LETTER

Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor

Xing-Yi Ge1,*, Jia-Lu Li1,*, Xing-Lou Yang1,*, Aleksei A. Chmura2, Guangjian Zhu2, Jonathan H. Epstein2, Jonna K. Mazet3, Ben Hu1, Wei Zhang1, Cheng Peng1, Yu-Ji Zhang1, Chu-Ming Luo1, Bing Tan1, Ning Wang1, Yan Zhu1, Gary Crameri4, Shu-Yi Zhang5, Lin-Fa Wang4,6, Peter Daszak2 & Zheng-Li Shi1

The 2002–3 pandemic caused by severe acute respiratory syndrome coronavirus (SARS-CoV) was one of the most significant public health events in recent history1. An ongoing outbreak of Middle East respiratory syndrome coronavirus2 suggests that this group of viruses remains a key threat and that their distribution is wider than previously recognized. Although bats have been suggested to be the natural reservoirs of both viruses3–5, attempts to isolate the progenitor virus of SARS-CoV from bats have been unsuccessful. Diverse SARS-like coronaviruses (SL-CoVs) have now been reported from bats in China, Europe and Africa6–8, but none is considered a direct progenitor of SARS-CoV because of their phylogenetic disparity from this virus and the inability of their spike proteins to use the SARS-CoV cellular receptor molecule, the human angiotensin converting enzyme II (ACE2)9,10. Here we report whole-genome sequences of two novel bat coronaviruses from Chinese horseshoe bats (family: Rhinolophidae) in Yunnan, China: RsSHC014 and Rs3367. These viruses are far more closely related to SARS-CoV than any previously identified bat coronaviruses, particularly in the receptor binding domain of the spike protein. Most importantly, we report the first recorded isolation of a live SL-CoV (bat SL-CoV-WIV1) from bat faecal samples in Vero E6 cells, which has typical coronavirus morphology, 99.9% sequence identity to Rs3367 and uses ACE2 from humans, civets and Chinese horseshoe bats for cell entry. Preliminary in vitro testing indicates that WIV1 also has a broad species tropism. Our results provide the strongest evidence to date that Chinese horseshoe bats are natural reservoirs of SARS-CoV, and that intermediate hosts may not be necessary for direct human infection by some bat SL-CoVs. They also highlight the importance of pathogen-discovery programs targeting high-risk wildlife groups in emerging disease hotspots as a strategy for pandemic preparedness.

The 2002–3 pandemic of SARS1 and the ongoing emergence of the Middle East respiratory syndrome coronavirus (MERS-CoV)2 demonstrate that CoVs are a significant public health threat. SARS-CoV was shown to use the human ACE2 molecule as its entry receptor, and this is considered a hallmark of its cross-species transmissibility11. The receptor binding domain (RBD) located in the amino-terminal region (amino acids 318–510) of the SARS-CoV spike (S) protein is directly involved in binding to ACE2 (ref. 12). However, despite phylogenetic evidence that SARS-CoV evolved from bat SL-CoVs, all previously identified SL-CoVs have major sequence differences from SARS-CoV in the RBD of their S proteins, including one or two deletions6,8. Replacing the RBD of one SL-CoV S protein with SARS-CoV S conferred the ability to use human ACE2 and replicate efficiently in mice6,13. However, to date, no SL-CoVs have been isolated from bats, and no wild-type SL-CoV of bat origin has been shown to use ACE2.

We conducted a 12-month longitudinal survey (April 2011–September 2012) of SL-CoVs in a colony of Rhinolophus sinicus at a single location in Kunming, Yunnan Province, China (Extended Data Table 1). A total of 117 anal swabs or faecal samples were collected from individual bats using a previously published method5,14. A one-step reverse transcription (RT)-nested PCR was conducted to amplify the RNA-dependent RNA polymerase (RdRP) motifs A and C, which are conserved among alphacoronaviruses and betacoronaviruses15.

Twenty-seven of the 117 samples (23%) were classed as positive by PCR and subsequently confirmed by sequencing. The species origin of all positive samples was confirmed to be R. sinicus by cytochrome b sequence analysis, as described previously16. A higher prevalence was observed in samples collected in October (30% in 2011 and 48.7% in 2012) than those in April (7.1% in 2011) or May (7.4% in 2012) (Extended Data Table 1). Analysis of the S protein RBD sequences indicated the presence of seven different strains of SL-CoVs (Fig. 1a and Extended Data Figs 1 and 2). In addition to RBD sequences, which closely matched previously described SL-CoVs (Rs672, Rf1 and HKU3)5,8,17,18, two novel strains (designated SL-CoV RsSHC014 and Rs3367) were discovered. Their full-length genome sequences were determined, and both were found to be 29,787 base pairs in size (excluding the poly(A) tail). The overall nucleotide sequence identity of these two genomes with human SARS-CoV (Tor2 strain) is 95%, higher than that observed previously for bat SL-CoVs in China (88–92%)5,8,17,18 or Europe (76%)6 (Extended Data Table 2 and Extended Data Figs 3 and 4). Higher sequence identities were observed at the protein level between these new SL-CoVs and SARS-CoVs (Extended Data Tables 3 and 4). To understand the evolutionary origin of these two novel SL-CoV strains, we conducted recombination analysis with the Recombination Detection Program 4.0 package19 using available genome sequences of bat SL-CoVs (Rs1, Rp3, Rs672, Rm1, HKU3 and BM48-31) and human and civet representative SARS-CoV strains (BJ01, SZ3, Tor2 and GZ02). Three breakpoints were detected with strong P values (<10−20) and supported by similarity plot and bootstrap analysis (Extended Data Fig. 5a, b). Breakpoints were located at nucleotides 20,827, 26,553 and 28,685 in the Rs3367 (and RsSHC014) genome, and generated recombination fragments covering nucleotides 20,827–26,553 (5,727 nucleotides) (including partial open reading frame (ORF) 1b, full-length S, ORF3, E and partial M gene) and nucleotides 26,534–28,685 (2,133 nucleotides) (including partial ORF M, full-length ORF6, ORF7, ORF8 and partial N gene). Phylogenetic analysis using the major and minor parental regions suggested that Rs3367, or RsSHC014, is the descendent of a recombination of lineages that ultimately lead to SARS-CoV and SL-CoV Rs672 (Fig. 1b).

