Guangzhou Institute of Biomedicine and Health, Guangzhou, China.

*To whom correspondence should be addressed. E-mail: zlshi@wh.iov.cn (Z.S.); zhangsy@ioz.ac.cn (S.Z.); linfa.wang@csiro.au (L-F.W.)

Severe acute respiratory syndrome (SARS) emerged in 2002-3 in southern China. The origin of its etiological agent, the SARS coronavirus (SARS-CoV), remains elusive. Here we report that species of bats are a natural host of coronaviruses closely related to those responsible for the SARS outbreak. These viruses, termed SARS-like coronaviruses (SL-CoV), display greater genetic variation than SARS-CoV isolated from humans or civets. The human and civet isolates of SARS-CoV nestle phylogenetically within the spectrum of SL-CoVs, indicating that the virus responsible for the SARS outbreak was a member of this coronavirus group.

Severe acute respiratory syndrome (SARS) was caused by a newly emerged and previously unrecognized coronavirus, now known as SARS coronavirus (SARS-CoV) (1, 2). In spite of the early success of etiological studies and molecular characterization of this new virus (3, 4), efforts to identify the origin of SARS-CoV have been less successful. Without knowledge of the reservoir host distribution and transmission routes of SARS-CoV, it will be difficult to prevent and control future outbreaks of SARS.

Studies conducted previously on animals sampled from live animal markets in Guangdong, China indicated that masked palm civets (*Paguma larvata*) and two other species had been infected by SARS-CoV (5). This led to a large scale culling of civets to prevent further SARS outbreaks. However, subsequent studies have revealed no widespread infection in wild or farmed civets (6, 7). Also, experimental infection of civets with two different human isolates of SARS-CoV resulted in overt clinical symptoms, rendering them unlikely to be the natural reservoir host (8). These data suggest that although *P. larvata* may have been the source of the human infection that precipitated the SARS outbreak, infection in this and other common species in animal markets was more likely a reflection of an 'artificial' market cycle in

naïve species than an indication of the natural reservoir of the virus.

Bats are reservoir hosts of several zoonotic viruses, including Hendra and Nipah viruses, which have recently emerged in Australia and East Asia, respectively (9-11). Bats may be persistently infected with many viruses but rarely display clinical symptoms (12). These characteristics and the increasing presence of bats and bat products in food and traditional medicine markets in southern China and elsewhere in Asia (13) led us to survey bats in the search for the natural reservoir of SARS-CoV.

In this study, conducted during March to December of 2004, 408 bats representing 9 species, 6 genera and 3 families, from four locations in China (Guangdong, Guangxi, Hubei and Tianjin) were sampled by trapping in their native habitat. Blood, fecal and throat swabs were collected; serum samples along with cDNA from fecal or throat samples were independently analyzed, double-blind, using different methods in Wuhan and Geelong (see Supporting Online Materials). The results are summarized in Table 1.

Among six genera of bat species surveyed (*Rousettus*, *Cynopterus*, *Myotis*, *Rhinolophus*, *Nyctalus*, and *Miniopterus*), three communal, cave-dwelling species from the genus *Rhinolophus* (horseshoe bats) in the family *Rhinolophidae* demonstrated a high SARS-CoV antibody prevalence: 13/46 (28%) in *R. pearsoni* from Guangxi; 2/6 (33%) in *R. pusillus* from Guangxi; and 5/7 (71%) in *R. macrotis* from Hubei. The high seroprevalence and wide distribution of seropositive bats is expected for a wildlife reservoir host for a pathogen (14).

The serological findings were corroborated by PCR analyses using primer pairs derived from the nucleocapsid (N) and polymerase (P) genes (table S1). A total of five positive fecal samples were detected, all of them from the genus *Rhinolophus*, three in *R. pearsoni* from Guangxi and one each in *R. macrotis* and *R. ferrumequinum*, respectively,

from Hubei. No virus has been isolated from fecal swabs of PCR-positive samples using Vero E6 cells.

