Evolutionary arms race between virus and host drives genetic diversity in bat SARS related coronavirus spike genes

Hua Guo[a,b,1], Bing-Jie Hu[a,1], Xing-Lou Yang[a], Lei-Ping Zeng[a], Bei Li[a], Song-Ying Ouyang[c], Zheng-Li Shi [a,#]

[a]CAS Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China

[b]University of Chinese Academy of Sciences, Beijing, China

[c]The Key Laboratory of Innate Immune Biology of Fujian Province, Provincial University Key Laboratory of Cellular Stress Response and Metabolic Regulation, Biomedical Research Center of South China, Key Laboratory of OptoElectronic Science and Technology for Medicine of the Ministry of Education, College of Life Sciences, Fujian Normal University, Fuzhou, China

[#]Correspondence: Zheng-Li Shi: zlshi@wh.iov.cn.

Keywords: SARS related coronavirus, spike gene, receptor, ACE2, Chinese horseshoe bat, genetic diversity.

Running title: Arms race between SARSr-CoVs spike gene and host receptor

Abstract

The Chinese horseshoe bat (*Rhinolophus sinicus*), reservoir host of severe acute respiratory syndrome coronavirus (SARS-CoV), carries many bat SARS-related CoVs (SARSr-CoVs) with high genetic diversity, particularly in the spike gene. Despite these variations, some bat SARSr-CoVs can utilize the orthologs of human SARS-CoV receptor, angiotensin-converting enzyme 2 (ACE2), for entry. It is speculated that the interaction between bat ACE2 and SARSr-CoV spike proteins drives diversity. Here, we have identified a series of *R. sinicus* ACE2 variants with some polymorphic sites involved in the interaction with the SARS-CoV spike protein. Pseudoviruses or SARSr-CoVs carrying different spike proteins showed different infection efficiency in cells transiently expressing bat ACE2 variants. Consistent results were observed by binding affinity assays between SARS- and SARSr-CoV spike proteins and receptor molecules from bats and humans. All tested bat SARSr-CoV spike proteins had a higher binding affinity to human ACE2 than to bat ACE2, although they showed a 10-fold lower binding affinity to human ACE2 compared with their SARS-CoV counterpart. Structure modeling revealed that the difference in binding affinity between spike and ACE2 might be caused by the alteration of some key residues in the interface of these two molecules. Molecular evolution analysis indicates that these residues were under strong positive selection. These results suggest that the SARSr-CoV spike protein and *R. sinicus* ACE2 may have coevolved over time and experienced selection pressure from each other,

44 triggering the evolutionary arms race dynamics. It further proves that *R. sinicus* is the

45 natural host of SARSr-CoVs.

46

Importance

Evolutionary arms race dynamics shape the diversity of viruses and their receptors. Identification of key residues which are involved in interspecies transmission is important to predict potential pathogen spillover from wildlife to humans. Previously, we have identified genetically diverse SARSr-CoV in Chinese horseshoe bats. Here, we show the highly polymorphic ACE2 in Chinese horseshoe bat populations. These ACE2 variants support SARS- and SARSr-CoV infection but with different binding affinity to different spike proteins. The higher binding affinity of SARSr-CoV spike to human ACE2 suggests that these viruses have the capacity of spillover to humans. The positive selection of residues at the interface between ACE2 and SARSr-CoV spike protein suggests a long-term and ongoing coevolutionary dynamics between them. Continued surveillance of this group of viruses in bats is necessary for the prevention of the next SARS-like disease.

Introduction

Coronaviruses belong to the *Orthocoronavirinae* subfamily and *Coronaviridae* family. They are enveloped viruses and contain a positive-sense and single-stranded RNA genome. There are four genera in this subfamily, *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus,* and *Deltacoronavirus*. *Alphacoronavirus* and *Betacoronavirus* are believed to have originated in bats or rodents while *Gammacoronavirus* and *Deltacoronavirus* in birds (1, 2). Since the beginning of the 21st century, three betacoronaviruses have caused outbreaks of severe pneumonia in humans. These are, the severe acute respiratory syndrome coronavirus (SARS-CoV) (3-5), the Middle-East respiratory syndrome coronavirus (MERS-CoV) (6) and the ongoing 2019 novel coronavirus (SARS-CoV-2) (7).

The SARS-CoV-2 outbreak has brought back memories of SARS-CoV that occurred 17 years ago in China. SARS is a zoonosis, as demonstrated by the identification of an almost identical coronavirus present in market civets in Guangdong province, China (8). In the following years, genetically diverse SARS-related coronaviruses (SARSr-CoV) were detected or isolated from horseshoe bats from different regions of China and Europe (9-18). Bat SARSr-CoVs share 78–96% nucleotide sequence identity with human and civet SARS-CoVs, with the most variable regions encoding the spike protein (S) and accessory protein ORF3 and 8 (17, 19). Moreover, we have identified all the building blocks of SARS-CoV in the genome of different bat SARSr-CoVs and suggest that the ancestor of SARS-CoV originated in bats through the recombination of bat SARSr-CoV genomes (17, 19).

The first and essential step of virus infection is cell receptor recognition. The entry of the coronavirus is mediated by specific interactions between the viral S protein and cell surface receptor, followed by fusion between the viral and host membrane. The coronavirus S protein is functionally divided into two subunits, a cell attachment subunit (S1) and a membrane-fusion subunit (S2). The S1 region contains an N-terminal domain (NTD) and a C-terminal domain (CTD); both can be used for coronavirus receptor binding (RBD) (20). For SARS-CoV, its S1-CTD serves as an RBD for binding to the cellular receptor, angiotensin-converting enzyme 2 (ACE2) (21). Biochemical and crystal structure analyses have identified a few key residues in the interface between the SARS-CoV S-RBD and human ACE2 (21-23).

SARSr-CoVs, detected in *Rhinolophus sinicus*, can be divided into two distinct clades based on the size of the S protein. Clade 1 includes viruses that have S proteins identical in size to the SARS-CoV S protein, whereas viruses belonging to clade 2 have a smaller S protein, due to 5, 12, or 13 amino acid deletions (17, 19). Despite the variations in the RBD, all clade 1 strains can use ACE2 for cell entry, whereas clade 2 strains, with deletions cannot (14, 16, 17). These results suggest that members of clade 1 are likely to be the direct source of SARS-CoV in terms of genome similarity and ACE2 usage.

ACE2 is functionally divided into two domains—the N-terminal domain is involved in SARS-CoV binding and the C-terminal domain in the regulation of heart function (21, 24). Previous results have indicated that the C-terminal domains of ACE2 from different origins are relatively well conserved, whereas the N-terminal

105 domains show much more diversity across species (25). Previously, we have shown
106 that SARS-CoV can utilize ACE2 derived from *Myotis daubentonii* and *R. sinicus*;
107 minor mutations in the RBD-binding site could convert ACE2 from being
108 unsusceptible to susceptible to SARS-CoV binding (26, 27). As all bat SARSr-CoV
109 belonging to clade 1 can utilize ACE2 and have been isolated from *R. sinicus*, we
110 asked whether variations in ACE2 of *R. sinicus* could contribute to the diversity of bat
111 SARSr-CoV.

112 Here, we have investigated the polymorphism of *R. sinicus* ACE2 genes and
113 assessed their susceptibility and binding affinity to different bat SARSr-CoV spike
114 proteins through a combination of molecular evolution analyses, protein affinity
115 assays, and virus infection assays. Our results showed that the diversity of the
116 SARSr-CoV spike protein may experience natural selection pressure from the *R.*
117 *sinicus* ACE2 variants; over long periods of co-existence, the SARSr-CoV spike
118 protein may be selected by *R. sinicus* ACE2 to maintain genetic diversity and to fit
119 with the population of *R. sinicus*.