The most notable sequence differences between these two new SL-CoVs and previously identified SL-CoVs is in the RBD regions of their S proteins. First, they have higher amino acid sequence identity to SARS-CoV (85% and 96% for RsSHC014 and Rs3367, respectively). Second, there are no deletions and they have perfect sequence alignment with the SARS-CoV RBD region (Extended Data Figs 1 and 2). Structural
Figure 1 | Phylogenetic tree based on amino acid sequences of the S RBD region and the two parental regions of bat SL-CoV Rs3367 or RsSHC014. 
a. SARS-CoV S protein amino acid residues 310–520 were aligned with homologous regions of bat SL-CoVs using the ClustalW software. A maximum-likelihood phylogenetic tree was constructed using a Poisson model with bootstrap values determined by 1,000 replicates in the MEGA5 software package. The RBD sequences identified in this study are in bold and named by the sample numbers. The key amino acid residues involved in interacting with the human ACE2 molecule are indicated on the right of the tree. SARS-CoV GZ02, BJ01 and Tor2 were isolated from patients in the early, middle and late phase, respectively, of the SARS outbreak in 2003. SARS-CoV SZ3 was identified from Guangdong, China in 2005/2006 and used as an outgroup. All sequences in bold and italics were identified in the current study. Filled triangles, circles and diamonds indicate samples with co-infection by two different SL-CoVs. ‘–’ indicates the amino acid deletion.

b. Phylogenetic origins of the two parental regions of Rs3367 or RsSHC014. Maximum likelihood phylogenetic trees were constructed from alignments of two fragments covering nucleotides 20,827–26,533 (5,727 nucleotides) and 26,534–28,685 (2,133 nucleotides) of the Rs3367 genome, respectively. For display purposes, the trees were midpoint rooted. The taxa were annotated according to strain names: SARS-CoV, SARS-like CoV, bat SARS-like coronavirus. The two novel SL-CoVs, Bat SL-CoV Rp3, Rs672 and HKU3-1 were identified from R. sinicus collected in Guangdong, China. SL-CoV Rp3, Rs672 and HKU3-1 were identified from R. sinicus collected in China (respectively: Guangxi, 2004; Guangdong, 2005; Hong Kong, 2005). Rf1 and Rm1 were identified from Paguma larvata in 2003 collected in Guangdong, China. SL-CoV Rs4084, Rs4085, Rs4087-1, Rs4079, Rs4087-2, Rs4081, Rs4087, Rs4080, Rs4081, Rs4097 were identified from bats collected in Bulgaria in 2008. Bat SL-CoV RsSHC014 was identified from Rousettus leschenaultii collected in Guangdong, China in 2005/2006 and used as an outgroup. All sequences in bold and italics were identified in the current study. Filled triangles, circles and diamonds indicate samples with co-infection by two different SL-CoVs. ‘–’ indicates the amino acid deletion. 

To determine whether WIV1 can use ACE2 as a cellular entry receptor, we conducted virus infectivity studies using HeLa cells expressing or not expressing ACE2 from humans, civets or Chinese horseshoe bats. We found that WIV1 is able to use ACE2 of different origins as an entry receptor and replicated efficiently in the ACE2-expressing cells (Fig. 3). This is, to our knowledge, the first identification of a wild-type bat SL-CoV capable of using ACE2 as an entry receptor.

To assess its cross-species transmission potential, we conducted infection assays in cell lines from a range of species. Our results (Fig. 4 and Extended Data Fig. 1) indicate that bat SL-CoV-WIV1 can grow in human alveolar basal epithelial (A549), pig kidney 15 (PK-15) and Rhinolophus sinicus kidney (RSKT) cell lines, but not in human cervix (HeLa), Syrian golden hamster kidney (BHK21), Myotis davidi kidney (BK), Myotis chimensis kidney (MCKT), Rousettus leschenaultii kidney (RLK) or Pteropus alecto kidney (PaKi) cell lines. Real-time RT–PCR indicated that WIV1 replicated much less efficiently in A549, PK-15 and RSKT cells than in Vero E6 cells (Fig. 4).
To assess the cross-neutralization activity of human SARS-CoV sera against WIV1, we conducted serum-neutralization assays using nine convalescent sera from SARS patients collected in 2003. The results showed that seven of these were able to completely neutralize 100 tissue culture infectious dose 50 (TCID$_{50}$) WIV1 at dilutions of 1:10 to 1:40, further confirming the close relationship between WIV1 and SARS-CoV.

