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

Supporting Online Material for

Bats Are Natural Reservoirs of SARS-like Coronaviruses

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

*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.)

Published 29 September 2005 on *Science* Express
DOI: 10.1126/science.1118391

This PDF file includes:

Supporting online materials

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*],and Lin-Fa Wang[3*]

Materials and Methods

Sampling. Bats were trapped from their natural habitat in China at four different locations up to 1,800 km apart. Throat and fecal swab samples were collected in viral transport medium composed of Hank's balanced salt solution at pH7.4 containing BSA (1%), amphotericin (15 μg/ml), penicillin G (100 units/ml), and streptomycin (50 μg/ml). Samples were aliquoted and kept at -70^{0}C until use. Blood samples were collected from the brachial artery or vein or, in some cases, heart, and mixed with anti-coagulation agent. Serum was separated by centrifugation at 3000 x g for 15 min within 24 h and preserved at 4^{0}C. One aliquot was analysed at the Wuhan Institute of Virology (WIV), and another sent to the Australian Animal Health Laboratory (AAHL) for parallel analysis. Most animals survived the sampling process and were released back into their habitat. Those which did not survive were anatomised.

ELISA tests. Two different ELISA tests were employed for detection of SARS-CoV antibodies in bat sera. At the WIV in China, a sandwich ELISA kit based on recombinant

N protein (Beijing Wantai Biological Pharmacy Enterprise Co., Ltd, China) was used. The ELISA was performed according to the manufacture's instructions. At AAHL in Australia, an indirect ELISA was employed for detection of SARS-CoV antibodies using inactivated total viral antigen prepared from SARS-CoV strain HKU-39849. Bound bat antibodies were detected using a 1:1 mixture of horseradish peroxidase (HRP)-conjugated protein-A and HRP-protein-G (Zymed Laboratories, USA).

Serum neutralization test. Due to cytotoxicity of bat sera previously observed in our group, a modified neutralization test was developed. Vero E6 cell monolayer was prepared in 8 well chamber slides seeded with 50,000 cells/well and incubated overnight at 37ºC in a humidified 5% CO_2 incubator. The cells were grown in 300 µl Minimal Essential Medium containing Earle's salts and supplemented with 2 mM glutamine, 500 µg/ml fungizone, 100 units/ml penicillin, 100 µg/ml streptomycin and 10% foetal calf serum (FCS). Under BSL 4 conditions, SARS-CoV stock (HKU-39849) at 5 x 10^6 $TCID_{50}$/ml was diluted to a concentration of 100 $TCID_{50}$/25 µl and 25 µl aliquots were added to wells in a 96-well incubation plate. Test sera were diluted 1:10 in cell medium, 25 µl added to test wells or 25 µl medium to control wells and incubated for 30 min at 37ºC. The medium was discarded from the cell monolayers and 50 µl of the pre-incubated virus/sera mix was added to appropriate wells in the chamber slide and incubated for 30 min at 37ºC. The inoculum was discarded and the cells washed 3 times with 200 µl cold PBSA and 200 µl cell medium added to each well. The cells were incubated at 37ºC for 16 h. After removing the medium, cells were fixed in 100% ice-cold methanol for 15 min and air-dried. Reduction or inhibition of infection foci

was determined by immunofluorescence as described previously (*S1*).