120

121 # Results

122 **ACE2 genes show high polymorphism among the *R. sinicus* populations**

123 Samples from three provinces (Hubei, Guangdong, and Yunnan) were used for
124 ACE2 amplification, based on the prevalence of bat SARSr-CoVs and tissue sample
125 availability and quality. In addition to previously sequenced bat ACE2 by our group
126 (sample ID 832, 411, and 3357, collected from Hubei, Guangxi, and Yunnan,

127 respectively) and others (GenBank accession no. ACT66275; sample collected from
128 Hong Kong), we obtained ACE2 gene sequences from 21 *R. sinicus* bat individuals:
129 five from Hubei, nine from Guangdong, and seven from Yunnan. The ACE2
130 sequences exhibited 98–100% amino acid (aa) identity within their species and 80-81%
131 aa identity with human ACE2 (**Table S1**). Major variations were observed at the
132 N-terminal region, including in some residues which were previously identified to be
133 in contact with SARS-CoV S-RBD (**Fig. 1A and Fig. S1**). Analysis based on
134 nonsynonymous SNPs helped identify eight residues, including 24, 27, 31, 34, 35, 38,
135 41, and 42. The combination of these 8 residues produced eight alleles, including
136 RIESEDYK, LIEFENYQ, RTESENYQ, RIKSEDYQ, QIKSEDYQ, RMTSEDYQ,
137 EMKTKDHQ, and EIKTKDHQ, named allele 1–8, respectively (**Fig. 1A**). In
138 addition to the ACE2 genotype data from previous studies (allele 4, 7, and 8), five
139 novel alleles were identified in the *R. sinicus* populations in this study. Alleles 2 and 4
140 were found in two and three provinces, respectively, whereas the other alleles seemed
141 to be geographically restricted. In summary, three alleles (4, 6, and 8) were found in
142 Guangdong, four (1, 2, 4, and 7) in Yunnan, three (2, 4, and 5) in Hubei, and one each
143 in Guangxi and Hong Kong. Coexistence of four alleles was found in the same bat
144 cave of Yunnan where the direct progenitor of SARS-CoV was found (**Fig. 1B**). Taken
145 together, these data suggest that ACE2 variants have been circulating within the *R.*
146 *sinicus* populations in different regions for a long time; substitutions at the sites
147 directly in contact with SARS-CoV S-RBD suggest that they may have important
148 functions during SARS-CoV evolution and transmission.

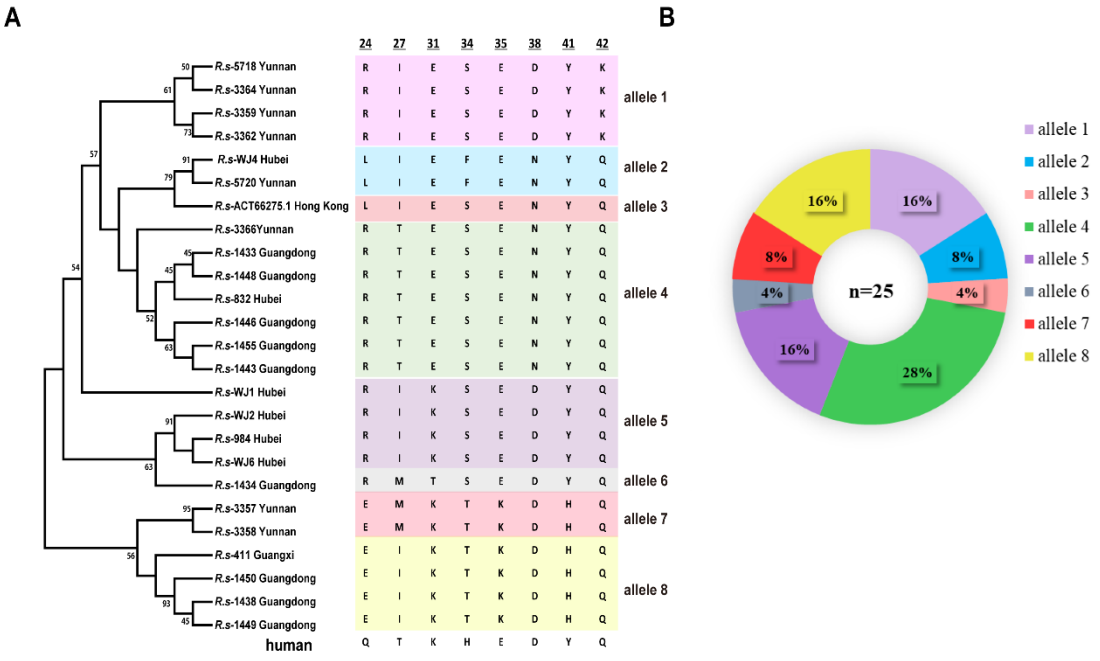

Fig. 1. Phylogenetic tree of *R. sinicus* ACE2.

(A) The maximum-likelihood tree (left panel) was produced using MEGA6 software, based on the alignment of ACE2 amino acid sequences of *R. sinicus* with the Jones-Taylor-Thornton (JTT+G+I) model and bootstrap value (%) of 1000 replicates (28). The eight key residues which are involved in interacting with the SARS-CoV spike are indicated in the panel on the right. The eight residues of human ACE2 are indicated at the bottom. The numbers at the top are amino acid positions in ACE2. (B) Frequencies of ACE2 alleles among the *R. sinicus* population. The number of *R. sinicus* ACE2 sequences is shown in the center of the pie chart. Different colored sectors with percentages in the pie chart indicate allele frequencies of ACE2 in the *R. sinicus* population used in this study. The alignments of complete amino acid sequences of *R. sinicus* ACE2 are shown in supplementary **Fig. S1**. ACE2 sequences of *Rs*-411, 832, and 3357 and *Rs*-ACT66275.1 have been published previously and were downloaded from GenBank. The accession number of *R. sinicus* ACE2

sequences obtained in this study are listed in Table S2

R. sinicus ACE2 variants show different susceptibility to SARSr-CoV infection

To assess if different *R. sinicus* ACE2 molecules affect the entry of SARS-CoV and bat SARSr-CoV, *R. sinicus* ACE2 variants were transiently expressed in HeLa cells and the entry efficiency of SARS-CoV pseudotyped or bat SARSr-CoVs carrying different S proteins were tested. The four tested bat SARSr-CoV strains can be divided into four genotypes according to their S1 sequences, as reported previously (17). Briefly, compared with SARS-CoV S protein, SARSr-CoV-RsWIV1 shares high aa identity at the RBD; RsWIV16 is the closest relative of SARS-CoV and shows high aa identity at both the NTD and RBD; Rs4231 shares high aa identity at the NTD; and RsSHC014 shows divergence at both the NTD and RBD regions.

Similar to our previous report, all four bat SARSr-CoV strains with the same genomic background but different S proteins could use human ACE2 and replicate at similar levels (17). However, there are some differences in how they utilize *R. sinicus* ACE2s (**Fig. 2 and Fig. S2**). All test viruses could efficiently use allele 1, 2, 4, 5 for entry. RsWIV1 and RsWIV16, which share an identical RBD, could not use allele 6 (sample ID 1434) from Guangdong. Rs4231 and RsSHC014, which share an identical RBD, could not use allele 7 (sample ID 3357) and 8 (sample ID 1438) from Yunnan and Guangdong, respectively. SARS-CoV-BJ01, which shares high similarity with WIV1 and WIV16 RBD, was able to use same bat ACE2 alleles as Rs4231 and

RsSHC014 in the pseudotyped infection assay (**Fig. 2 and Fig. S3**). These results indicate that cell entry was affected by both spike RBD and *R. sinicus* ACE2 variants.

