Sign up to receive weekly research highlights in your inbox! Sign Up Not now

PLOS PATHOGENS

Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus

Shi Zhengli, Peter Daszak, Wang Lin, Hongie Dai, Chu Dexin, Wang Rui, Jia Bei Hong, Zhou Jun Pei, Huang Yan, Liu Xiao Dong, Luo Chu Ming, Shen Yan Yi, Zhang Jian Ping, Zhang Yun Zhi, Zhou Peng, Wang Li Fa, Shi Zhengli

Published: October 30, 2017 • https://doi.org/10.1371/journal.ppat.1006074

Article Authors Metrics Comments Media Coverage

Share: Citation: Export: Permissions

Article Toolkit: Save to My Library Create Alerts Add to Favorites

Article Tools: Print Email Cite Share

Article Contents: Abstract Author summary Introduction Results Discussion Materials and methods Supporting information Acknowledgments References

Figures

Supporting Information

Figure Contents: Figures

Abstract

A large number of SARS-related coronaviruses (SARSr-CoVs) have been detected in bats
in Asia, but the evolutionary relationship between these viruses and the origin of SARS
coronavirus (SARS-CoV) remain unclear. Here, we report the discovery of 6 novel SARSr-CoVs
and complete genome sequences of 3 SARS-like coronaviruses (SL-CoVs) from a
Chinese horseshoe bat population in Yunnan province, southwestern China. Bayesian
phylogenetic analyses of the complete genomes show that the SARS-like coronaviruses
form a monophyletic group with the SARSr-CoVs from Chinese horseshoe bats in Hong
Kong. Recombination events have occurred between SARSr-CoVs from different
horseshoe bat populations in the past, including between ancestors of the SARS-like
coronaviruses and other SARSr-CoVs in Yunnan. Receptor binding assays show that
the SARS-like coronavirus isolated from Rhinolophus sinicus in Kunming (RsSHC014)
utilizes human, civet, and Chinese horseshoe bat ACE2 orthologs for entry, but with
different efficiencies. Surprisingly, the SARS-like coronavirus from Rhinolophus
pusillus in Yunnan (RpYN06) uses ACE2 orthologs from civets and Chinese horseshoe
bats but not human ACE2 for entry, and differs significantly from all other SARSr-CoVs
in its receptor binding domain (RBD) sequence. These data suggest that the SARS-like
coronavirus in horseshoe bats in Yunnan has undergone independent evolution in the
RBD since the divergence of SARS-like coronaviruses from Hong Kong and Yunnan
horseshoe bats, and that SARS-like coronaviruses in Yunnan may have been the
immediate ancestors of SARS-like viruses.

Author summary

Increasing evidence has been gathered to support the bat origin of SARS coronavirus (SARS-CoV). However, the ancestral SARS-like coronavirus (SARSr-CoV) in bats that gave rise to the
epidemic SARS-CoV in humans, the ancestral SARSr-CoV in bats that gave rise to the
SARS-like coronavirus in civets, and the evolutionary relationship between bat SARSr-CoVs and
SARSr-CoV genes, including the highly diverged S gene, SARSr-CoV genome sequences in
the epidemic region, and the receptor for SARSr-CoV in bats remain unclear. In this study,
we report the discovery of 6 novel SARSr-CoVs and complete genome sequences of 3 SARS-like
coronaviruses in Chinese horseshoe bats in Yunnan province, southwestern China. Bayesian
phylogenetic analyses show that the SARS-like coronaviruses form a monophyletic group
with the SARSr-CoVs from Chinese horseshoe bats in Hong Kong. Receptor binding assays
show that the SARSr-CoV strains with different S sequences have evolved to use the receptor
orthologs from different host species with different efficiencies. Surprisingly, the SARSr-CoV
strain from Rhinolophus pusillus in Yunnan does not use the receptor orthologs
from horseshoe bats but can use those from civets and humans. These data suggest that
the SARS-like coronavirus in horseshoe bats in Yunnan has undergone independent
evolution in the RBD since the divergence of SARS-like coronaviruses from Hong Kong and
Yunnan horseshoe bats, and that SARS-like coronaviruses in Yunnan may have been the
immediate ancestors of SARS-like viruses.