Our findings have important implications for public health. First, they provide the clearest evidence yet that SARS-CoV originated in bats. Our previous work provided phylogenetic evidence of this$^5$, but the lack of an isolate or evidence that bat SL-CoVs can naturally infect human cells, until now, had cast doubt on this hypothesis. Second, the lack of capacity of SL-CoVs to use of ACE2 receptors has previously been considered as the key barrier for their direct spillover into humans, supporting the suggestion that civets were intermediate hosts for SARS-CoV adaptation to human transmission during the SARS outbreak$^{24}$. However, the ability of SL-CoV-WIV1 to use human ACE2 argues against the necessity of this step for SL-CoV-WIV1 and suggests that direct bat-to-human infection is a plausible scenario for some bat SL-CoVs. This has implications for public health control measures in the face of potential spillover of a diverse and growing pool of recently discovered SARS-like CoVs with a wide geographic distribution.

Our findings suggest that the diversity of bat CoVs is substantially higher than that previously reported. In this study we were able to demonstrate the circulation of at least seven different strains of SL-CoVs within a single colony of $R$. sinicus during a 12-month period. The high genetic diversity of SL-CoVs within this colony was mirrored by high phenotypic diversity in the differential use of ACE2 by different strains. It would therefore not be surprising if further surveillance reveals a broad diversity of bat SL-CoVs that are able to use ACE2, some of which may have even closer homology to SARS-CoV than SL-CoV-WIV1. Our results—in addition to the recent demonstration of MERS-CoV in a Saudi Arabian bat$^{25}$, and of bat CoVs closely related to MERS-CoV in China, Africa, Europe and North America$^{3,26,27}$—suggest that bat coronaviruses remain a substantial global threat to public health.

Finally, this study demonstrates the public health importance of pathogen discovery programs targeting wildlife that aim to identify the ‘known unknowns’—previously unknown viral strains closely related to known pathogens. These programs, focused on specific high-risk wildlife groups and hotspots of disease emergence, may be a critical part of future global strategies to predict, prepare for, and prevent pandemic emergence$^{28}$.

---

**Figure 2 | Electron micrograph of purified virions.** Virions from a 10-ml culture were collected, fixed and concentrated/purified by sucrose gradient centrifugation. The pelleted viral particles were suspended in 100 µl PBS, stained with 2% phosphotungstic acid (pH 7.0) and examined directly using a Tecnai transmission electron microscope (FEI) at 200 kV.

---

**Figure 3 | Analysis of receptor usage of SL-CoV-WIV1 determined by immunofluorescence assay and real-time PCR.** Determination of virus infectivity in HeLa cells with and without the expression of ACE2. b, bat; c, civet; h, human. ACE2 expression was detected with goat anti-humanACE2 antibody followed by fluorescein isothiocyanate (FITC)-conjugated donkey anti-goat IgG. Virus replication was detected with rabbit antibody against the SL-CoV Rp3 nucleocapsid protein followed by cyanine 3 (Cy3)-conjugated mouse anti-rabbit IgG. Nuclei were stained with DAPI (4′,6-diamidino-2-phenylindole). The columns (from left to right) show staining of nuclei (blue), ACE2 expression (green), virus replication (red), merged triple-stained images and real-time PCR results, respectively. ($n = 3$); error bars represent standard deviation.
Throat and faecal swabs or fresh faecal samples were collected in viral transport medium as described previously. All PCR was conducted with the One-Step RT-PCR kit (Invitrogen). Primers targeting the highly conserved regions of the RDRP gene were used for detection of all alphacoronaviruses and betacoronaviruses as described previously. Degenerate primers were designed on the basis of all available genomic sequences of SARS-CoVs and SL-CoVs and used for amplification of the RBD sequences of 5 genes or full-length genomic sequences. Degenerate primers were used for amplification of the bat ACE2 gene as described previously. PCR products were gel purified and cloned into pGEM-T Easy Vector (Promega). At least four independent clones were sequenced to obtain a consensus sequence. PCR-positive faecal samples (in 200 µl buffer) were gradient centrifuged at 3,000–12,000 g and supernatant diluted at 1:10 in DMEM before being added to Vero E6 cells. After incubation at 37 °C for 1 h, inocula were removed and replaced with fresh DMEM with 2% FCS. Cells were incubated at 37 °C and checked daily for cytopathic effect. Cell lines from different origins were grown on coverslips in 24-well plates and inoculated with the novel SL-CoV at a multiplicity of infection of 10. Virus replication was detected at 24 h after infection using rabbit antibodies against the SL-CoV Rp3 nucleocapsid protein followed by Cy3-conjugated goat anti-rabbit IgG.

**METHODS SUMMARY**

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 May; accepted 18 September 2013.