RNA extraction and PCR analysis. RNA was extracted from 140 µl of swab samples with QIA amp Viral RNA Mini Kit (Qiagen, Germany) following the manufacture's instructions. RNA was eluted in 60 µl AVE buffer and stored at -70℃. Reverse transcription was carried out in a 25-µl reaction containing 2 µl RNA, 1 µg/ml random primer $(dNp)_6$, 5 µl reaction buffer, 0.6 mM dNTP, 20 units RNase inhibitor and 200 units of M-MLV reverse transcriptase (Promega, USA). Aliquots of cDNA were then analysed by two different groups as follows. At WIV, RT-PCR was used to detect the presence of SARS-CoV sequences using three pairs of primers kindly designed by Prof. Yi Guan (The University of Hong Kong). The sequences of the primers are given in Table S1. The 25 µl reaction mix contained 2.5 µl PCR Buffer, 20 pmol of each primer, 0.2 mM dNTP and 0.5 units Taq DNA polymerase (BioStar, Canada). The amplification of N-gene fragment was performed as follows: 94℃ for 5 min followed by 35 cycles consisting of 94℃ for 30 sec, 53℃ for 30 sec, 72℃ for 1 min, and a final extension of 72℃ for 10 min. The amplification of S and P fragments was conducted as follows: 94℃ for 5 min followed by 35 cycles consisting of 94℃ for 30 sec, 54℃ for 30 sec, 72℃ for 40 sec, and a final extension of 72℃ for 10 min. At AAHL, a real-time TaqMan PCR assay was used with a pair of primers targeting the N gene region (forward primer: 5'-CCC AGA TGG TAC TTC TAT TAC CTA GGA; reverse primer: 5'-CCA TAC GAT GCC

TTC TTT GTT AG) and a FAM-labeled probe (5'-AAG CTT CAC TTC CCT ACG G). The assays were conducted on the ABI PRISM® 7700 Sequence Detection System (Applied Biosystems, USA) for 45 cycles following standard conditions recommended with the TaqMan Universal PCR Master Mix (ABI). The 25 µl reaction mixture contained 12.5 µl TaqMan PCR mix, 2 µl random cDNA, 1.25 µl of each primer (20 µM), 1.25 ul probe (5 µM) and 6.75 µl dH_2O.

Virus isolation. Vero E6 cell Monolayer was maintained in DMEM medium supplemented with 10% FCS. PCR-positive fecal samples (in 200 µl buffer) were filtered and diluted 1:10 in the same medium before being added to Vero E6 cells. After incubation at 37 ^oC for 1 h, the inoculum was removed and replaced with fresh DMEM medium with 2% FCS. The cells were incubated at 37 ^oC and checked daily for CPE. Two blind passages were carried out for each sample. After each passage, both the culture supernatant and cell pellet were examined for presence of virus by RP-PCR using the SARS-CoV primers listed in Table S1. Penicillin and streptomycin, at 100 IU per ml each, were included in all tissue culture media.

Characterization of genome sequence. The randomly primed cDNA prepared above was used as template for PCR amplification to analyse the genome sequence. Due to the genetic variation between SARS-CoV and SL-CoVs, primers designed based on SARS-CoV sequences did not always work. Hence, a combination of SARS-CoV primers, degenerate coronavirus primers, and SL-CoV-specific primers (designed using partial sequences obtained from small PCR fragments) were used to complete the

full-length genome sequencing of SL-CoV Rp3 (primer sequences will be provided upon request). These primers were then used to obtain partial sequences for the other four bat samples (Rf1, Rm1, Rp1 and Rp2). PCR products with expected size were gel-purified with E.Z.N.A Gel Extraction Kit (Omega Bio-Tek, USA) or QIAquick Gel Extraction Kit (Qiagen), and sequenced directly. For some fragments, they were first cloned into pGEM-T Easy Vector (Promega) before being sequenced. For each PCR fragment cloned, three independent clones were subjected to sequencing to obtain consensus sequence.

Databank accession numbers. All sequences obtained in this study have been deposited in GenBank and their accession numbers are as follows (given in parenthesis): full-length genome sequence of SL-CoV Rp3 (DQ71615); partial sequences of SL-CoV Rf1 covering the ORF10'-N region (DQ71611) and the S1 coding region (DQ159956), respectively; partial sequences of SL-CoV Rm1 covering the ORF10'-N region (DQ71612) and the S1 coding region (DQ159957), respectively; sequence of the SL-CoV Rp1 N gene (DQ71613); and sequence of the SL-CoV Rp2 N gene (DQ71614). SARS-CoV sequences used in this study: human SARS-CoV strains Urbani (AY278741), Tor2 (AY274119), and GD01 (AY278489), and civet SARS-CoV strain SZ3 (AY304486). Other coronavirus sequences used in this study: HCoV-229E (AF304460), HCoV-OC43 (AY391777), HCoV-NL63 (AY567487), PEDV (AF353511), MHV (AY700211), IBV (AY851295), and bat-CoV P1b (AY864196) and S2 (AY864197) regions. Nipah virus sequences used in this study: human isolates UMCC1 (AY029767), Malaysia (AF212302), and Bangladesh (AY988601); bat isolates *Pteropus hypomelanus* or Ph