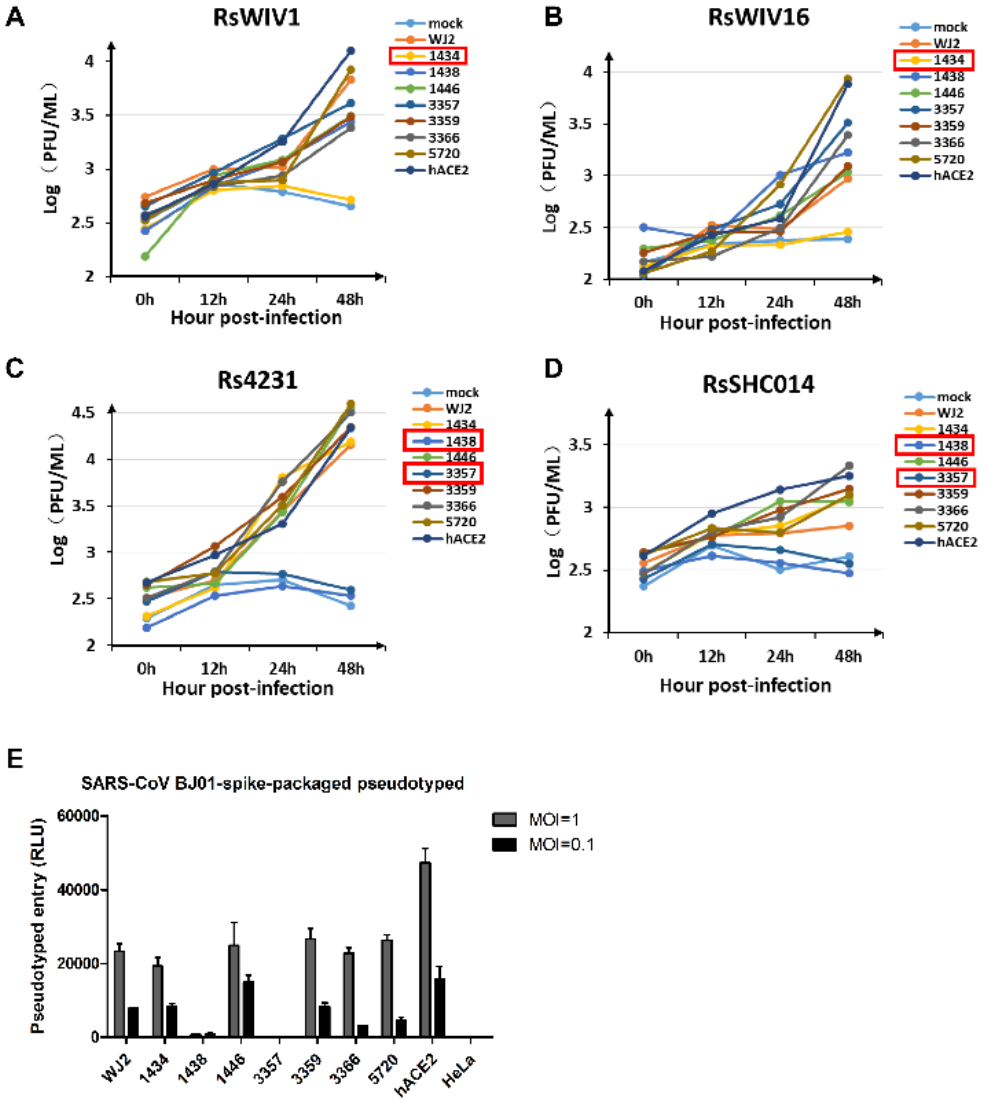

Fig. 2. Infectivity of SARSr-CoV in HeLa cells expressing *R. sinicus* ACE2. (A–D) Determination of bat SARSr-CoV infectivity in HeLa cells with and without the expression of ACE2 from *R. sinicus* or human (hACE2) at an MOI = 0.01. The growth curves were determined by real-time PCR. The number in a red square indicates the sample which is not susceptible to bat SARSr-CoV. (E) The infectivity of the SARS-CoV-BJ01 pseudotyped was used for the assay at an MOI = 1 and 0.1, due

194 to biosafety regulation in China. Error bars represent the SEM from two independent
195 transfections; each assay was performed in triplicate.
196
197 **Mutation of *R. sinicus* ACE2 residues affects its binding affinity with**
198 **SARSr-CoV RBDs**
199 To further explain the different ability of SARS and SARSr-CoVs in ACE2 usage,
200 we expressed RBD from SARS-CoV BJ01, RsWIV1, and RsSHC014 and three *R.*
201 *sinicus* ACE2s, allele 6 (sample 1434), allele 7 (sample 3357), and allele 2 (sample
202 5720), and tested the binding affinity between them. BJ01 RBD with human ACE2
203 and RsWIV1 RBD with human DPP4 were used as the positive and negative control,
204 respectively. Real-time analysis showed that the binding affinity between different
205 RBDs and ACE2s was different based on the equilibrium dissociation constant (KD)
206 (**Fig. 3 and Fig. S4**). Consistent with the results of the virus infection experiments,
207 1434ACE2 (allele 6) was found to bind RsSHC014 and BJ01 but not RsWIV1 RBD;
208 3357ACE2 (allele 7) was found to bind RsWIV1 but not RsSHC014 and BJ01 RBD;
209 5720ACE2 (allele 2) was found to bind all tested RBDs. All tested RBDs had a high
210 binding affinity to human or bat ACE2. BJ01 RBD had a higher binding affinity for
211 human ACE2 than did RsWIV1 and RsSHC014 RBDs (**Fig. 3A, E, and I**); however,
212 it had a lower binding affinity to bat ACE2 than the two bat SARSr-CoV RBDs (**Fig.**
213 **3B, D, F, G, J, K**). These results demonstrated that the binding affinity between spike
214 RBD and ACE2 affects the ability of the virus to infect.

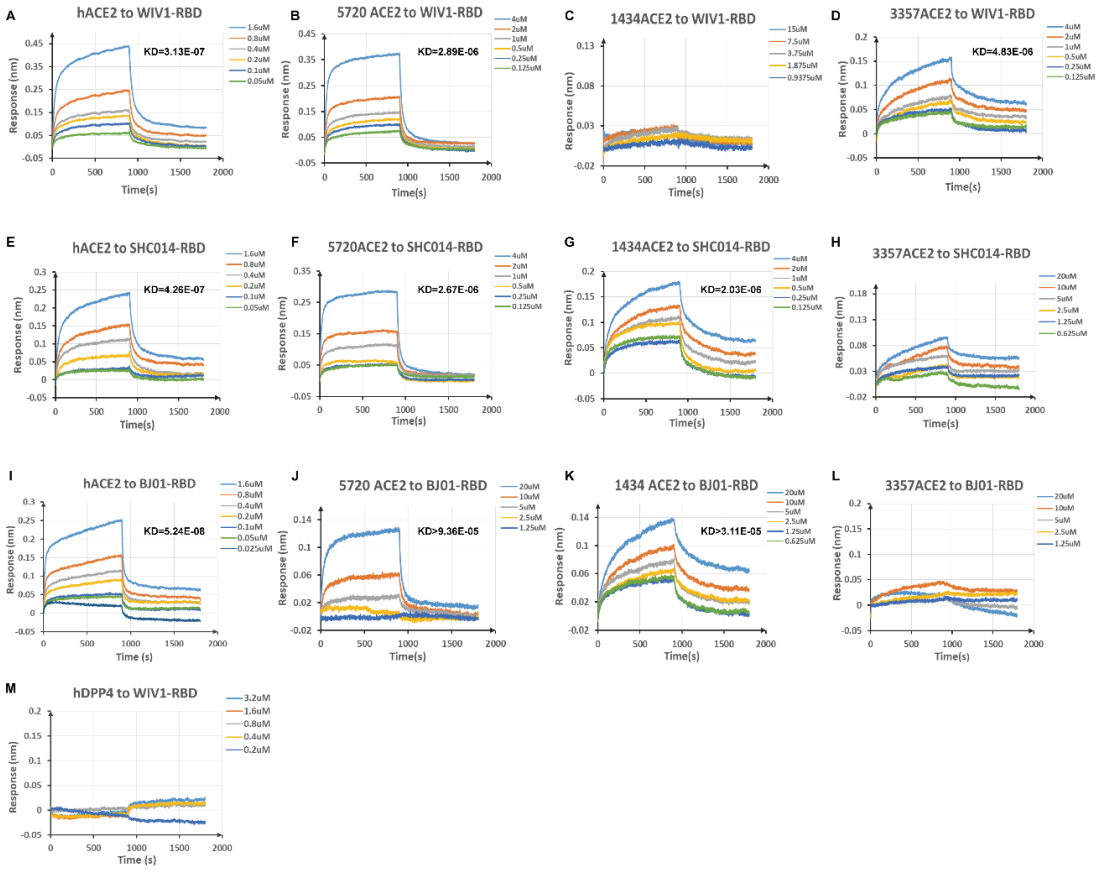

Fig. 3. Binding affinity assay between different RBDs and ACE2s.

(A–D) Binding assay of human ACE2 or bat ACE2 to WIV1 RBD. (E–H) Binding assay of human ACE2 or bat ACE2 to SHC014 RBD. (I–L) Binding assay of human ACE2 or bat ACE2 to BJ01 RBD. (M) Binding assay of human DPP4 to WIV1 RBD was used as the negative control. The parameters of KD value (M) are shown on the upper right side of the picture. Different RBD proteins were immobilized on the sensors and tested for binding with graded concentrations of *R. sinicus* ACE2s, hACE2, or hDPP4. The Y-axis shows the real-time binding response.

Structure modeling of the interaction between SARSr-CoV RBDs and *R. sinicus* ACE2s

227 The four tested spike proteins of bat SARSr-CoV are identical in size and share
228 over 90% aa identity with SARS-CoV, which suggests that these proteins have a
229 similar structure. In this study, we built structural complex models of bat
230 SARSr-CoV-RsWIV1 RBD with *R. sinicus* ACE2 3357 (allele 7) and RsSHC014
231 RBD with *R. sinicus* ACE2 1434 (allele 6), in concordance with the results of the
232 binding affinity assay between SARS-CoV RBD and human ACE2 (**Fig. 4**).
233 Compared with the contact residues in the interface between SARS-CoV RBD and
234 human ACE2, changes in or near the two virus-binding hotspots on ACE2 (hotspots
235 Lys31 and Lys353 consist of a salt bridge with Glu35 and Asp38, respectively) were
236 observed. As reported previously, the two hot spots are buried in a hydrophobic
237 environment. They provide a substantial amount of energy to the virus-receptor
238 binding interactions as well as fill critical voids in the hydrophobic stacking
239 interactions at the binding interface (21, 29).
240 Compared with human ACE2 (**Fig. 4A, C, D**), *R. sinicus* ACE2-3357 contains
241 two main residue changes, E35K and Y41H. The E35K breaks the salt bridge with
242 K31 and Arg479 in RsSHC014 RBD, which may influence the binding affinity
243 between them. A histidine at position 41 may weaken the support for the K353–D38
244 salt bridge because it is a weaker hydrophobic stacker than tyrosine, resulting in
245 decreased binding affinity with BJ01 RBD (30). Consequently, BJ01 and RsSHC014
246 RBD showed a lower binding affinity with *R. sinicus* ACE2-3357.
247 In *R. sinicus* ACE2-1434, we found a threonine at 31, unlike human ACE2, which
248 has a lysine at this position (**Fig. 4**). Although the K31T change would fail to form a

249 salt bridge with Glu35, Tyr442 in BJ01 RBD can form a hydrogen bond with it (29).
250 However, a serine at 442 in RsWIV1 RBD cannot. Besides, RsSHC014 contains an
251 arginine at 479, and Thr31 cannot form a salt bridge with Glu35, making Glu35
252 available to form a salt bridge with Arg479, but the RBD residue Asn479 in RsWIV1
253 may lose this ability. Consequently, BJ01 and RsSHC014 RBD can bind with *R.*
254 *sinicus* ACE2-1434, but not RsWIV1.

