AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

Evolutionary Arms Race between Virus and Host Drives Genetic Diversity in Bat Severe Acute Respiratory Syndrome-Related Coronavirus Spike Genes

Hua Guo,[a,b] Bing-Jie Hu,[a] Xing-Lou Yang,[a] Lei-Ping Zeng,[a] Bei Li,[a] Songying 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

Hua Guo and Bing-Jie Hu contributed equally to this work. Author order was determined by project originator and work time.

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 the 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 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 efficiencies in cells transiently expressing bat ACE2 variants. Consistent results were observed by binding affinity assays between SARS-CoV 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 that of 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 some key residues were under 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, triggering the evolutionary arms race dynamics.

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-CoVs in Chinese horseshoe bats. Here, we show the highly polymorphic ACE2 in Chinese horseshoe bat populations. These ACE2 variants support SARS-CoV and SARSr-CoV infection but with different binding affinities to different spike proteins. The higher binding affinity of SARSr-CoV spike to human ACE2 suggests that these viruses have the capacity for spillover to humans. The positive selection of residues at the interface between ACE2 and SARSr-CoV spike protein suggests 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.

KEYWORDS SARS-related coronavirus, spike gene, receptor, ACE2, Chinese horseshoe bat, genetic diversity

Citation Guo H, Hu B-J, Yang X-L, Zeng L-P, Li B, Ouyang S, Shi Z-L. 2020. Evolutionary arms race between virus and host drives genetic diversity in bat severe acute respiratory syndrome-related coronavirus spike genes. J Virol 94:e00902-20. https://doi.org/10.1128/JVI.00902-20.

Editor Julie K. Pfeiffer, University of Texas Southwestern Medical Center

Address correspondence to Zheng-Li Shi, zlshi@wh.iov.cn.

Received 8 May 2020
Accepted 17 July 2020
Accepted manuscript posted online 22 July 2020
Published 29 September 2020

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* are believed to have originated 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, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), outbreaks (7).

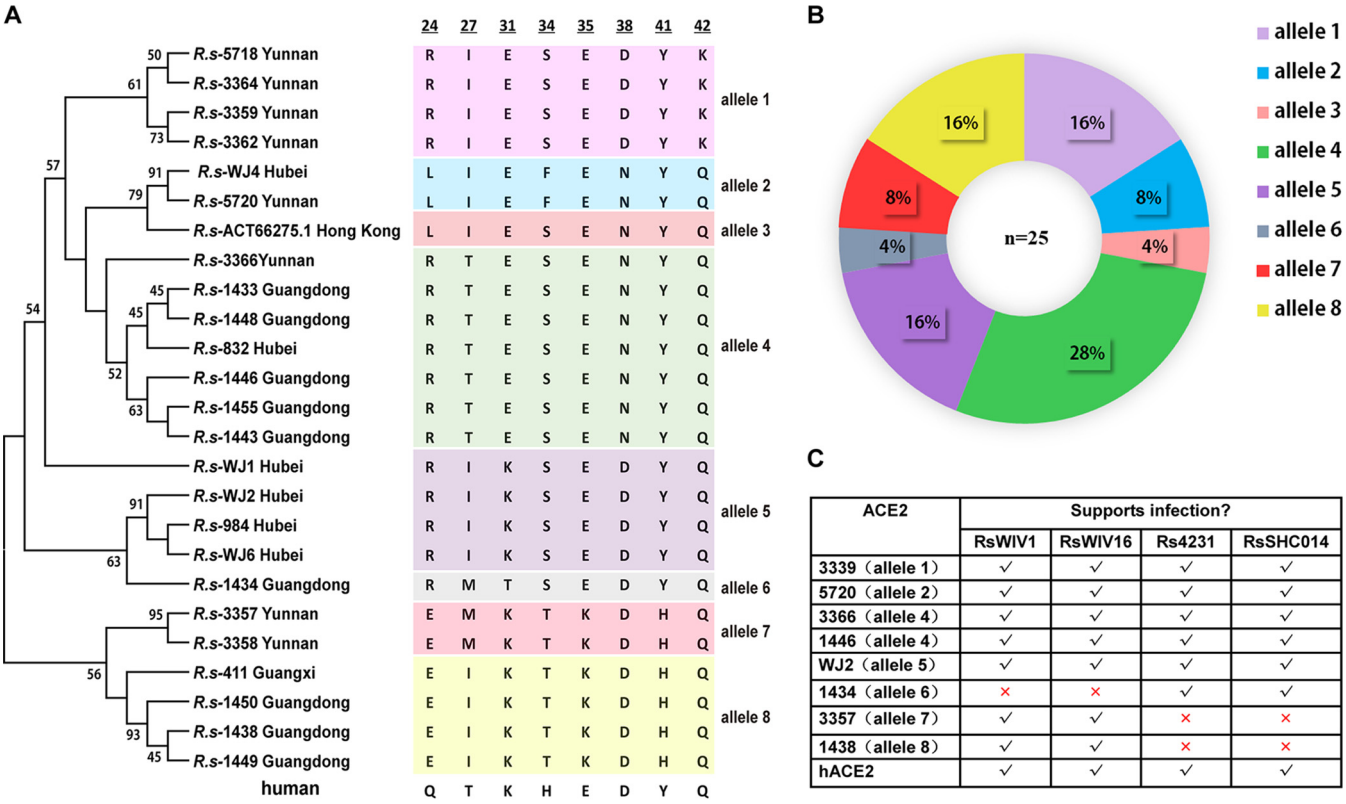
The SARS-CoV-2 outbreak has brought back memories of SARS-CoV that occurred 17 years ago in China. SARS is a zoonosis, as demonstrated by the identification of an almost identical coronavirus present in market civets in Guangdong province, China (8). In the following years, genetically diverse SARS-related coronaviruses (SARSr-CoV) were detected or isolated from horseshoe bats from different regions of China and Europe (9–18). Bat SARSr-CoVs share 78 to 96% nucleotide sequence identity with human and civet SARS-CoVs, with the most variable regions encoding the spike protein (S) and accessory proteins open reading frame 3 (ORF3) and ORF8 (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 receptors, 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 (receptor binding domain [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 (21–24).