Complete genome sequence was determined directly from PCR products from one of the fecal samples (Rp3) that contained relatively high levels of genetic material. As shown in Fig. 1, the genome organization of this virus, tentatively named SARS-like coronavirus isolate Rp3 (SL-CoV Rp3), was essentially identical to that of SARS-CoV, with the exception of three regions (indicated by the shaded boxes in Fig. 1). The overall nucleotide sequence identity between SL-CoV Rp3 and SARS-CoV Tor2 was 92% and increased to approximately 94% when the three variable regions were excluded. The variable regions are located at the 5' end of the S gene (equivalent to the S1 coding region of coronavirus S protein) and the region immediately upstream of the N gene. These regions have been identified as "high mutation" regions among different SARS-CoVs (5, 15, 16). The region upstream of the N gene is known to be prone to deletions of various sizes (5, 15, 17).

Predicted protein products from each gene or putative ORF of SL-CoV Rp3 and SARS-CoV Tor2 were compared (table S1). The P, S, E, M and N proteins, which are present in all coronaviruses, were similarly sized in the two viruses with sequence identities ranging from 96% to 100%. The only exception was the S1 domain of the S protein where sequence identity fell to 64%. The S1 domain is involved in receptor-binding, whereas the S2 domain is responsible for the fusion of virus and host cell membranes (18). The sequence divergence in the S1 domain corroborated our serum neutralization studies, which indicated that although bat sera have a significant level of cross-reactive antibodies (ELISA titers ranging from 1:100 to 1:6,400), they failed to neutralize SARS-CoV when tested on Vero E6 cells. This finding suggests that S1 is the main target for antibody-mediated virus neutralization of this group of viruses, which is consistent with previous reports indicating that major SARS-CoV neutralization epitopes are located in the S1 region (19, 20).

In addition to the five genes present in all coronavirus genomes, coronaviruses also have several ORFs between the P gene and the 3' end of the genome which code for non-structural proteins. The function of these non-structural proteins is largely unknown. The location and sequence of ORFs are group- or virus-specific, and hence can serve as important molecular markers for studying virus evolution and classification (18, 21). SARS-CoV has a unique set of ORFs not shared by any of the known coronaviruses (3, 4). As detailed in Fig. 1 and table S2, most of these ORFs were also present in SL-CoV, confirming the extremely close genetic relationship between SARS-CoV and SL-CoV.

Coronaviruses produce subgenomic mRNAs through a discontinuous transcription process not fully characterized

(18). Conserved nucleotide sequences functioning as transcription regulatory sequences (TRSs) are required for the production of the subgenomic mRNAs. In SARS-CoV, such TRSs were identified at each of the predicted gene start sites (3, 4). All of these TRSs were absolutely conserved between SARS-CoV Tor2 and SL-CoV Rp3 (table S2), further demonstrating that these two viruses are very closely related.

It is important to note that SL-CoV is completely different from a bat coronavirus (bat-CoV) recently identified by Poon *et al.* (7) from species of bats in the genus *Miniopterus* during a wildlife surveillance study in Hong Kong (Fig. 2). Since the complete genome sequence was not available for bat-CoV, only the trees covering the common sequences (i.e., parts of the P1b protein and S2 protein) are shown. The results demonstrated that while SL-CoV Rp3 and SARS-CoVs are closely related, bat-CoV is placed among the relatively distant group 1 viruses. Hereafter, SARS-CoVs and SL-CoVs will be collectively called the SARS cluster of coronaviruses.

In addition to the complete genome sequence of SL-CoV Rp3, partial genome sequences for the other four PCR-positive bat samples were also determined. Phylogenetic analysis based on the N protein sequences (Fig. 3A) revealed that the genetic variation among the SL-CoV sequences was much greater than that exhibited by SARS-CoVs (for simplicity, only three human and civet SARS-CoV isolates were used since the remainder are almost identical to those shown). This was especially obvious when SL-CoVs isolated from different bat species were compared. Moreover, the results suggested that SARS-CoVs nestle phylogenetically within the spectrum of SL-CoVs.