255 In addition, all *R. sinicus* ACE2s in this study contain an asparagine at position
256 82 rather than a methionine in human ACE2. The M82N change introduces an
257 unfavorable hydrophilic residue which would weaken the hydrophobic network
258 around hot spot 31. Moreover, Asn82 introduces a glycosylation site, like that in rat
259 ACE2, which cannot support SARS-CoV infection efficiently; the glycan at position
260 82 of ACE2 may lead to steric interference with viral RBD binding. Hence, M82N
261 may have significant effects on the interaction between SARS-CoV and SARSr-CoV
262 RBD with *R. sinicus* ACE2s. As previously described, residue 487 of the RBD
263 interacts with hot spot 353 on ACE2. RsWIV1 contains an Asn487 in its RBD, the
264 polar side chain of Asn487 may have unfavorable interactions with the aliphatic
265 portion of residue Lys353 in ACE2, which may affect the hot spot interaction between
266 K353 and D38. Moreover, the RsSHC014 RBD contains an alanine at position 487;
267 the small side chain of Ala 487 does not provide support to the structure of hot spot
268 353 (19). Therefore, both RsWIV1 and RsSHC014 RBD had a lower binding affinity
269 to human ACE2 than did BJ01, but they both showed a higher binding affinity with

human ACE2 than *R. sinicus* ACE2s, which correlates well with the results from virus infection and binding assays.

A

	Residues in ACE2 contact with RBD								Supports infection?		
	24	27	31	34	35	38	41	82	WIV1	SHC014	BJ01
human-ACE2	Q	T	K	H	E	D	Y	M	YES	YES	YES
1434-ACE2	R	M	T	S	E	D	Y	N	NO	YES	YES
3357-ACE2	E	M	K	T	K	D	H	N	YES	NO	NO
5720-ACE2	L	I	E	F	E	N	Y	N	YES	YES	YES
Residues in SARS-CoV RBD contact with human ACE2	N473	Y475	Y475/ Y442	Y440/ N479	N479	Y436	Y484/ T486/ T487	L472			

B

RBD	Host	Residues in RBD contact with ACE2				
		442	472	479	480	487
hBJ01	human	Y	L	N	D	T
WIV1	*R.sinicus*	S	F	N	D	N
SHC014	*R.sinicus*	W	P	R	P	A

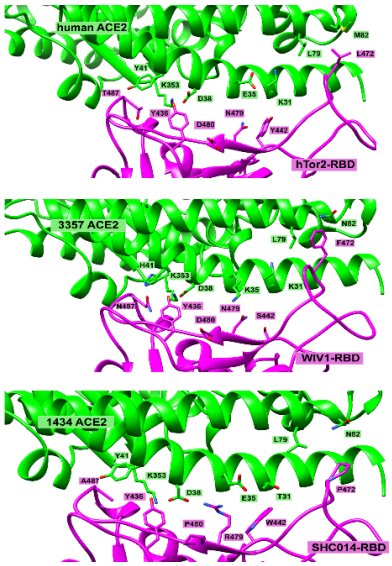

Fig. 4. Structure modeling at the interface between SARSr-CoV spike and human or bat ACE2.

Detailed view of the interaction between RBD and ACE2. Several important residues in RBD and ACE2 that are involved in the interactions are shown. ACE2 residues are in green and RBD residues are in magenta. (A) List of eight contact residues in ACE2 from human and *R. sinicus* that are directly involved in RBM/ACE2 binding. (B) Alignment of five residues from SARS-CoV and two SARSr-CoVs RBD that are critical for receptor binding. (C) Structure complex of the SARS-CoV hTor2 RBD and human ACE2 (protein data code 2AJF). (D) Predicted structure complex of *R. sinicus* ACE2 3357 with RsWIV1 RBD. (E) Predicted structure complex of *R. sinicus* ACE2 1434 with RsSHC014 RBD. The model for RsWIV1 RBD and ACE2 was built based on the structure of hTor2 RBD with hACE2 (protein data code 2AJF). The model for RsSHC014 RBD and ACE2 was built based on the structure of civet-optimized RBD

with hACE2 (protein data code 3CSJ). The model for (C) has been adapted from previous studies (19, 29).

SARSr-CoV spike genes have coevolved with *R. sinicus* ACE2 through positive selection

To test the possible selection pressure acting on the SARSr-CoV spike and *R. sinicus* ACE2 gene, we used the codeml program of the PAML software package (31) to analyze the ratio of nonsynonymous to synonymous mutations (dN/dS ratios) at individual codons. We analyzed the complete gene encoding the SARSr-CoV spike protein by aligning nine SARSr-CoV spike gene sequences from *R. sinicus* samples. We found that the models allowed codons to evolve under positive selection (M2a and M8). In model M8 (initial seed value for ω (dN/dS) = 1.6, codon frequency = F3X4), 20 codons (p>0.95) were under positive selection with dN/dS>1 (**Table 1**), 17 of those were found to be located on the RBD region, which faces its receptor ACE2, according to the crystal structure (**Fig. S5)**. Moreover, five of those (442, 472, 479, 480, and 487), present in the SARS-CoV spike, have been previously identified to have a significant impact on binding affinity to human ACE2 (**Fig. 4, Table S3 and Fig. S5**) (32). We next analyzed the ACE2 gene of *R. sinicus* by aligning 25 *R. sinicus* ACE2 gene sequences obtained in this study and downloaded from GenBank. We found that 12 codons (2.3%, p>0.95) were under positive selection with dN/dS>1 in model M8 (**Table 1**), and 8 of them (24, 27, 31, 34, 35, 38, 41, and 42) correspond to the residues in human ACE2, which were previously identified to be involved in

direct contact with the human SARS-CoV spike protein (**Fig. 4, Fig. S5**) (32). We also analyzed the ACE2 gene of *Rhinolophus affinis* (*R. affinis*), which has been reported to carry SARSr-CoV occasionally (15). Used an alignment of 23 ACE2 gene sequences from *R. affinis* obtained in this study, we found that *R. affinis* ACE2 was more conserved between different individuals in the entire coding region than *R. sinicus* ACE2 (**Fig. S6**) and no obvious positive selection sites were observed (data not shown). Additionally, by querying single nucleotide polymorphism (SNP) databases, we found that SNPs in human ACE2 gene arose randomly through the entire coding region (https://www.ncbi.nlm.nih.gov/snp/). Although we identified two SNP sites with nonsynonymous mutation (T27A and E35K) in human ACE2, both displayed a rare frequency in the global population (frequency 0.00001 and 0.00002, respectively). These results indicate that positive selection has happened at the interface between bat SARSr-CoV spike protein and *R. sinicus* ACE2.