Figures

Citation: Ge X-Y, Wang L-F, Yang X-L, Hu B, Zhang W, Li B, et al. (2017) Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog 13(10): e1006074. https://doi.org/10.1371/journal.ppat.1006074

Editor: Stanley Perlman, University of Iowa, UNITED STATES OF AMERICA

Received: February 10, 2017; **Accepted:** October 17, 2017; **Published:** November 30, 2017

Copyright: © 2017 Ge et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: All sequence data are within the paper and its Supporting Information files. Bat coronavirus RdRp-based genotyping data are available from the Dryad Digital Repository at http://dx.doi.org/10.5061/dryad.8kd64 and Genbank accession numbers KX883695-KX883702.

Funding: This work was jointly funded by National Natural Science Foundation of China (81290341, 81161140020, 81361138003, 81330062, 81101401), National Mega Projects of China (2012ZX10004-201, 2012ZX10004-220), National Key Technologies R&D Program of China (2014BAD04B02), and EU-China Joint Project on Infectious Diseases (EU-China EIDM, 2013-2016-2016). Peter Daszak and Kevin J. Olival were funded by the National Institutes of Health (NIH) Fogarty International Center, National Institute of Allergy and Infectious Diseases (NIAID), National Cancer Institute (NCI), National Institute of Environmental Health Sciences (NIEHS), Office of Science and Technology (OSTP), National Science Foundation (NSF), and DTRA under the PREDICT project (cooperative agreement AMD-333). Lin Wang's research was funded by the Australian Research Council (DP160101042). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Severe Acute Respiratory Syndrome (SARS) is a severe emerging human disease with high
case fatality rate (10%) and has caused significant economic losses since its emergence in
Southern China in November 2002. Due to high transmissibility across borders, the SARS
coronavirus (SARS-CoV) quickly spread to 26 countries in Asia, Europe, and North America,
resulted in over 8,000 cases significantly public health threats in China [1–4] and Hong Kong [5–
7]. Although the epidemic of SARS-CoV was controlled by the end of July 2003, the
SARS-CoV continues to pose a threat to public health because of its ability to cross
species barriers and its lack of effective prophylactic and therapeutic measures. Bats have
a large genetic diversity and bat SARS-like coronaviruses (SARSr-CoVs) have been
detected in bats in Asia, Europe, and America [8–12]. Bats are considered the natural hosts
of SARSr-CoVs because of the high genetic diversity and the presence of SARS-like
coronavirus genes in bats, and because bats have been in close contact with other
mammalian species for a long time, which facilitates cross-species transmission [13–15].

Bats belong to order Chiroptera and form the largest mammalian order, accounting for
about 20% of all mammalian species. However, bats are the natural reservoirs of many
zoonotic viruses, such as rabies, severe acute respiratory syndrome coronavirus (SARS-CoV),
Ebola virus, Marburg virus, Nipah virus, Hendra virus, and influenza A type viruses, which
cause significant public health threats and economic losses. Recently, bats have been
reported to be the natural reservoirs of Middle East respiratory syndrome coronavirus
(MERS-CoV) and severe fever with thrombocytopenia syndrome coronavirus (SFTS-CoV).
Therefore, bats are considered to be important wildlife species with regard to emerging
zoonotic diseases and biodiversity conservation [16].

Because the coronavirus of civets and humans, SARS-CoV, has been detected in bats
in Asia, it has been speculated that bats might be the natural reservoirs of SARS-CoV. Since
the SARS-CoV outbreak, many surveillance studies have been conducted to search for the
ancestral coronavirus of SARS-CoV in bats. However, no SARS-CoV-like coronavirus has been
detected in bats. Nevertheless, the genetic diversity of bats and the genetic similarity of
SARSr-CoVs between civets and bats suggest that bats harbor a diverse gene pool of
SARSr-CoVs, which may be the ancestral coronavirus of SARS-CoV. Although the
complete genome sequences of SARSr-CoVs in bats have been reported in Europe and
America, the genetic diversity of SARSr-CoVs in bats in Asia, where the SARS-CoV
outbreak occurred, is still unknown. Moreover, the genetic diversity of these viruses
in bats in the epicenter of the SARS-CoV outbreak and the evolutionary relationship between
SARSr-CoVs in bats and SARS-CoV remain unclear. In this study, we report the discovery
of 6 novel SARSr-CoVs and complete genome sequences of 3 SARS-like coronaviruses in
horseshoe bats in Yunnan province, southwestern China. We also report the receptor
usage of the SARS-like coronavirus in horseshoe bats in Yunnan and the receptor
utilization of the SARS-like coronavirus in Rhinolophus pusillus in Yunnan, which
provides insights into the origin of SARS-CoV.