We also compared the "high mutation" regions in Rf1, Rm1, and Rp3. For the region upstream of the N gene, SL-CoVs from all three bat species contained a single ORF (ORF10'), similar to that found in SARS-CoV isolates from civets (5) and patients in the early phase of the outbreaks (15, 17), but different from most human isolates which have a 29-nt deletion in this region (3, 4, 15). ORF10' in Rf1 codes for a protein having the same size (122 aa) as and more than 80% sequence identity to ORF10' proteins of SARS-CoVs, but those in Rm1 and Rp3 code for a 121-aa protein with only 35% sequence identity (see fig. S2 and Fig. 3B). By contrast, analysis of the S1 protein regions (fig. S3 and Fig. 3C) indicated that Rf1 was more closely related to SL-CoVs from two other bat species than to SARS-CoVs, suggesting that the SARS cluster of coronaviruses could recombine to increase genetic diversity and fitness, as is well documented for other coronaviruses (18). We were unable to sequence these regions for Rp1 or Rp2 owing to the poor quality of the fecal materials from these two animals. The limited amount of cDNA was used up for N gene analysis and in the initial sequencing trials using SARS-CoV-derived primers, which were largely unsuccessful. Judging from the close

relationship of the N genes between Rp1, Rp2 and Rp3 (fig. S1), it is unlikely that Rp1 or Rp2 will have major sequence differences from Rp3 in the S1 or ORF 10' regions. This is not unexpected considering that these three positive samples were obtained from the same bat species in the same location.

The genetic diversity of bat-derived sequences supported the notion that bats are a natural reservoir host of the SARS cluster of coronaviruses. A similar observation has been made for henipaviruses, another important group of emerging zoonotic viruses of bat origin, which show greater genetic diversity in fruit bats in the genus *Pteropus*, than that observed among viruses isolated during the initial disease outbreaks in Malaysia (22–25). It is interesting to note that the overall nucleotide sequence identity of 92% between SL-CoVs and SARS-CoVs is very similar to that observed between Nipah viruses isolated from Malaysia and Bangladesh in 1999 and 2004, respectively (24) (fig. S4). The discovery of SL-CoVs presents a new challenge to the diagnosis and treatment of future disease outbreaks since the current tests and therapeutic strategies may not work effectively against all viruses in this group owing to their great genetic variability in the S1 domain region of the S gene.

The genus *Rhinolophus* contains 69 species and has a wide distribution from Australia to Europe (26). They roost primarily in caves and feed mainly on moths and beetles. However, notwithstanding the predominant *Rhinolophus* findings in this study, it is highly likely that there are more SARS-related coronaviruses to be discovered in bats. Indeed, our positive serological findings in the cave-dwelling fruit bat *Rousettus leschenaulti* indicate that infection by a related virus could occur in fruit bats as well, albeit at a much lower frequency. A plausible mechanism for emergence from a natural bat reservoir can be readily envisaged. Fruit bats including *R. leschenaulti*, and less frequently insectivorous bats, are found in markets in southern China. An infectious consignment of bats serendipitously juxtaposed with a susceptible amplifying species such as *P. larvata* at some point in the wildlife supply chain could result in spillover and establishment of a market cycle while susceptible animals were available to maintain infection. Further studies in field epidemiology, laboratory infection, and receptor distribution and usage are being conducted to assess potential roles played by different bat species in SARS emergence.

These findings on coronaviruses, together with data on henipaviruses (22–24, 27), suggest that genetic diversity exists among zoonotic viruses in bats increasing the possibility of variants crossing the species barrier and causing outbreaks of disease in human populations. It is therefore essential that we enhance our knowledge and understanding of reservoir host distribution, animal-animal and human-animal interaction, particularly within the wet-market system,

and virus genetic diversity of bat-borne viruses to prevent future outbreaks.