322

323 **Table 1. PAML analysis of *R.sinicus* ACE2 and SARSr-CoV spike sequences**

324

dataset [a]	ω 0, codon freq [b]	Model comparison [c]						dN/dS (%) [d]	Residues with dN/dS>1 *p>0.95 **p>0.99 [e]
		M1a vs M2a		M7 vs M8		M8a vs M8			
		2δ	p-value	2δ	p-value	2δ	p-value		
ACE2 Full length	0.4, F61	70.92	p=4.441E-16	60.66	p=6.728E-14	11.82	p=1.356E-3	16.5(1.9)	S5*,L24**,I27**,E31**,F34**,E35*,N38**,Y41*,Q42*,P84**,H139*,N159**
	0.4, F3X4	72.26	p=2.22E-16	74.32	p=1.11E-16	72.27	p=1.11E-16	15.6(2.3)	S5*,L24**,I27**,E31**,F34**,E35*,N38**,Y41*,Q42*,P84**,H139*,N159**
	1.6, F61	70.92	p=4.441E-16	74.58	p=1.11E-16	11.82	p=1.356E-3	16.5(1.9)	S5*,L24**,I27**,E31**,F34**,E35*,N38**,Y41*,Q42*,P84**,H139*,N159**
	1.6, F3X4	72.26	P=2.22E-16	74.32	p=1.11E-16	72.27	p=1.11E-16	15.4(2.3)	S5*,L24**,I27**,E31**,F34**,E35*,N38**,Y41*,Q42*,P84**,H139*,N159**
Spike Full length	0.4, F61	59.52	p=1.189E-13	80.54	p<1.11E-16	59.52	p=5.945E-14	3.9(4.9)	Y151*,T417**,R426**,N427**,I428*,Q432**,K439**,S442*,H445**,G446**,R449**,I455*,V458*,P459**,D463*,P470**,F472**,W476**,N479**,D480**,N487**
	0.4, F3X4	54.22	p=1.684E-12	61.86	p=4.041E-14	49.86	p=7.445E-12	7.6(2)	H37*,Y151*,S198*,T417**,R426**,N427**,Q432**,K439**,S442*,H445**,G446**,R449**,V458*,P459*,P470*,F472*,W476**,N479*,D480**,N487*
	1.6, F61	59.52	p=1.189E-13	80.54	p<1.11E-16	59.52	p=5.945E-14	3.9(4.9)	Y151*,T417**,R426**,N427**,I428*,Q432**,K439**,S442*,H445**,G446**,R449**,I455*,V458*,P459**,D463*,P470**,F472**,W476**,N479**,D480**,N487**
	1.6, F3X4	54.22	p=1.684E-12	61.86	p=4.041E-14	49.86	p=7.445E-12	7.6(2)	H37*,Y151*,S198*,T417**,R426**,N427**,Q432**,K439**,S442*,H445**,G446**,R449**,V458*,P459*,P470*,F472*,W476**,N479*,D480**,N487*

325

326 a, Dataset consisted of the aligned sequences of *R.sinicus* ACE2 and SARSr-CoV spike sequences were list in table S2.

327 b, Initial seed value for ω (dN/dS) used in the maximum likelihood simulation and model of codon frequency.

328 c, Twice the difference in the natural logs of the likelihoods (ΔlnL x 2) of the two models being compared. This value is used in a
329 likelihood ratio test along with the degrees of freedom. In all cases (M1a vs M2a), (M7 vs M8), (M8a vs M8), a model that
330 allows positive selection is compared to a null model. The p-value indicates the confidence with which the null model can be
331 rejected.

332 d, dN/dS value of the class of codons evolving under positive selection in model M8, and the percent of codons falling in that
333 class

334 e, Residues corresponding to codons assigned to the dN/dS>1 class in model M8 (p>0.95 by BEB).

335

Discussion

Chinese horseshoe bats are widely distributed in China and carry genetically diverse SARSr-CoVs. Two different clades of bat SARSr-CoVs were discovered, based on the size and similarity of S protein to that of human SARS-CoV. SARSr-CoVs of clade-1, only found in Yunnan province, have an S protein identical in size to SARS-CoV and use the ACE2 receptor, whereas SARSr-CoVs of clade-2 are widely distributed in China and cannot use ACE2 as the receptor (19). In this study, we analyzed ACE2 sequences of *R. sinicus*, collected from four provinces and one city, and observed highly polymorphic sites, which correspond to those that interact with SARS-CoV RBD, at the N-terminal region. Despite these variations, most ACE2s supported the viral entry of clade-1 bat SARSr-CoVs, but with different susceptibilities and binding affinities. In addition, we have identified key residues involved in the interaction between the bat ACE2 variants and SARSr-CoV RBD by structural modeling and positive selection analysis. These results indicate that SARSr-CoV has co-evolved with its natural host, *R. sinicus*, for a very long time. Members of clade-1 use ACE2 as the receptor, whereas members of clade-2 exploit different receptor(s), due to the deletions in S protein.

Genes with important functions usually display a dN/dS ratio of less than 1 (negative selection) because most amino acid alterations in a protein are deleterious. In a host-virus arms race situation, the genes involved tend to display dN/dS ratios greater than 1 (positive selection), specifically in the codons involved in the interaction interface between the virus and its host, with minimal effect on their

358 physical function (20, 33). In this study, our analysis of the SARSr-CoV spike protein
359 and *R. sinicus* ACE2 gene sequences showed that some codons related to the
360 interaction interface between them were under positive selection. Similar rapid
361 adaptation for SARS-CoV occurred during the SARS outbreak in 2002–2003 (21, 34).
362 When SARS-CoV was transmitted from market civet to human, the spike gene
363 experienced positive selection, where mutations in two critical residues (amino acids
364 479 and 487) of the spike protein changed the binding affinity of the virus to human
365 ACE2 from low to high; subsequently, turning into a pandemic strain (34).

366 An increasing number of studies have suggested that co-evolution is driven by
367 specific interactions between high levels of virus sequence divergence and
368 polymorphic host receptors (20, 33, 35-38). The first example was observed for the
369 avian leucosis virus infection in which receptor polymorphism in chicken could alter
370 the sensitivity to virus entry (36). Arms race can also occur between the glycoproteins
371 of two different viruses, as seen for the rodent arenavirus and the mouse mammary
372 tumor virus, with their receptor transferrin receptor 1 (TfR1) (39).

373 In a continuous coevolutionary process, viruses often tend to decrease their
374 virulence to fit both the host and the virus population. The “trade-off hypothesis” is
375 the most popular explanation for why pathogens often do not reach their maximum
376 reproductive potential (40, 41). In this study, we have found that bat SARSr-CoV
377 RBDs showed a lower binding affinity to *R. sinicus* ACE2 than to human ACE2, but
378 did not display obvious differences among the *R. sinicus* ACE2s, indicating that bat
379 SARSr-CoVs may decrease their virulence to fit both the host and themselves. When

they adapt to fit other host receptors better, cross-species transmission events could happen. This situation can be demonstrated by the higher binding affinity seen between SARS-CoV-BJ01 RBDs and human ACE2s than bat SARSr-CoV strains in this study. Similar examples can be found in other cases of coevolution between host receptor and virus (36, 39, 42, 43). Previous studies have shown adaptive evolution of the coronavirus spike protein to host receptor *in vivo* and *in vitro* (44-48). Recently, the MERS-CoV spike protein was shown to rapidly adapt to *Desmodus rotundus* DPP4 (*dr*DPP4) receptor from semi-permissive to permissive state after several passages *in vitro*. Furthermore, mutations detected in the RBD of MERS-CoV spike protein enhanced viral entry and replication on cells expressing *dr*DPP4 within three passages (49). Similarly, adaptive mutations have occurred in the Ebola virus envelope protein, which contacts its receptor, NPC intracellular cholesterol transporter 1, when it crossed the host-virus barriers, was transmitted to different hosts, and entered the human population (50-53). These examples are a warning that the SARSr-CoVs circulating in *R. sinicus* may adapt to other animal hosts, including humans, and cause the next SARS-like disease. Therefore, it is important to continually monitor SARSr-CoVs in *R. sinicus* populations to avoid future spillover events. Our study provides an example to assess the risk of a potential spillover from viruses circulating in animal populations through the positive selection analysis, structural modeling, and experimental verification.

ACE2, a multifunctional enzyme involved in the negative regulation of the renin-angiotensin system, also interacts with amino acid transporters and integrins (54,

55). ACE2 functions as a carboxypeptidase and its role as the SARS-CoV receptor does not affect its peptidase activity. Moreover, the structural complex shows that the SARS-CoV spike RBD does not block the catalytically-active site of ACE2 (21, 56). Considering the conserved recognition site in receptor of SARSr-CoV clade 1, the interface of SARS-CoV and ACE2 interaction would be an ideal target for drug screening against a broad range of infections caused by these viruses.

Materials and Methods

All work with the infectious virus was performed under biosafety level 2 conditions with appropriate personal protection.

Bat samples, cells, and viruses

Bat samples were from the biobank of our laboratory. Vero E6, HeLa, and HEK 293T/17 cells (ATCC) were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS). Cells were cultured at 37°C with 5% CO_2. Bat SARSr-CoV-RsWIV1 and recombinant viruses with RsWIV1 as the backbone and replaced by different spike genes of bat SARSr-CoVs (RsWIV16, RsSHC014, and Rs4231) were cultured as previously described (16, 57). Titers of all viruses used in this study were determined by plaque assays on Vero E6 cells, as previously described (17).