(AF376747) and *Pteropus lylei* or Pl (AY858110); pig isolates VRI2794 (AJ564621), VRI1413 (AJ564622), VRI0626 (AJ627196).

Sequence analysis. Routine sequence management and analysis were carried out using DNAStar, GeneDoc and Clone Manager 7, respectively. Phylogenetic trees were constructed using the neighbour-joining algorithm with bootstrap values determined by 1,050 replicates in the MEGA3 software package (*S2*).

Table S1. Primers used in RT-PCR analysis of bat swab samples. The location of the primers refers to the numbering used by Rota *et al*. (*S3*) for SARS-CoV strain Urbani.

Gene	Primer name and location	Primer sequence
P gene	Co13-80F (18118-18138) Co13-368R(18401-18382)	CCTACCGTAGACTCATCTCTA GTCACCTGGTGGAGGTTTTG
N gene	Co19-1418F(28433-28452) RV19-296 (28810-28791)	TCAGCCCCAGATGGTACTTC GCTGGTTCAATCTGTCTAGC
S gene	Co15-904F (21918-21937) Co15-1213R(22227-22208)	CCCATGGGTACACAGACACA CCCAAA TGTCTTGAGCAGGT

Table S2. Comparison of predicted protein size and sequence homology between SARS-CoV and SL-CoV Rp3. Data for SARS-CoV were derived from the human isolate Tor2 with the exception of ORF10', which was derived from the civet isolate SZ3.The ORF nomenclature used in the table follows that of Marra *et al*. (*S4*). The equivalent ORF nomenclature used by Rota *et al*. (*S3*) is included in the square brackets. (na: not available; np: not present).

Gene/ORF	Gene product size (aa)		Amino Acid Sequence identity
	SARS-CoV	SL-CoV	
P1a	4,382	4,380	96%
P1b	2,628	2,628	99%
S	1,255	1,241	78%
(S1)*	680	666	64%
(S2)*	575	575	96%
ORF3 [X1]	274	274	83%
ORF4 [X2]	154	np	na
E	76	76	100%
M	221	221	97%
ORF7 [X3]	63	63	92%
ORF8 [X4]	122	122	96%
ORF9	44	44	93%
ORF10	39	np	na
ORF11 [X5]	84	np	na
ORF10'†	122	121	35%
N	422	421	97%
OF13	98	97	85%
ORF14	70	70	91%

* 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.
†ORF 10' is not present in isolate Tor2, the sequence from the civet isolate SZ3 (*S5*) is used instead.

Table S3. Comparison of putative transcription regulatory sequences (TRS) between SARS-CoV and SL-CoV Rp3. The TRS data for the human isolate Tor2 was obtained from Marra *et al*. (*S4*). Nucleotides which differ in the two viruses are shaded. (nt: nucleotides).