**Acknowledgements** We acknowledge financial support from the State Key Program for Basic Research (2011CB504701 and 2010CB530100), National Natural Science Foundation of China (81290341 and 31321001), Scientific and technological basis for Basic Research (2011CB504701 and 2010CB530100), National Natural Science Foundation of China (81290341 and 31321001), Scientific and technological basis for Basic Research (2011CB504701 and 2010CB530100), National Natural Science Foundation of China (81290341 and 31321001), Scientific and technological basis for Basic Research (2011CB504701 and 2010CB530100), National Natural Science Foundation of China (81290341 and 31321001), Scientific and technological basis for Basic Research (2011CB504701 and 2010CB530100), and United States Agency for International Development (USAID) Emerging Pandemic Threats PREDICT. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID, NIH, NSF, USAID or the United States Government. We thank C. Che from Zhujian Hospital, Southern Medical University, for providing human SARS patient sera.

**Author Contributions** Z.-L.S. and P.D. designed and coordinated the study. X.-Y.G., J.-L. and X.-L.Y. conducted majority of experiments and contributed equally to the study. X.-Y.G. and J.-L. supervised part of the experiments. All authors contributed to the interpretations and conclusions presented. Z.-L.S. and X.-Y.G. wrote the manuscript with significant contributions from P.D. and L.F.W. and input from all authors.

**Author Information**

**Funding Information** The study was supported by grants from the National Natural Science Foundation of China (81290341 and 31321001), National Institutes of Health (NIH)/National Science Foundation (NSF) ‘Ecology and Evolution of Infectious Diseases’ the beginning of another SARS-like pandemic? (27057), NIH/Fogarty International Center (D43 TW002856), and the Rockefeller Foundation (K08TW002856). The contents are the responsibility of the authors and do not necessarily reflect the views of NIH, NIH, NSF, USAID or the United States Government.

**Contact Information** Correspondence and requests for materials should be addressed to P.D. (daszak@ecolealthalliance.org) or Z.-L.S. (zhli@whi.ucn).
METHODS

Sampling. Bats were trapped in their natural habitat as described previously. Throat and fecal snap samples were collected in viral transport medium (VTM) and placed in VTM. Samples were transported to the laboratory and stored at −80 °C until use. All animals trapped for this study were released back to their habitat after sample collection. All sampling processes were performed by veterinarians with approval from Animal Ethics Committee of the Wuhan Institute of Virology (WIV-HW2012001) and EcoHealth Alliance under an inter-institutional agreement with University of California, Davis (UC Davis protocol no. 16048).

RNA extraction, PCR and sequencing. RNA was extracted from 140 µl of saliva or fecal samples with a Viral RNA Mini Kit (Qiagen) following the manufacturer’s instructions. RNA was eluted in 60 µl RNAase-free buffer (buffer AVE, Qiagen) and aliquoted and stored at −80 °C. One-step RT–PCR (Invitrogen) was used to detect coronavirus sequences as described previously. First round PCR was conducted in a 25-µl reaction mix containing 1.25 µl PCR reaction mix buffer, 10 pmol of each primer, 2.5 mM MgSO4, 20 U RNase inhibitor, 1 µl SuperScript III Platinum Taq Enzyme Mix and 5 µl RNA. Amplification of the RdRP-gene fragment was performed as follows: 50 °C for 30 min, 94 °C for 2 min, followed by 40 cycles consisting of 94 °C for 15 s, 62 °C for 15 s, 68 °C for 40 s, and a final extension of 68 °C for 5 min. Second round PCR was conducted in a 25-µl reaction mix containing 2.5 µl PCR reaction buffer, 5 pmol of each primer, 50 mM MgCl2, 0.5 mM dNTP, 0.1 µl Platinum Taq Enzyme (Invitrogen) and 1 µl first round PCR product. The amplification of RdRP-gene fragment was performed as follows: 94 °C for 5 min followed by 35 cycles consisting of 94 °C for 30 s, 52 °C for 30 s, 72 °C for 40 s, and a final extension of 72 °C for 5 min.

To amplify the RdR region, one-step RT–PCR was performed with primers designed based on available SARS-CoV or bat SL-CoV sequences (first round PCR primers: F, forward, R, reverse: CoVS931F-5'-VWAGDGTGKTGKRCTTYCT-3' and CoVS16995-5'-TAARACAVCCWGCYTGTW-3'). Second round PCR was conducted in a 25-µl reaction mix containing 2.5 µl PCR reaction buffer, 5 pmol of each primer, 50 mM MgCl2, 0.5 mM dNTP, 0.1 µl Platinum Taq Enzyme (Invitrogen) and 1 µl first round PCR product. The amplification of RdRP-gene fragment was performed as follows: 94 °C for 5 min followed by 40 cycles consisting of 94 °C for 15 s, 60 °C for 30 s, 72 °C for 40 s, and a final extension of 72 °C for 5 min.

Sequencing full-length genomes. Degenerate coronavirus primers were designed based on all available SARS-CoV and bat SL-CoV sequences in GenBank and specific primers were designed from genome sequences generated from previous rounds of sequencing in this study (primer sequences will be provided upon request). All PCRs were conducted using the One-step RT–PCR kit (Invitrogen). The 5′ and 3′ genomic ends were determined using the 5′ or 3′ RACE kit (Roche), respectively. PCR products were gel purified and sequenced directly or following cloning into pGEM-T Easy Vector (Promega). At least four independent clones were sequenced to obtain a consensus sequence for each of the amplified regions.