Amplification, cloning, and expression of bats ACE2 gene

Bat ACE2 gene was amplified using DNA from the intestinal tissue of bats. In brief, total RNA was extracted from bat intestine tissue using the RNAprep pure Kit (for Cell/Bacteria) (TIANGEN, Beijing, China). First-strand complementary DNA was synthesized from total RNA by reverse transcription with random hexamers; full-length bat ACE2 fragments were amplified by reverse-transcription nested polymerase chain reaction (RT-PCR). Primers were designed based on available ACE2 sequences from NCBI. First-round primers are as follows: F-ACE2-out-AATGGGGTTTTGGCGCTCAG, R-ACE2-out-CATACAATGAAATCACCTCAAGAG, second-round primers: F-ACE2-in-CAGGGAAAGATGTCAGGCTC, R-ACE2-in-TTCTAAAABGAVGTYTGAAC. The ACE2 gene was cloned into the pcDNA3.1 vector with XhoI and BamHI. N-terminal mouse Igk or human IFNα1 signal peptide and S-tag were inserted. The plasmids were verified by sequencing. The expression of human and *R. sinicus* ACE2 was confirmed by western blotting and immunofluorescence assay.

Expression constructs, protein expression, and purification

The constructs used for protein expression and purification were individually prepared by inserting coding sequences for SARS-CoV-BJ01 RBD (spike residues: aa 317–569，accession number: AY278488.2), RsWIV1 RBD (spike residues: aa 318–570，accession number: KF367457.1), RsSHC014 RBD (spike residues: aa 318–570, accession number: KC881005.1), *R. sinicus* ACE2 (aa 19–615, accession numbers are

447 listed in Table S2), human ACE2 (aa 19–615, accession number: BAJ21180), and
448 human DPP4 (aa 39–766, accession number: NP_001926) into the expression vector,
449 as described previously (58). For each protein, an N-terminal signal peptide and an
450 S-tag were added to facilitate protein secretion and purification. The proteins used for
451 the Octet RED binding assay were expressed in HEK 293T/17 cells. Cells were
452 washed twice with D-Hanks solution 6 h post-transfection, then cultured in fresh
453 293T FreeStyle expression medium. The cells were collected 48 h post-transfection
454 and centrifuged at 4000 × *g* for 10 min at 4°C. The supernatant was incubated with
455 S-tag agarose beads overnight at 4°C, passed through a column, and the protein was
456 eluted from the column with 3 M $MgCl_2$. The protein was finally buffered with PBS
457 and stored at -80 °C until further use.
458
459 **Immunofluorescence assay**
460 HeLa cells were seeded in 24-well plates and transfected with equal amounts of
461 human or *R. sinicus* ACE2 plasmids. After 24 h, the cells were incubated with
462 different SARSr-CoV strains at a multiplicity of infection (MOI) = 1 for 1 h.
463 Thereafter, the cells were washed with D-Hanks solution thrice and supplemented
464 with the medium. At 24 h after infection, cells were washed with PBS and fixed with
465 4% formaldehyde for 30 min at room temperature (about 25°C). ACE2 expression
466 was detected using mouse anti-S-tag monoclonal antibody (Kyab Biotech Co. Ltd.,
467 Wuhan, China), followed by FITC-labelled goat anti-mouse IgG H&L (Abcam,
468 Cambridge Biomedical Campus, Cambridge, UK). Virus replication was detected

using rabbit serum against the SARSr-CoV-Rp3 NP, as previously described (14), and Cy3-conjugated goat anti-rabbit IgG (Abcam, Cambridge Biomedical Campus, Cambridge, UK). Nuclei were stained with 4′ ,6-diamidino-2-phenylindole. Staining patterns were examined using a fluorescence imaging system (EVOS™ FL Color Imaging System, Thermo Fisher Scientific, Waltham, MA, USA)

Real-time PCR

SARSr-CoV infection was detected by RT-PCR, as described previously (14). In brief, HeLa cells were transfected with equal amounts of human or *R. sinicus* ACE2 plasmids, 24 h before they were infected with the virus, at an MOI = 0.01 with 1 h adsorption. Thereafter, the cells were washed with D-Hanks solution three times and cultured in 1 mL DMEM+2% FBS. The viral supernatants were harvested at 0, 12, 24, and 48 h post-infection—200 μL supernatant was removed and an equal amount of medium was added back at each time point. Viral RNA was extracted from the supernatant using the Viral RNA Mini Kit (Qiagen, Waltham, MA, USA) and then quantified on the ABI Step One system (Thermo Fisher Scientific, Waltham, MA, USA), with the AgPath-ID One-Step RT-PCR Kit (Applied Biosystems life technologies, Waltham, MA, USA). RNA dilutions from purified SARSr-CoV-RsWIV1 stock were used as a standard (with a titer of 6.5×10^6 PFU/mL). Every sample was analyzed in triplicate on two independent occasions.

Binding assay

Binding assays between SARS/SARSr-CoV RBD and human or *R. sinicus* ACE2 protein were performed using the Octet RED system (ForteBio, Menlo Park, CA, USA) in 96-well microplates at 30°C, as described previously (59, 60). Briefly, the RBD was biotinylated using EZ-Link NHS-LC-LC-Biotin (Thermo Fisher Scientific, Waltham, MA, USA). The assay was carried out by placing the Streptavidin Biosensors (ForteBio, Menlo Park, CA, USA) in the wells and measuring changes in layer thickness (nm) over time (s). First, the sensors were rinsed in the kinetic buffer (1 M NaCl, 0.1% BSA, 0.02% Tween-20; pH 6.5) for 120 s, which served as the baseline. Thereafter, the sensors were immobilized for 600 s, with 200 µL buffer containing biotinylated SARSr-CoV RBD (50 µg/mL). Subsequently, the sensors were washed in the kinetic buffer for 200 s. Finally, the sensors were exposed to a series of diluted ACE2 protein run in 200 µL volumes. The association between each RBD and its corresponding ACE2 binding partner was monitored for 900 s. This was followed by monitoring their dissociation in the kinetic buffer for 900 s. During the entire experiment, the sample plate was kept shaking at 1000 rpm, and an RBD-loaded biosensor was used in association with the buffer as a baseline. Data analysis from the ForteBio Octet RED instrument includes reference subtraction. Inter-step correction and Y-alignment were used to minimize tip-dependent variability. Data were globally fitted in a 1:1 model using the Data Analysis Software v7.1 (ForteBio, Menlo Park, CA, USA).

Pseudotyped production and infection of ACE2-transfected HeLa cells

513 SARS-CoV-BJ01 pseudotyped particles were produced by HEK 293T/17 cells.
514 Plasmids, pcDNA3.1-BJ01-S and pNL4-3.luc.R-E-, were co-transfected into HEK
515 293T/17 cells. At 6 h post-transfection, the medium was replaced with fresh 293T
516 FreeStyle expression medium. The supernatant, which contained the pseudotyped
517 particles, was harvested 48 h post-transfection and filtered using 0.45 μm filters.
518 Thereafter, the pseudotyped particles were aliquoted and stored at -80°C until further
519 use. The viral titer was determined by the HIV p24 Quantitation ELISA Kits (Kyab
520 Biotech Co. Ltd., Wuhan, China) before the viruses were used for the infection assay.
521 HeLa cells were transfected with the same amount of human or *R. sinicus* ACE2
522 plasmids. After 24 h, the cells were incubated with SARS-CoV-BJ01 pseudotyped at
523 an MOI = 1 or 0.1 for 1 h at 37°C, washed twice with D-Hanks solution, and
524 supplemented with DMEM containing 10% FBS. Luciferase activity was determined
525 using a GloMax luminometer (Promega Biotech Co. Ltd., Beijing, China) 48 h after
526 infection. Each sample was analyzed in triplicate on two independent occasions.
527

528 **Structure modeling**

529 The structure complex of SARS-CoV/SARSr-CoVs RBD and *R. sinicus* ACE2 was
530 homology modeled using SWISS-MODEL (https://swissmodel.expasy.org/), based on
531 the structure of human SARS-CoV RBD (hTor2 RBD) complexed with human ACE2
532 (hACE2) and civet-optimized RBD complexed with hACE2 (Protein Data Bank ID:
533 2AJF, 3SCJ) (29). Molecular graphics visualization and analyses were performed
534 using the UCSF Chimera software (http://www.rbvi.ucsf.edu/chimera).

535

536 **Codon-based analysis of molecular evolution**

537 Bat ACE2 and SARSr-CoV spike sequences were analyzed for positive selection. In
538 this study, bat ACE2 sequences were either amplified or downloaded from NCBI and
539 SARSr-CoV spike sequences were downloaded from NCBI; the database accession
540 numbers are listed in Table S2. Sequences were aligned in Clustal X. Phylogenetic
541 trees were built by the maximum likelihood method implemented in RAxML program
542 in CIPRES Science Gateway (https://www.phylo.org/). Maximum likelihood analysis
543 of dN/dS was performed using the codeml program in the PAML4.7 software package,
544 as previously described (31, 39). In brief, multiple alignments were fit to NSsites
545 models M1a, M2a, M7, M8a, and M8. Model fitting was performed with multiple
546 seed values for dN/dS and assuming either the F61 or F3x4 model of codon
547 frequencies. We used likelihood ratio tests (LRTs) to assess a better fit of codons that
548 allowed positive selection. Posterior probabilities of codons under positive selection
549 in different models were inferred using the Bayes empirical Bayes (BEB) algorithm.