References and Notes

1. J. S. M. Peiris *et al.*, *Lancet*, **361**, 1319 (2003).
2. T. G. Ksiazek *et al.*, *N. Engl. J. Med.* **348**, 1953 (2003).
3. P. A. Rota *et al.*, *Science* **300**, 1394 (2003).
4. M. A. Marra *et al.*, *Science* **300**, 1399 (2003).
5. Y. Guan *et al.*, *Science* **320**, 276 (2003).
6. C. Tu *et al.*, *Emerg. Infect. Dis.* **10**, 2244 (2004).
7. L. L. M. Poon *et al.*, *J. Virol.* **79**, 2110 (2005).
8. D. Wu *et al.*, *J. Virol.* **79**, 2620 (2005).
9. K. Murray *et al.*, *Science*, **268**, 94 (1995).
10. K. B. Chua *et al.*, *Science* **288**, 1432 (2000).
11. L.-F. Wang, B. T. Eaton, *Infect. Dis. Rev.* **3**, 52 (2001).
12. S. E. Sulkin, R. Allen, in *Monograph Virol.*, Melnick, Ed. (1974), vol. 8.
13. S.P. Mickleburgh, A. M. Huston, P. A. Racey, *Oryx* **36**, 18 (2002).
14. P. J. Hudson, A. Rizzoli, G. Bryan, Eds. *The Ecology of Wildlife Diseases*, Oxford University Press (2002).
15. Chinese SARS Molecular Epidemiology Consortium, (2004) *Science* **303**, 1666 (2004).
16. P. Li, N. Goldman, *Trends Microbiol.* **12**, 106 (2004).
17. H.-D. Song *et al.*, *Proc. Natl. Acad. Sci. USA* **102**, 2430 (2005).
18. M. C. Lai, K. V. Holmes, in *Fields Virology*, D.M. Knipe, P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, and S.E. Straus, Eds (Lippincott Williams & Wilkins, Philadelphia, 2001) vol. 2, chapter 35.
19. R. A. Tripp *et al.*, *J. Virol. Meth.* **128**, 21 (2005).
20. Y. He, H. Lu, P. Siddiqui, Y. Zhou, S. Jiang, *J. Immunol.* **174**, 4908 (2005).
21. D. A., Brian, R. S. Baric, *Curr. Topics Microbiol. Immunol.* **287**, 1 (2005).
22. S. AbuBakar, L.-Y. Chang, A. R. M. Ali, S. H. Sharifah, K. Yusoff, Z. Zamrod, *Emerg. Infect. Dis.* **10**, 2228 (2004).
23. J.-M. Reynes *et al.*, *Emerg. Infect. Dis.* **11**, 1042 (2005).
24. B. H. Harcourt *et al.*, *Emerg. Infect. Dis.* [serial on the Internet]. Available from <http://www.cdc.gov/ncidod/EID/vol11no10/05-0513.htm> (2005).
25. V. P. Hsu *et al.*, *Emerg. Infect. Dis.* **10**, 2082 (2004).
26. T. H. Kunz, M. B. Fenton, Eds. *Bat Ecology*, University Of Chicago Press (2003).
27. L.-F. Wang, K. B. Chua, M. Yu, B. T. Eaton, *Curr. Genomics* **4**, 263 (2003).
28. This work was jointly funded by a special grant for “Animal Reservoir of SARS-CoV”, State Key Program for Basic Research Grant 2005CB523004, and a State High Technology Development Program Grant 2005AA219070 from the Ministry of Science and Technology, People’s

Republic of China, the Sixth Framework Program “EPISARS” from the European Commission (No. 51163), the Australian Biosecurity CRC for Emerging Infectious Diseases (Project 1.007R), an NIH/NSF “Ecology of Infectious Diseases” (R01-TW05869) award from the John E. Fogarty International Center, and the V. Kann Rasmussen Foundation. For the full-length genome sequence of SL-CoV Rp3, see GenBank accession # DQ71615. Please see SOM for full details of additional GenBank accession numbers.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1118391/DC1

Material and Methods

Figs. S1 to S4

Tables S1 to S3

References. and Notes

4 August 2005; accepted 20 September 2005

Published online 29 September 2005; 10.1126/science.1118391

Include this information when citing this paper.