ORF		Transcription-regulating Sequences (TRS)
Leader	(SCV)	UCUCUAAACGAACUUUAAAAUCUGUG
	(SLCV)	UCUCUAAACGAACUUUAAAAUCUGUG
S	(SCV)	CAACUAAACGAAC**AUG**
	(SLCV)	CAACUAAACGAAC**AUG**
ORF3	(SCV)	CACAUAAACGAACUU**AUG**
	(SLCV)	CACAUAAACGAACUUA**AUG**
E	(SCV)	UGAGUACGAACUU**AUG**
	(SLCV)	UGAGUACGAACUU**AUG**
M	(SCV)	GGUCUAAACGAACU (43 nt)**AUG**
	(SLCV)	GGUCUAAACGAACU (43 nt)**AUG**
ORF7	(SCV)	AACUAUAAA(64 nt)**AUG**
	(SLCV)	AACUAUAAA(64 nt)**AUG**
ORF8	(SCV)	CCAUAAAACGAAC**AUG**
	(SLCV)	CCAUAAAACGAAC**AUG**
ORF9	(SCV)	CUCUA--GUAUUUUUAA (31 nt)**AUG**
	(SLCV)	CUCUA--GUAUUUAUAA (31 nt)**AUG**
ORF10	(SCV)	AGUCUAAACGAAC**AUG**
	(SLCV)	AGUCUAAACGAAC**AUG**
N	(SCV)	UAAACGAACAAAUUAAA**AUG**
	(SLCV)	UAAACGAACAAACUAAA**AUG**

```
Tor2   1 MSDNGPQSNQRSAPRITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA
GD01   1 MSDNGPQSNQRSAPRITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA
SZ3    1 MSDNGPQSNQRSAPRITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA
Rp3    1 MSDNGPQ-NQRSAPRITFGGPTDSTDNNQDGGRSGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNSGKDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA
Rp2    1 MSDNGPQ-NQRSAPRITFGGPTDSTDNNQDGGRSGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNSGKDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA
Rp1    1 MSDNGPQ-NQRSAPRITFGGPTDSTDNNQDGGRSGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNSGKDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA
Rm1    1 MSDNGPQ-NQRSAPRITFGGPSDSTDNNQDGGRSGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNSGKDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA
Rf1    1 MSDNGPQ-NQCSAPRITFGGPSDSTDNNQDGGRSGARPKQRRPQGLPNNTASWFTALTQHGKEGLKFPQGQGVPINTNSGRDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA
             *

Tor2 121 SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLPKGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESKVSGKGQ
GD01 121 SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLPKGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESKVSGKGQ
SZ3  121 SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLPKGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESKVSGKGQ
Rp3  120 SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAAIVLQLPQGTTLPKGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESKVSGRSQ
Rp2  120 SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAAIVLQLPQGTTLPKGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESKVSGRSQ
Rp1  120 SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAAIVLQLPQGTTLPKGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESKVSGRSQ
Rm1  120 SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAAIVLQLPQGTTLPKGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGSGETALALLLLDRLNQLESKVSGKGQ
Rf1  120 SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAAIVLQLPQGTTLPKGFYAEGSRNGSQASSRSSSRSRGNSRTSTPGSSRGNSPARVASGGGETALALLLLDRLNQLESKVSGKGQ

Tor2 241 QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA
GD01 241 QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA
SZ3  241 QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA
Rp3  240 QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA
Rp2  240 QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA
Rp1  240 QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELVRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA
Rm1  240 QQQGQTVTKKSAAEASKKPRQKRTATKSYNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYKYWPQIAQFAPSASAFFGMSRIGMEVTPLGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA
Rf1  240 QQQGQTVTKKSTSEASKKPRQKRTATKQYNVTQAFGRRGPDQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRISMEVTPSGTWLIYHGAIKLDDKDPQFKDNVILLNKHIDA

Tor2 361 YKTFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA
GD01 361 YKTFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA
SZ3  361 YKTFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA
Rp3  360 YKIFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA
Rp2  360 YKIFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA
Rp1  360 YKTFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA
Rm1  360 YKTFPPTEPKKDKKKKTDEAQPLPQR-KKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA
Rf1  360 YKTFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA
                                    *
```

Figure S1. Alignment of amino acid sequences of N proteins from representative members of the SARS cluster of coronaviruses. Identical residues are shaded while dashes indicated gaps introduced to maximize homology regions. The locations of single amino acid deletion were indicated by * below the sequence. For abbreviations and sources of the sequences used see ***Databank accession numbers.***