550

551 **Accession numbers**

552 The complete ACE2 sequences of *R. sinicus* and *R. affinis* ACE2 obtained in this
553 study have been deposited in the GenBank database and the accession numbers are
554 listed in Table S2.

555

556 **Acknowledgments**

This work was jointly funded by the strategic priority research program of the Chinese Academy of Sciences (XDB29010101), National Natural Science Foundation of China (31770175 and 31621061) to Z.L.S.

We would like to thank Dr. Bing Yan for providing S-tag Agarose beads and mouse anti-S-tag monoclonal antibodies. We also thank Dr. Ding Gao in the Core Facility and Technical Support of the Wuhan Institute of Virology for his help in Octet RED technology. We thank Microorganisms & Viruses Culture Collection Center, Wuhan Institute of Virology, Chinese Academy of Sciences for providing HeLa and HEK 293T/17 cells.

Declaration of Interests

The authors declare no competing interests.

Author contributions:

H.G. and Z.L.S. designed the research; H.G., B.J.H., L.P.Z., and B.L. performed the research; H.G. and Z.L.S. analyzed the data; H.G., X.L.Y., S.Y.O.Y., and Z.L.S. wrote the paper.

H.G. and B.J.H. contributed equally to this work.

References

1. Woo PCY, Lau SKP, Lam CSF, Lau CCY, Tsang AKL, Lau JHN, Bai R, Teng JLL, Tsang CCC, Wang M, Zheng B-J, Chan K-H, Yuen K-Y. 2012. Discovery of seven novel Mammalian and avian

coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol 86:3995-4008.

2. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF. 2016. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends Microbiol 24:490-502.

3. Zhong NS, Zheng BJ, Li YM, Poon, Xie ZH, Chan KH, Li PH, Tan SY, Chang Q, Xie JP, Liu XQ, Xu J, Li DX, Yuen KY, Peiris, Guan Y. 2003. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet 362:1353-8.

4. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Berger A, Burguiere AM, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra JC, Muller S, Rickerts V, Sturmer M, Vieth S, Klenk HD, Osterhaus AD, Schmitz H, Doerr HW. 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 348:1967-76.

5. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ, Group SW. 2003. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 348:1953-66.

6. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. 2012. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. New England Journal of Medicine 367:1814-1820.

7. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270-273.

8. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL. 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302:276-8.

9. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY. 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A 102:14040-5.

10. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang LF. 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science 310:676-9.

11. Yuan J, Hon CC, Li Y, Wang D, Xu G, Zhang H, Zhou P, Poon LL, Lam TT, Leung FC, Shi Z. 2010. Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans. J Gen Virol 91:1058-62.

12. Gouilh MA, Puechmaille SJ, Gonzalez JP, Teeling E, Kittayapong P, Manuguerra JC. 2011. SARS-Coronavirus ancestor's foot-prints in South-East Asian bat colonies and the refuge theory. Infect Genet Evol 11:1690-702.

13. Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, Goettsche M, Seebens A, Niedrig M, Pfefferle S, Yordanov S, Zhelyazkov L, Hermanns U, Vallo P, Lukashev A, Muller MA, Deng H,

624 Herrler G, Drosten C. 2010. Genomic characterization of severe acute respiratory
625 syndrome-related coronavirus in European bats and classification of coronaviruses based on
626 partial RNA-dependent RNA polymerase gene sequences. J Virol 84:11336-49.
627 14. Ge X-Y, Li J-L, Yang X-L, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang
628 Y-J, Luo C-M, Tan B, Wang N, Zhu Y, Crameri G, Zhang S-Y, Wang L-F, Daszak P, Shi Z-L. 2013.
629 Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor.
630 Nature 503:535-538.
631 15. He B, Zhang Y, Xu L, Yang W, Yang F, Feng Y, Xia L, Zhou J, Zhen W, Feng Y, Guo H, Zhang H, Tu C.
632 2014. Identification of diverse alphacoronaviruses and genomic characterization of a novel
633 severe acute respiratory syndrome-like coronavirus from bats in China. J Virol 88:7070-82.
634 16. Yang XL, Hu B, Wang B, Wang MN, Zhang Q, Zhang W, Wu LJ, Ge XY, Zhang YZ, Daszak P, Wang
635 LF, Shi ZL. 2015. Isolation and Characterization of a Novel Bat Coronavirus Closely Related to
636 the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. J Virol 90:3253-6.
637 17. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ, Wang N, Luo DS,
638 Zheng XS, Wang MN, Daszak P, Wang LF, Cui J, Shi ZL. 2017. Discovery of a rich gene pool of
639 bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus.
640 PLoS Pathog 13:e1006698.
641 18. Wacharapluesadee S, Duengkae P, Rodpan A, Kaewpom T, Maneeorn P, Kanchanasaka B,
642 Yingsakmongkon S, Sittidetboripat N, Chareesaen C, Khlangsap N, Pidthong A, Leadprathom K,
643 Ghai S, Epstein JH, Daszak P, Olival KJ, Blair PJ, Callahan MV, Hemachudha T. 2015. Diversity of
644 coronavirus in bats from Eastern Thailand. Virol J 12:57.
645 19. Cui J, Li F, Shi ZL. 2019. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol
646 17:181-192.
647 20. Daugherty MD, Malik HS. 2012. Rules of engagement: molecular insights from host-virus
648 arms races. Annu Rev Genet 46:677-700.
649 21. Song HD, Tu CC, Zhang GW, Wang SY, Zheng K, Lei LC, Chen QX, Gao YW, Zhou HQ, Xiang H,
650 Zheng HJ, Chern SW, Cheng F, Pan CM, Xuan H, Chen SJ, Luo HM, Zhou DH, Liu YF, He JF, Qin
651 PZ, Li LH, Ren YQ, Liang WJ, Yu YD, Anderson L, Wang M, Xu RH, Wu XW, Zheng HY, Chen JD,
652 Liang G, Gao Y, Liao M, Fang L, Jiang LY, Li H, Chen F, Di B, He LJ, Lin JY, Tong S, Kong X, Du L,
653 Hao P, Tang H, Bernini A, Yu XJ, Spiga O, Guo ZM, et al. 2005. Cross-host evolution of severe
654 acute respiratory syndrome coronavirus in palm civet and human. Proc Natl Acad Sci U S A
655 102:2430-5.
656 22. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong SK, Huang IC, Xu K, Vasilieva N,
657 Murakami A, He Y, Marasco WA, Guan Y, Choe H, Farzan M. 2005. Receptor and viral
658 determinants of SARS-coronavirus adaptation to human ACE2. Embo J 24:1634-43.
659 23. Li F, Berardi M, Li W, Farzan M, Dormitzer PR, Harrison SC. 2006. Conformational states of the
660 severe acute respiratory syndrome coronavirus spike protein ectodomain. J Virol
661 80:6794-800.
662 24. Crackower MA, Sarao R, Oudit GY, Yagil C, Kozieradzki I, Scanga SE, Oliveira-dos-Santos AJ, da
663 Costa J, Zhang L, Pei Y, Scholey J, Ferrario CM, Manoukian AS, Chappell MC, Backx PH, Yagil Y,
664 Penninger JM. 2002. Angiotensin-converting enzyme 2 is an essential regulator of heart
665 function. Nature 417:822-828.
666 25. Demogines A, Farzan M, Sawyer SL. 2012. Evidence for ACE2-utilizing coronaviruses (CoVs)
667 related to severe acute respiratory syndrome CoV in bats. J Virol 86:6350-3.