Fig. 1. Genome organization of, and comparison between, SL-CoV and SARS-CoV. (A) Overall genome organization of SL-CoV Rp3. (B) Expanded diagram of the 3' region of the genome in comparison with SARS-CoV strains Tor2 and SZ3, following the same nomenclature used by Marra *et al.* (4). The genes (named by letters, P, S, E, M and N) present in all coronaviruses are shown in dark-colored arrows whereas the SARS-CoV-specific ORFs are numbered and illustrated in light-colored arrows. ORF10' follows the nomenclature by Guan *et al.* (5) to indicate that the single ORF present between ORF9 and N in SL-CoV is equivalent to the fusion of ORF10 and ORF11 in the same region in SARS-CoV Tor2. The shaded boxes mark the only three regions displaying significant sequence difference between the two viruses. For more detailed comparison of each predicted gene product, please see table S2.

Fig. 2. Phylogenetic trees. (A) and (B) are trees based on deduced amino acid sequences of the same regions in P1b and S, respectively, as used by Poon *et al.* (7) for bat coronavirus (bat-CoV). Abbreviations: Tor2 and SZ3, SARS-CoV strains Tor2 and SZ3; Rp3, SL-CoV Rp3; HCoV, human coronavirus; MHV, mouse hepatitis virus; PEDV, porcine epidemic diarrhea virus; IBV, avian infectious bronchitis virus.

Fig. 3. Phylogenetic trees based on deduced amino acid sequences of N (A), ORF10' (B) and S1 (C) proteins. Abbreviations: Tor2, SZ3 and GD01, different SARS-CoV strains; Rf1, Rm1 and Rp1-3, different SL-CoV sequences. The genetic distance scale shown for A is different from those for B or C.

Table 1. Detection of antibodies to SARS-CoV and PCR amplification of N and P gene fragments using SARS-CoV specific primers. (ND: not determined due to poor sample quality or unavailability of specimens from individual animals.)

Sampling		Bat Species	Antibody Test: Positive/Total (%)	PCR Analysis: Positive/Total (%)	
Time	Location			Fecal swabs	Respiratory swabs
Mar 04	Nanning, Guangxi	<i>Rousettus leschenaulti</i>	1/84 (1.2%)	0/110	ND
	Maoming, Guangdong	<i>Rousettus leschenaulti</i>	0/42	0/45	ND
		<i>Cynopterus sphinx</i>	0/17	0/27	ND
July 04	Nanning, Guangxi	<i>Rousettus leschenaulti</i>	ND	0/55	0/55
	Tianjin	<i>Myotis ricketti</i>	ND	0/21	0/21
		<i>Rhinolophus pusillus</i>	ND	0/15	ND
		<i>Rhinolophus ferrumequinum</i>	0/4	1/8 (12.5%)*	ND
Nov 04	Yichang, Hubei	<i>Rhinolophus macrotis</i>	5/7 (71%)	1/8 (12.5%) [†]	0/3
		<i>Nyctalus plancyi</i>	0/1	0/1	ND
		<i>Miniopterus schreibersi</i>	0/1	0/1	ND
		<i>Myotis altarium</i>	0/1	0/1	ND
		<i>Rousettus leschenaulti</i>	1/58 (1.8)	ND	ND
Dec 04	Nanning, Guangxi	<i>Rhinolophus pearsoni</i>	13/46 (28.3%)	3/30 (10%) [‡]	0/11
		<i>Rhinolophus pussilus</i>	2/6 (33.3%)	0/6	0/2

*Positive fecal sample designated Rf1 [†]Positive fecal sample designated Rm1 [‡]Positive fecal samples designated Rp1, Rp2 and Rp3, respectively.

A **B** **C**