Sequence analysis and database accession numbers. Routine sequence management and analysis was carried out using DNAStar or Geneious. Sequence alignment and editing was conducted using ClustalW, BioEdit or GeneDoc. Maximum Likelihood phylogenetic trees based on the protein sequences were constructed using a Poisson model with bootstrap values determined by 1,000 replicates in the MEGA5 software package.

Sequences obtained in this study have been deposited in GenBank as follows (accession numbers given in parenthesis): full-length genome sequence of SL-CoV RsSHC014 and Rs3367 (KC881105, KC881106); full-length sequence of WIV1 S (KC881007); RBD (KC880984, KC881003); ACE2 (KC881004). SARS-CoV sequences (accession numbers given in parenthesis) from this study: human SARS-CoV-2 (AY278488), GZ02 (AY390556) and civet SARS-CoV strain SZ3 (AY304846). Coronavirus sequences used in this study: Rs672 (FJ588686), Rp3 (Q071615), Rf1 (Q071614), Rf2 (Q071612), Rf3 (Q071611), Rf4 (Q071610), Rf5 (Q071609), Rf6 (Q071608), Rf7 (Q071607), Rf8 (Q071606). Amplification parameters were 10 min at 50 °C, 10 min at 95 °C and 50 cycles of 15 s at 95 °C and 20 s at 60 °C. RNA dilutions from purified WIV-1 stock were used as a standard.
and then serially twofold diluted in 96-well cell plates to 1:40. Each 100 μl serum dilution was mixed with 100 μl viral supernatant containing 100 TCID₅₀ of WIV1 and incubated at 37 °C for 1 h. The mixture was added in triplicate wells of 96-well cell plates with plated monolayers of Vero E6 cells and further incubated at 37 °C for 2 days. Serum from a healthy blood donor was used as a negative control in each experiment. CPE was observed using an inverted microscope 2 days after inoculation. The neutralizing antibody titre was read as the highest dilution of serum which completely suppressed CPE in infected wells. The neutralization test was repeated twice.

**Recombination analysis.** Full-length genomic sequences of SL-CoV Rs3367 or RsSHC014 were aligned with those of selected SARS-CoVs and bat SL-CoVs using Clustal X. The aligned sequences were preliminarily scanned for recombination events using Recombination Detection Program (RDP) 4.0 (ref. 19). The potential recombination events suggested by RDP owing to their strong P values (<10⁻⁵) were investigated further by similarity plot and bootscan analyses implemented in Simplot 3.5.1. Phylogenetic origin of the major and minor parental regions of Rs3367 or RsSHC014 were constructed from the concatenated sequences of the essential ORFs of the major and minor parental regions of selected SARS-CoV and SL-CoVs. Two genome regions between three estimated breakpoints (20,827–26,553 and 26,554–28,685) were aligned independently using ClustalX and generated two alignments of 5,727 base pairs and 2,133 base pairs. The two alignments were used to construct maximum likelihood trees to better infer the fragment parents. All nucleotide numberings in this study are based on Rs3367 genome position.
Extended Data Figure 1 | Sequence alignment of CoV S protein RBD.
SARS-CoV S protein (amino acids 310–520) is aligned with homologous regions of bat SL-CoVs using ClustalW. The newly discovered bat SL-CoVs are indicated with a bold vertical line on the left. The key amino acid residues involved in the interaction with human ACE2 are numbered on the top of the aligned sequences.
Extended Data Figure 2 | Alignment of CoV S protein S1 sequences.

Alignment of S1 sequences (amino acids 1–660) of the two novel bat SL-CoV S proteins with those of previously reported bat SL-CoVs and human and civet SARS-CoVs. The newly discovered bat SL-CoVs are boxed in red. SARS-CoV GZ02, BJ01 and Tor2 were isolated from patients in the early, middle and late phase, respectively, of the SARS outbreak in 2003. SARS-CoV SZ3 was identified from *P. larvata* in 2003 collected in Guangdong, China. SL-CoV Rp3, Rs 672 and HKU3-1 were identified from *R. sinicus* collected in Guangxi, Guizhou and Hong Kong, China, respectively. Rf1 and Rm1 were identified from *R. ferrumequinum* and *R. macrotis*, respectively, collected in Hubei Province, China. Bat SARS-related CoV BM48-31 was identified from *R. blasii* collected in Bulgaria.
Extended Data Figure 3 | Complete RdRP sequence phylogeny. Phylogenetic tree of bat SL-CoVs and SARS-CoVs on the basis of complete RdRP sequences (2,796 nucleotides). Bat SL-CoVs RsSHC014 and Rs3367 are highlighted by filled circles. Three established coronavirus genera, *Alphacoronavirus*, *Betacoronavirus* and *Gammacoronavirus* are marked as α, β and γ, respectively. Four CoV groups in the genus *Betacoronavirus* are indicated as A, B, C and D, respectively. MHV, murine hepatitis virus; PHEV, porcine haemagglutinating encephalomyelitis virus; PRCV, porcine respiratory coronavirus; FIPV, feline infectious peritonitis virus; IBV, infectious bronchitis coronavirus; BW, beluga whale coronavirus.
Extended Data Figure 4 | Sequence phylogeny of the complete S protein of SL-CoVs and SARS-CoV. Phylogenetic tree of bat SL-CoVs and SARS-CoVs on the basis of complete S protein sequences (1,256 amino acids).