668 26. Hou Y, Peng C, Yu M, Li Y, Han Z, Li F, Wang LF, Shi Z. 2010. Angiotensin-converting enzyme 2
669 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. Arch
670 Virol 155:1563-9.
671 27. Yu M, Tachedjian M, Crameri G, Shi Z, Wang LF. 2010. Identification of key amino acid
672 residues required for horseshoe bat angiotensin-I converting enzyme 2 to function as a
673 receptor for severe acute respiratory syndrome coronavirus. J Gen Virol 91:1708-12.
674 28. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary
675 Genetics Analysis version 6.0. Mol Biol Evol 30:2725-9.
676 29. Kailang Wu GP, Matthew Wilken, Robert J. Geraghty, Fang Li. 2012. Mechanisms of Host
677 Receptor Adaptation by Severe Acute Respiratory Syndrome Coronavirus. J Biol Chem
678 287:8904–8911.
679 30. Li F. 2013. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral
680 Res 100:246-54.
681 31. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol
682 24:1586-91.
683 32. Li F, Li W, Farzan M, Harrison SC. 2005. Structure of SARS coronavirus spike receptor-binding
684 domain complexed with receptor. Science 309:1864-8.
685 33. Meyerson NR, Sawyer SL. 2011. Two-stepping through time: mammals and viruses. Trends
686 Microbiol 19:286-94.
687 34. Chinese SMEC. 2004. Molecular evolution of the SARS coronavirus during the course of the
688 SARS epidemic in China. Science 303:1666-9.
689 35. Gupta RK, Hue S, Schaller T, Verschoor E, Pillay D, Towers GJ. 2009. Mutation of a single
690 residue renders human tetherin resistant to HIV-1 Vpu-mediated depletion. PLoS Pathog
691 5:e1000443.
692 36. Crittenden LB, Stone HA, Reamer RH, Okazaki W. 1967. Two loci controlling genetic cellular
693 resistance to avian leukosis-sarcoma viruses. J Virol 1:898-904.
694 37. Warren CJ, Meyerson NR, Stabell AC, Fattor WT, Wilkerson GK, Sawyer SL. 2019. A glycan
695 shield on chimpanzee CD4 protects against infection by primate lentiviruses (HIV/SIV).
696 Proceedings of the National Academy of Sciences 116:11460.
697 38. Bibollet-Ruche F, Russell RM, Liu W, Stewart-Jones GBE, Sherrill-Mix S, Li Y, Learn GH, Smith
698 AG, Gondim MVP, Plenderleith LJ, Decker JM, Easlick JL, Wetzel KS, Collman RG, Ding S, Finzi A,
699 Ayouba A, Peeters M, Leendertz FH, van Schijndel J, Goedmakers A, Ton E, Boesch C, Kuehl H,
700 Arandjelovic M, Dieguez P, Murai M, Colin C, Koops K, Speede S, Gonder MK, Muller MN, Sanz
701 CM, Morgan DB, Atencia R, Cox D, Piel AK, Stewart FA, Ndjango J-BN, Mjungu D, Lonsdorf EV,
702 Pusey AE, Kwong PD, Sharp PM, Shaw GM, Hahn BH. 2019. CD4 receptor diversity in
703 chimpanzees protects against SIV infection. Proceedings of the National Academy of Sciences
704 116:3229.
705 39. Demogines A, Abraham J, Choe H, Farzan M, Sawyer SL. 2013. Dual host-virus arms races
706 shape an essential housekeeping protein. PLoS Biol 11:e1001571.
707 40. Day T. 2001. Parasite Transmission Modes and the Evolution of Virulence. Evolution
708 55:2389-2400.
709 41. Morgan AD, Koskella B. 2017. Coevolution of Host and Pathogen, p 115-140. *In* Tibayrenc M
710 (ed), Genetics and Evolution of Infectious Diseases, Second Edition ed. Elsevier, Netherlands.
711 42. Adkins HB, Brojatsch J, Naughton J, Rolls MM, Pesola JM, Young JA. 1997. Identification of a

cellular receptor for subgroup E avian leukosis virus. Proceedings of the National Academy of Sciences of the United States of America 94:11617-11622.

43. O'Brien SJ, Nelson GW. 2004. Human genes that limit AIDS. Nature Genetics 36:565-574.

44. Baric RS, Sullivan E, Hensley L, Yount B, Chen W. 1999. Persistent infection promotes cross-species transmissibility of mouse hepatitis virus. J Virol 73:638-49.

45. Roberts A, Deming D, Paddock CD, Cheng A, Yount B, Vogel L, Herman BD, Sheahan T, Heise M, Genrich GL, Zaki SR, Baric R, Subbarao K. 2007. A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mice. PLoS Pathog 3:e5.

46. McRoy WC, Baric RS. 2008. Amino acid substitutions in the S2 subunit of mouse hepatitis virus variant V51 encode determinants of host range expansion. J Virol 82:1414-24.

47. Sheahan T, Rockx B, Donaldson E, Sims A, Pickles R, Corti D, Baric R. 2008. Mechanisms of zoonotic severe acute respiratory syndrome coronavirus host range expansion in human airway epithelium. J Virol 82:2274-85.

48. Wu K, Peng G, Wilken M, Geraghty RJ, Li F. 2012. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. J Biol Chem 287:8904-11.

49. Letko M, Miazgowicz K, McMinn R, Seifert SN, Sola I, Enjuanes L, Carmody A, van Doremalen N, Munster V. 2018. Adaptive Evolution of MERS-CoV to Species Variation in DPP4. Cell Rep 24:1730-1737.

50. Ng M, Ndungo E, Kaczmarek ME, Herbert AS, Binger T, Kuehne AI, Jangra RK, Hawkins JA, Gifford RJ, Biswas R, Demogines A, James RM, Yu M, Brummelkamp TR, Drosten C, Wang LF, Kuhn JH, Muller MA, Dye JM, Sawyer SL, Chandran K. 2015. Filovirus receptor NPC1 contributes to species-specific patterns of ebolavirus susceptibility in bats. Elife 4.

51. Urbanowicz RA, McClure CP, Sakuntabhai A, Sall AA, Kobinger G, Muller MA, Holmes EC, Rey FA, Simon-Loriere E, Ball JK. 2016. Human Adaptation of Ebola Virus during the West African Outbreak. Cell 167:1079-1087 e5.

52. Diehl WE, Lin AE, Grubaugh ND, Carvalho LM, Kim K, Kyawe PP, McCauley SM, Donnard E, Kucukural A, McDonel P, Schaffner SF, Garber M, Rambaut A, Andersen KG, Sabeti PC, Luban J. 2016. Ebola Virus Glycoprotein with Increased Infectivity Dominated the 2013-2016 Epidemic. Cell 167:1088-1098 e6.

53. Yang XL, Tan CW, Anderson DE, Jiang RD, Li B, Zhang W, Zhu Y, Lim XF, Zhou P, Liu XL, Guan W, Zhang L, Li SY, Zhang YZ, Wang LF, Shi ZL. 2019. Characterization of a filovirus (Mengla virus) from Rousettus bats in China. Nat Microbiol 4:390-395.

54. Kuba K, Imai Y, Ohto-Nakanishi T, Penninger JM. 2010. Trilogy of ACE2: A peptidase in the renin–angiotensin system, a SARS receptor, and a partner for amino acid transporters. Pharmacology & Therapeutics 128:119-128.

55. Crackower MA, Sarao R, Oudit GY, Yagil C, Kozieradzki I, Scanga SE, Oliveira-dos-Santos AJ, da Costa J, Zhang L, Pei Y, Scholey J, Ferrario CM, Manoukian AS, Chappell MC, Backx PH, Yagil Y, Penninger JM. 2002. Angiotensin-converting enzyme 2 is an essential regulator of heart function. Nature 417:822-8.

56. Moore MJ, Dorfman T, Li W, Wong SK, Li Y, Kuhn JH, Coderre J, Vasilieva N, Han Z, Greenough TC, Farzan M, Choe H. 2004. Retroviruses pseudotyped with the severe acute respiratory syndrome coronavirus spike protein efficiently infect cells expressing angiotensin-converting enzyme 2. J Virol 78:10628-35.

57. Zeng LP, Gao YT, Ge XY, Zhang Q, Peng C, Yang XL, Tan B, Chen J, Chmura AA, Daszak P, Shi ZL.

2016. Bat Severe Acute Respiratory Syndrome-Like Coronavirus WIV1 Encodes an Extra Accessory Protein, ORFX, Involved in Modulation of the Host Immune Response. J Virol 90:6573-6582.

58. Wang Q, Qi J, Yuan Y, Xuan Y, Han P, Wan Y, Ji W, Li Y, Wu Y, Wang J, Iwamoto A, Woo PC, Yuen KY, Yan J, Lu G, Gao GF. 2014. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. Cell Host Microbe 16:328-37.

59. Scietti L, Sampieri K, Pinzuti I, Bartolini E, Benucci B, Liguori A, Haag AF, Lo Surdo P, Pansegrau W, Nardi-Dei V, Santini L, Arora S, Leber X, Rindi S, Savino S, Costantino P, Maione D, Merola M, Speziale P, Bottomley MJ, Bagnoli F, Masignani V, Pizza M, Scharenberg M, Schlaeppi JM, Nissum M, Liberatori S. 2016. Exploring host-pathogen interactions through genome wide protein microarray analysis. Sci Rep 6:27996.

60. Stephenson RJ, Toth I, Liang J, Mangat A, McManus DP, You H. 2016. Identification of Host Insulin Binding Sites on Schistosoma japonicum Insulin Receptors. PLoS One 11:e0159704.