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—a peptidase domain at the N terminus (amino acids [aa] 19 to 615) and a collectrin-like domain at the C terminus (aa 616 to 768) (25). The peptidase domain is involved in controlling vasoconstriction and blood pressure homeostasis (26–28). The collectrin-like domain participates in regulating the absorption of the intestinal neutral amino acid (29, 30), and the N-terminal domain is also involved in binding to the spike protein during SARS-CoV entry into the cells; however, they do not influence the normal peptidase activity (14, 16, 17, 31, 32). Previous results have indicated that the C-terminal domains of ACE2 from different origins are relatively well conserved, whereas the N-terminal domains show much more diversity across species (33). Previously, we have shown that SARS-CoV can utilize ACE2 derived from *Myotis daubentonii* and *R. sinicus*; minor mutations in the RBD-binding site could convert ACE2 from being unsusceptible to susceptible to SARS-CoV binding (34, 35). As all bat SARSr-CoVs belonging to clade 1 can utilize ACE2 and have been isolated from *R. sinicus*, we asked whether variations in ACE2 of *R. sinicus* could contribute to the diversity of bat SARSr-CoV.

Here, we have investigated the polymorphism of *R. sinicus* ACE2 genes and assessed



ACE2	Supports infection?			
	RsWIV1	RsWIV16	Rs4231	RsSHC014
3339（allele 1）	✓	✓	✓	✓
5720（allele 2）	✓	✓	✓	✓
3366（allele 4）	✓	✓	✓	✓
1446（allele 4）	✓	✓	✓	✓
WJ2（allele 5）	✓	✓	✓	✓
1434（allele 6）	×	×	✓	✓
3357（allele 7）	✓	✓	×	×
1438（allele 8）	✓	✓	×	×
hACE2	✓	✓	✓	✓

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 a bootstrap value (%) of 1,000 replicates (72). 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. ACE2 sequences of *Rs*-411, *Rs*-832, *Rs*-3357, and *Rs*-ACT66275 have been published previously and were downloaded from GenBank. (C) Analysis of different *R. sinicus* ACE2 usage of SARSr-CoVs determined by immunofluorescence assay. Determination of bat SARSr-CoV infectivity in HeLa cells with and without the expression of ACE2 from *R. sinicus* or human (hACE2) at a multiplicity of infection (MOI) of 1. The original immunofluorescence assay images are not shown.The cross indicates that the allele is not susceptible to bat SARSr-CoV.

their susceptibility and binding affinity to different bat SARSr-CoV spike proteins through a combination of molecular evolution analyses, protein affinity assays, and virus infection assays. Our results showed that the diversity of the SARSr-CoV spike protein may experience natural selection pressure from the *R. sinicus* ACE2 variants; over long periods of coexistence, the SARSr-CoV spike protein may be selected by *R. sinicus* ACE2 to maintain genetic diversity and to fit with the population of *R. sinicus*.