Bat SL-CoVs RsSHC014 and Rs3367 are highlighted by filled circles. Bat CoV HKU9 was used as an outgroup.
Extended Data Figure 5 | Detection of potential recombination events.

a, b, Similarity plot (a) and bootscan analysis (b) detected three recombination breakpoints in the bat SL-CoV Rs3367 or SHC014 genome. The three breakpoints were located at the ORF1b (nt 20,827), M (nucleotides 26,553) and N (nucleotides 28,685) genes, respectively. Both analyses were performed with an F84 distance model, a window size of 1,500 base pairs and a step size of 300 base pairs.
## Extended Data Table 1 | Summary of sampling detail and CoV prevalence

| Sampling time  | Total number of swab or fecal samples collected | Number of CoV PCR positive samples (%) |
|----------------|-----------------------------------------------|----------------------------------------|
| April, 2011    | 14                                            | 1 (7.1)                                |
| October, 2011  | 10                                            | 3 (30)                                 |
| May, 2012      | 54                                            | 4 (7.4)                                 |
| September, 2012| 39                                            | 19 (48.7)                               |
## Extended Data Table 2 | Genomic sequence identities of bat SL-CoVs with SARS-CoVs

| CoVs     | Genome size (nt) | SHC014 | Rs672 | Rp3 | Rf1 | Rm1 | HKU3-1 | BM48-31 | GZ02 | BJ01 | Tor2 | SZ3 |
|----------|------------------|--------|-------|-----|-----|-----|--------|---------|------|------|------|-----|
| 3387     | 29,787           | 96.8   | 92.5  | 93.2| 87.3| 88.0| 87.8   | 76.9    | 95.4 | 95.3 | 95.4 | 95.3|
| SHC014   | 29,787           | -      | 92.6  | 93.2| 87.3| 88.1| 87.8   | 77.0    | 95.2 | 95.1 | 95.1 | 95.1|
| Rs672    | 29,059           | -      | 92.4  | 86.2| 87.4| 87.0| 75.2   |         | 90.9 | 90.9 | 90.8 | 91.0|
| Rp3      | 29,736           | -      | 88.3  | 90.3| 89.6| 77.0|        |         | 92.1 | 92.0 | 92.1 | 92.0|
| Rf1      | 29,709           | -      | 89.4  | 88.4| 76.6|     |        |         | 87.2 | 87.1 | 87.2 | 87.1|
| Rm1      | 29,749           | -      | 90.1  | 76.4|     |     |        |         | 87.6 | 87.5 | 87.5 | 87.5|
| HKU3-1   | 29,728           | -      | 76.8  |     |     |     |        |         | 87.4 | 87.3 | 87.4 | 87.3|
| BM48-31  | 29,276           | -      |       |     |     |     |        |         | 76.9 | 77.1 | 77.0 | 76.9|
| GZ02     | 29,760           |        |       |     |     |     |        |         | -    | 99.6 | 99.6 | 99.7|
| BJ01     | 29,725           | -      |       |     |     |     |        |         | -    | 99.8 | 99.6 |      |
| Tor2     | 29,751           |        |       |     |     |     |        |         | -    |      | 99.5 |      |
| SZ3      | 29,741           |        |       |     |     |     |        |         |      |      |      |      |
Extended Data Table 3 | Genomic annotation and comparison of bat SL-CoV Rs3367 with human/civet SARS-CoVs and other bat SL-CoVs