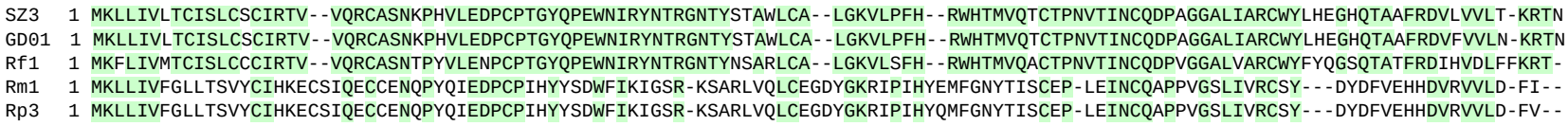

Figure S2. Alignment of amino acid sequences of putative proteins encoded by ORF10' in selected members of the SARS cluster of coronaviruses. Identical residues are shaded while dashes indicated gaps introduced to maximize homology regions. For abbreviations and sources of the sequences used see ***Databank accession numbers.***

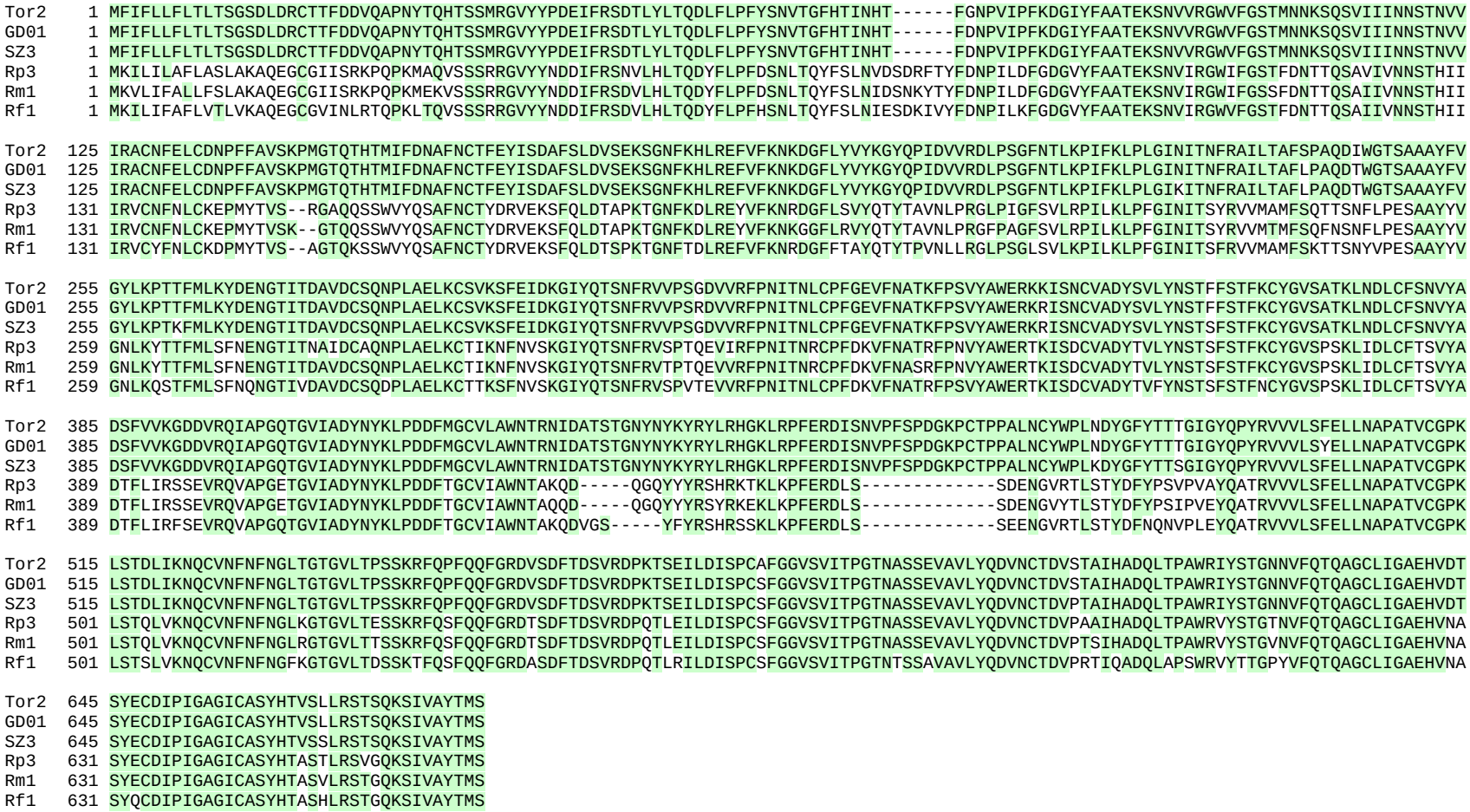

Figure S3. Alignment of amino acid sequences of S1 protein regions of selected members of the SARS cluster of coronaviruses. Identical residues are shaded while dashes indicated gaps introduced to maximize homology regions. For abbreviations and sources of the sequences used see ***Databank accession numbers.***

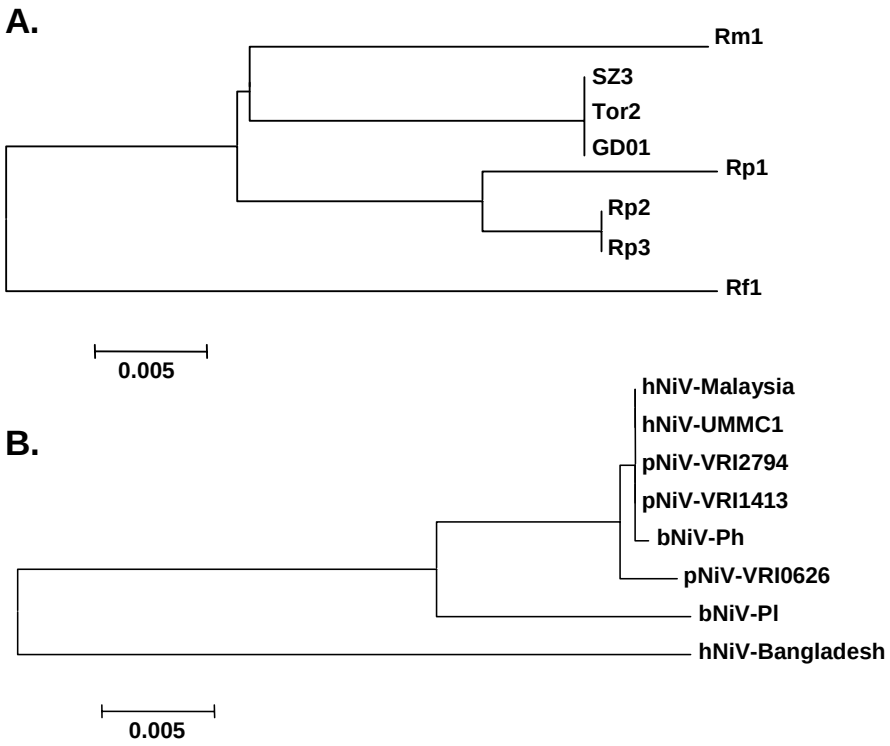

Figure S4. Comparison of genetic diversity between two groups of emerging zoonotic viruses of bat origin. (**A**) Trees based on N gene sequences of selected members of the SARS cluster of coronaviruses. (**B**) Trees based on N gene sequences of selected henipaviruses. Abbreviations: Tor2, SZ3 and GD01, different SARS-CoV isolates; Rf1, Rm1 and Rp1-3, different SL-CoV sequences; NiV, Nipah virus; h, b and p, human, bat and pig isolates. The genetic distance scale in **A** is identical to that in **B**.

References

S1. C. Tu *et al.*, *Emerg. Infect. Dis.* **10**, 2244 (2004)

S2. S. Kumar, K. Tamura, M. Nei, *Brief. Bioinformatics* 5:150 (2004)

S3. P. A. Rota *et al.*, *Science* **300**, 1394 (2003)

S4. M. A. Marra *et al.*, *Science* **300**, 1399 (2003)

S5. Y. Guan *et al. Science* **320**, 276 (2003)