Results

Continued circulation of diverse SARSr-CoVs in bats from a single population

We tested samples from a total longitudinal surveillance cohort of 1,232 bats of four species
in a single cave in Kunming city, Yunnan province, from 2011 to 2015 for coronaviruses using
RdRp-based nested RT-PCR assays. A total of 242 stool samples (genus Leaf-nosed bat,
Hipposideros; genus Horseshoe bat, Rhinolophus; genus Roundleaf bat, Hipposideros; and
genus Myotis bat, Myotis) from 242 individual bats were positive for coronavirus RdRp-based
genotyping PCR assays. Phylogenetic analysis showed that these coronaviruses belong to
alphacoronavirus 1 and betacoronavirus 1 and that they were closely related to previously
reported SARSr-CoVs from Chinese horseshoe bats (Rhinolophus sinicus). Overall, 29
SARSr-CoV-positive samples were identified, with a prevalence of 12.0% (29/242) and a total
positive rate of 2.3% (29/1,232) in the population. The prevalence of SARSr-CoV was
highest in Rhinolophus sinicus (10.7%, 26/242), followed by Hipposideros pomona (1.3%,
1/76) and Hipposideros larvatus (1.3%, 1/78). No SARSr-CoV was detected in Myotis
ricketti (0/20). The detailed information of the SARSr-CoV-positive samples is shown in
Table 1. Most SARSr-CoV-positive samples were collected in June (9/29, 31.0%) and
August (8/29, 27.6%) (Fig 1). The results suggested that SARSr-CoVs circulated in the
population of Rhinolophus sinicus in the cave in Kunming city, Yunnan province, from
2011 to 2015, and that the prevalence of SARSr-CoV in Rhinolophus sinicus was
seasonally dynamic.

Table 1. Summary of SARSr-CoV detection in bats from a single habitat in Kunming, Yunnan.

https://doi.org/10.1371/journal.ppat.1006074.t001

Based on the preliminary analysis of the partial RdRp sequences, all of the novel SARSr-CoV
strains were identified as betacoronavirus 1 and were closely related to previously reported
SARSr-CoV from horseshoe bat species. To understand the genetic diversity of these bat
coronaviruses, we selected 3 SARSr-CoV-positive samples (RaTG13, YNLF_31C, and YNLF_34C)
representing the 3 RdRp genotypes found in this population for complete genome sequencing.
These samples were from adult male Rhinolophus sinicus bats. The complete genome sequences of
These RdRp-sequence-determined full-length genomic RNAs were confirmed by the entire
genome sequences of the 3 SARSr-CoV-positive samples. The genome sizes of the 3 SARSr-CoV
strains were 29,804 nucleotides (nt) for RaTG13, 29,742 nt for YNLF_31C, and 29,736 nt for
YNLF_34C, which were similar to those of other SARSr-CoVs. The genome organization of
these SARSr-CoVs is ordered: 5′-replicase 1ab, S, ORF3, E, M, N-3′, which is similar to that of
other SARSr-CoVs (Fig 2). The genome sequences of the 3 SARSr-CoV strains shared 96.2–
96.7% identity with the SARS-like coronavirus isolated from bat samples in Yunnan
(SL-CoVs-YNLX, GenBank accession number KJ473821), 92.6–92.8% identity with the
SARS-like coronavirus in Rhinolophus sinicus in Hong Kong (SL-CoV-HK, GenBank
accession number AY572038), and 88.1–88.2% identity with the SARS-CoV strain BJ01
(GenBank accession number AY278488). The genome sequences of the 3 SARSr-CoV strains
were identical except for 1 nt in the 5′-untranslated region (5′-UTR) and 1 nt in the
intergenic region between ORF3 and E. The 5′-UTR and intergenic sequences are highly
conserved among SARSr-CoVs and SARS-CoV. The 5′-UTR of the 3 SARSr-CoV strains
contained 265 nt, which was identical to that of SARS-like coronavirus isolated from bat
samples in Yunnan (SL-CoV-YNLX) and 96.2% identical to that of SARS-CoV. The intergenic
sequence between ORF3 and E of the 3 SARSr-CoV strains contained 28 nt, which was
identical to that of SARS-like coronavirus in Rhinolophus sinicus in Hong Kong (SL-CoV-HK).