| ORFs | Start-End (nt) | No. of Nt | No. of As | TRS | G202 | BJ01 | Tor2 | S23 | RsS72 | Rp3 | Rfl | Rm1 | HKU3-1 | BM40-31 |
|------|----------------|-----------|-----------|-----|-----|-----|-----|-----|------|-----|-----|-----|---------|----------|
| P1a  | 265-13,398     | 13,134    | 4,377     |     |     |     |     |     |      |     |     |     |         |          |
| P1b  | 13,398-21,485  | 8,088     | 2,695     |     | 96.6/97.9 | 96.6/97.9 | 96.6/97.9 | 96.6/98.1 | 93.3/94.2 | 95.5/96.9 | 88.1/94.0 | 87.4/93.3 | 87.9/94.2 | 78.3/80.8 |
| S    | 21,482-25,262  | 3,771     | 1,256     |     | 88.3/92.2 | 88.3/92.2 | 88.3/92.2 | 88.3/92.2 | 97.2/99.2 | 97.2/99.2 | 90.6/98.4 | 91.3/98.7 | 90.7/98.5 | 83.4/93.7 |
| (S1)*| 21,493-23,535  | 2,043     | 681       |     | 78.2/81.1 | 78.2/81.1 | 78.2/81.1 | 78.2/81.1 | 85.1/82.2 | 63.9/63.0 | 62.5/62.5 | 64.7/63.3 | 65.2/63.4 | 62.2/64.7 |
| (S2)*| 23,536-25,263  | 1,728     | 575       |     | 98.7/99.3 | 98.7/99.3 | 98.7/99.3 | 98.7/99.3 | 87.9/94.8 | 88.1/94.8 | 85.1/82.7 | 87.9/94.5 | 88.9/93.5 | 78.6/88.2 |
| ORF3a| 25,271-30,955  | 825       | 274       |     | 99.2/98.1 | 99.2/98.1 | 99.2/98.1 | 99.2/98.1 | 90.4/90.8 | 84.1/84.3 | 88.8/86.8 | 83.5/86.4 | 83.1/86.4 | 72.1/71.2 |
| ORF3b| 25,692-28,036  | 345       | 114       |     | 99.1/99.1 | 99.1/99.1 | 99.1/99.1 | 99.1/99.1 | 99.1/99.1 | 99.1/99.1 | 99.1/99.1 | 99.1/99.1 | 99.1/99.1 | 78.5/88.1 |
| E    | 26,123-28,350  | 231       | 76        |     | 99.7/99.8 | 99.7/99.8 | 99.7/99.8 | 99.7/99.8 | 99.1/99.6 | 97.8/98.6 | 96.5/96.0 | 96.1/97.3 | 97.4/99.6 | 91.3/93.4 |
| M    | 28,491-27,966  | 666       | 221       |     | 97.4/98.1 | 97.4/98.1 | 97.4/98.1 | 97.4/98.1 | 98.7/99.5 | 93.3/98.1 | 96.3/98.8 | 93.2/98.5 | 93.9/98.5 | 78.5/88.1 |
| ORF6 | 27,077-27,286  | 192       | 63        |     | 99.7/99.5 | 99.7/99.5 | 99.7/99.5 | 99.7/99.5 | 97.3/99.8 | 95.8/96.2 | 94.2/96.2 | 95.3/96.2 | 94.7/96.4 | 63.5/49.2 |
| ORF7a| 27,276-27,644  | 369       | 122       |     | 94.5/95.9 | 94.5/95.9 | 94.5/95.9 | 94.5/95.9 | 97.8/100 | 96.2/96.1 | 92.6/96.0 | 93.4/96.5 | 93.2/96.0 | 93.2/96.0 |
| ORF7b| 27,641-27,776  | 135       | 44        |     | 98.2/98.1 | 98.2/98.1 | 98.2/98.1 | 98.2/98.1 | 98.2/100 | 99.2/100 | 97.7/97.7 | 99.2/100 | 93.3/96.4 | 92.8/96.8 |
| ORF8 | 27,762-28,147  | 356       | 121       |     | 47.1/48.3 | N/A       | N/A       | 47.1/48.3 | 97.8/100 | 85.2/90.2 | 46.2/90.0 | 85.7/90.2 | 85.7/90.3 | N/A     |
| N    | 28,162-29,430  | 1,269     | 422       |     | 98.3/99.5 | 98.4/99.5 | 98.4/99.5 | 98.4/99.5 | 99.9/99.5 | 99.6/97.6 | 93.7/95.2 | 96.2/97.1 | 90.9/99.2 | 77.9/87.2 |
| s2m  | 29,629-29,666  | 41        | 75        |     | 97.5 | 97.5 | 97.5 | 97.5 | 100 | 100 | 100 | 100 | 95.1 |

* (S1), the N-terminal domain of the coronavirus S protein responsible for receptor binding; (S2), the S protein C-terminal domain responsible for membrane fusion.