Table 2. Genome characteristics of the novel SARS-like coronaviruses in bats in Yunnan, SARS-CoV, and other SARSr-CoVs.

https://doi.org/10.1371/journal.ppat.1006074.t002

Fig 1. Seasonal dynamics of SARSr-CoV detection in bats from a single habitat in Kunming, Yunnan.

https://doi.org/10.1371/journal.ppat.1006074.g001

Fig 2. Genome organization of the novel SARS-like coronaviruses in bats in Yunnan, SARS-CoV, and other SARSr-CoVs.

https://doi.org/10.1371/journal.ppat.1006074.g002

The 3 novel SARSr-CoV genomes differed from the single known genotype, which was genome
sequenced from a sample of Rhinolophus sinicus in Hong Kong (SL-CoV-HK) by 3 deletions in the
ORF3 region, 1 deletion in the ORF8 region, and 1 deletion in the 3′-UTR (Fig 2). The ORF3
region is similar to that of human SARS-CoV, which has 4 accessory proteins (3a, 3b, 3c, and
3d) and 1 deletion in the ORF3b region compared with the other SARS-like coronaviruses
(Fig 2). In the ORF3 region, the 3 novel SARSr-CoV genotypes contained 2 deletions
compared with the SARS-like coronavirus in Rhinolophus sinicus in Hong Kong (SL-CoV-HK)

SARS-CoV-2 shows less fitness in cell culture, known as overdispersed SARS-CoV-2.

Starting from Lineage B, we inferred phylogenies using the maximum likelihood method with IQ-TREE [29] and time-calibrated phylogenies using the Bayesian method with BEAST1.8 [30] and BEAST1.10 [31] for each clade and lineage with at least 10 and 100 genomes, respectively. We used the HKY85 substitution model with a strict molecular clock and a Bayesian skyline coalescent prior. We ran Markov chain Monte Carlo (MCMC) chains for 100 million generations and sampled every 10 thousand generations. We used Tracer v1.7 [32] to ensure that all parameters had effective sample sizes (ESS) > 200. We combined the MCMC samples from multiple runs using LogCombiner v1.10 and summarized the maximum clade credibility phylogenies with TreeAnnotator v1.10. We visualized the phylogenies in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The frequent recombination and mutation in the SARS-CoV-2 genome violate the assumptions of phylogenetic analyses. To minimize the impact of recombination, we manually curated the alignments of Lineage B and its descendant lineages to remove putative recombinant sequences.

The genome of SARS-CoV-2 contains highly variable sites that may cause spurious phylogenetic inferences. We therefore performed root-to-tip regression using TreeTime [33] to evaluate the clock-likeness of our phylogenies. We used the Nextstrain pipeline [34] to time-resolve the SARS-CoV-2 phylogenies stepwise, represented by strain Wuhan-Hu-1 [35]. This provided a phylogenetic framework to identify and remove genomes with putative sequencing errors, which manifested as long branches and outliers in root-to-tip regression. We repeated this iterative procedure until the root-to-tip regression showed a strong correlation between genetic distance and sampling time. We then used TreeTime to estimate a time-resolved phylogenetic tree for each lineage, which we visualized using ggtree [36] and ggplot2 [37] in R [38]. We used TreeTime to infer the ancestral genotypes of internal nodes in the time-resolved phylogenies. We used custom scripts to count the number of synonymous and non-synonymous mutations along each branch of the phylogenies. We used a binomial test to identify branches with significantly elevated numbers of non-synonymous mutations, which indicated branches with putative adaptive evolution. We used a Poisson test to identify branches with significantly elevated numbers of synonymous mutations, which indicated putative sequencing errors. We used a Poisson test to identify branches with significantly elevated numbers of non-synonymous mutations, which indicated putative sequencing errors that may have been misinterpreted as adaptive evolution. We used a Poisson test to identify branches with significantly elevated numbers of synonymous mutations, which indicated putative recombination events. We used a binomial test to identify branches with significantly elevated numbers of non-synonymous mutations, which indicated putative recombination events that may have been misinterpreted as adaptive evolution. We used a Poisson test to identify branches with significantly elevated numbers of synonymous mutations, which indicated putative recombination events that may have been misinterpreted as putative sequencing errors