The ORFs in the genome were predicted and potential protein sequences were translated. The pairwise comparisons were conducted for all ORFs at nucleotide acids (nt) and amino acids (aa) levels. The s2m were compared at nt level. TRS: Transcription regulating sequences. N.D.: not done; N.A.: not available.
| ORFs | Start-End (nt) | No. of nt | No. of aa | TRS | GZ2 | BJ41 | Tyr2 | S2Z | Rs(12) | Rp3 | RH | Rm1 | HKU3-1 | BMAP-31 |
|------|--------------|-----------|----------|-----|-----|-----|------|-----|--------|-----|-----|-----|--------|--------|
| PSa  | 265-13,398   | 13,134    | 4,377    |     | 96.7/67.9 | 96.8/67.9 | 96.8/67.0 | 96.8/68.1 | 93.3/84.2 | 95.5/66.0 | 88.1/64.0 | 87.8/63.3 | 87.9/64.2 | 76.3/68.8 |
| PSb  | 13,398-21,495| 8,086     | 2,695    |     | 53.8/60.2 | 96.3/63.9 | 56.3/63.9 | 95.3/63.9 | 93.2/200.2 | 96.3/63.9 | 90.2/63.9 | 90.2/63.9 | 90.3/63.9 | 83.4/63.7 |
| S    | 21,492-25,262| 3,771     | 1,256    |     | 88.3/60.1 | 88.2/60.0 | 88.1/60.0 | 88.2/60.0 | 76.5/67.2 | 76.0/67.1 | 74.9/67.4 | 76.3/67.1 | 76.0/67.2 | 72.0/74.5 |
| (S1)* | 21,493-23,535| 2,043     | 681      |     | 78.5/61.7 | 78.2/60.9 | 78.1/60.6 | 78.4/61.1 | 65.1/62.2 | 63.6/63.0 | 62.2/62.5 | 64.7/63.3 | 66.2/63.4 | 62.6/64.7 |
| (S2)* | 23,536-25,263| 1,728     | 575      |     | 98.4/60.3 | 98.3/60.2 | 98.4/60.3 | 98.4/60.3 | 87.7/84.8 | 88.1/85.8 | 85.1/84.7 | 87.7/84.5 | 86.2/85.5 | 78.6/84.8 |
| ORF3a | 25,271-20,505| 4,245     | 525      |     | 98.2/60.1 | 98.3/60.0 | 98.7/60.0 | 95.3/60.0 | 92.5/69.0 | 84.5/64.3 | 85.5/66.0 | 83.7/64.3 | 53.1/64.7 | 72.0/71.2 |
| ORF3b | 29,692-26,038| 645       | 114      |     | 99.6/60.1 | 98.4/60.2 | 98.5/60.2 | 97.9/60.7 | 96.1/60.2 | 92.6/62.1 | N/D | 82.6/62.1 | N/D | N/D | N/D |
| E    | 26,123-26,300| 1,776     | 76       |     | 98.7/60.6 | 98.7/60.6 | 98.7/60.6 | 98.7/60.6 | 96.1/60.1 | 97.8/60.3 | 96.5/60.3 | 96.1/60.3 | 96.5/60.3 | 91.3/65.3 |
| M    | 26,401-27,068| 666       | 221      |     | 97.4/60.1 | 97.2/60.1 | 97.5/60.1 | 97.2/60.1 | 98.7/69.5 | 93.3/60.1 | 98.3/69.6 | 93.2/69.5 | 93.2/69.5 | 78.5/68.1 |
| ORF6  | 27,077-27,268| 192       | 63       |     | 97.3/60.2 | 95.8/60.2 | 97.5/60.2 | 97.9/60.2 | 97.3/60.2 | 95.8/60.2 | 95.2/60.2 | 94.2/60.2 | 95.4/60.0 | 63.5/64.2 |
| ORF7a | 27,276-27,544| 296       | 122      |     | 94.5/60.5 | 94.5/60.5 | 94.5/60.5 | 94.5/60.5 | 97.8/60.1 | 96.2/60.1 | 92.9/60.0 | 93.4/60.0 | 93.2/60.0 | 62.3/68.1 |
| ORF7b | 27,641-27,776| 135       | 44       |     | 96.2/60.1 | 95.2/60.1 | 96.2/60.1 | 95.2/60.1 | 99.2/60.1 | 99.2/60.1 | 97.7/60.1 | 97.2/60.1 | 93.3/60.4 | 62.4/63.6 |
| ORF8  | 27,783-28,147| 366       | 121      |     | 97.1/60.5 | N/A | N/A | 97.1/60.5 | 97.1/60.5 | 97.1/60.5 | 97.1/60.5 | 97.1/60.5 | 97.1/60.5 | 97.1/60.5 |
| N    | 28,162-29,430| 1,269     | 422      |     | 98.3/60.5 | 98.4/60.5 | 98.4/60.5 | 98.4/60.5 | 98.3/60.5 | 96.5/60.7 | 93.7/60.2 | 96.2/60.7 | 95.9/60.2 | 77.9/67.2 |
| s2m  | 29,023-29,608| 41        | 33       |     | 97.5 | 97.5 | 97.5 | 97.5 | 100 | 100 | 100 | 100 | 100 | 95.1 |

* S1, the N-terminal domain of the coronavirus S protein responsible for receptor binding; S2, the S protein C-terminal domain responsible for membrane fusion.

The ORFs in the genome were predicted and potential protein sequences were translated. The pairwise comparisons were conducted for all ORFs at nucleotide (nt) and amino acids (aa) levels. The s2m were compared at nt level. TRS, Transcription regulating sequences. N/D, not done; N/A, not available.
Extended Data Table 5 | Cell lines used for virus isolation and susceptibility tests

| Cell lines | Species (organ) origin          | Medium          | Infectivity |
|------------|---------------------------------|-----------------|-------------|
| 293T       | Human (kidney)                  |                 | -           |
| Hela       | Human (cervix)                  |                 | -           |
| VeroE6     | Monkey (kidney)                 |                 | +           |
| PK15       | Pig (kidney)                    | DMEM+10%FBS     | +           |
| BHK21      | Hamster (kidney)                |                 | -           |
| A549       | Human (alveolar basal epithelial)|                 | +           |
| BK         | Myotis davidi (kidney)          | RPMI1640+10%FBS | -           |
| RSKT       | Rhinolophus sinicus (kidney)    |                 | +           |
| MCKT       | Myotis chinensis (kidney)       | DMEM/F12+10%FBS | -           |
| PaKi       | Pteropus alecto (kidney)        |                 | -           |
| RLK        | Rousettus leschenaulti (kidney) |                 | -           |

* Infectivity was determined by the presence of viral antigen detected by immunofluorescence assay.