RESULTS

ACE2 genes show high polymorphism among the *R. sinicus* populations. Samples from three provinces (Hubei, Guangdong, and Yunnan) were used for ACE2 amplification, based on the prevalence of bat SARSr-CoVs and tissue sample availability and quality. In addition to previously sequenced bat ACE2 by our group (sample identifiers [IDs] 832, 411, and 3357, collected from Hubei, Guangxi, and Yunnan, respectively) and others (GenBank accession no. ACT66275; sample collected from Hong Kong), we obtained ACE2 gene sequences from 21 *R. sinicus* bat individuals five from Hubei, nine from Guangdong, and seven from Yunnan. The ACE2 sequences exhibited 98 to100% amino acid identity within their species and 80 to 81% amino acid identity with human ACE2 (see Table S1 in the supplemental material). Major variations were observed at the N-terminal region, including in some residues which were previously identified to be in contact with SARS-CoV S-RBD (Fig. 1A). Analysis based on nonsynonymous single-nucleotide polymorphisms (SNPs) helped identify eight residues, including residues 24, 27, 31, 34, 35, 38, 41, and 42. The combination of these 8

residues produced eight alleles, including RIESEDYK, LIEFENYQ, RTESENYQ, RIKSEDYQ, QIKSEDYQ, RMTSEDYQ, EMKTKDHQ, and EIKTKDHQ, named alleles 1 to 8, respectively (Fig. 1A). In addition to the ACE2 genotype data from previous studies (alleles 4, 7, and 8), five novel alleles were identified in the *R. sinicus* populations in this study. Alleles 2 and 4 were found in two and three provinces, respectively, whereas the other alleles seemed to be geographically restricted. In summary, three alleles (4, 6, and 8) were found in Guangdong, four (1, 2, 4, and 7) in Yunnan, three (2, 4, and 5) in Hubei, and one each in Guangxi and Hong Kong. Coexistence of four alleles was found in the same bat cave of Yunnan where the direct progenitor of SARS-CoV was found (Fig. 1B). Taken together, these data suggest that ACE2 variants have been circulating within the *R. sinicus* populations in different regions for a long time; substitutions at the sites directly in contact with SARS-CoV S-RBD suggest that they may have important functions during SARS-CoV evolution and transmission.

***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. For SARS-CoV, we used the pseudotyped virus carrying the S protein of strain BJ01 with HIV backbone, which was isolated from a human patient in the middle phase of 2002 to 2003, due to the biosafety regulations in China. 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 amino acid identity at the RBD, RsWIV16 is the closest relative of SARS-CoV and shows high amino acid identity at both the NTD and the RBD, Rs4231 shares high amino acid identity at the NTD, and RsSHC014 shows divergence at both the NTD and RBD regions.

Similarly 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. 1C and Fig. 2). All test viruses could efficiently use alleles 1, 2, 4, and 5 for entry. RsWIV1 and RsWIV16, which share an identical RBD, could not use allele 6 (sample ID 1434) from Guangdong (Fig. 1C and Fig. 2A and B). Rs4231 and RsSHC014, which share an identical RBD, could not use alleles 7 (sample ID 3357) and 8 (sample ID 1438) from Yunnan and Guangdong, respectively (Fig. 1C and Fig. 2C and D). 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. 2E and F). These results indicate that cell entry was affected by both spike RBD and *R. sinicus* ACE2 variants.

Mutation of *R. sinicus* ACE2 residues affects its binding affinity with SARSr-CoV RBDs. To further explain the different ability of SARS and SARSr-CoVs in ACE2 usage, we expressed the RBDs from SARS-CoV BJ01, RsWIV1, and RsSHC014 and three *R. sinicus* ACE2s, allele 6 (sample 1434), allele 7 (sample 3357), and allele 2 (sample 5720), and tested the binding affinity between them. BJ01 RBD with human ACE2 and RsWIV1 RBD with human DPP4 were used as the positive and negative controls, respectively. Biolayer interferometry results showed that the binding affinity between different RBDs and ACE2s was different based on the equilibrium dissociation constant (K_D) (Fig. 3 and Table 1). Consistent with the results of the virus infection experiments, 1434ACE2 (allele 6) was found to bind RsSHC014 and BJ01 but not RsWIV1 RBD; 3357ACE2 (allele 7) was found to bind RsWIV1 but not RsSHC014 and BJ01 RBD; 5720ACE2 (allele 2) was found to bind all tested RBDs. All tested RBDs had a higher binding affinity to human ACE2 than to that of bats. BJ01 RBD had a higher binding affinity to human ACE2 than did RsWIV1 and RsSHC014 RBDs (Table 1 and Fig. 3); however, it had a lower binding affinity to bat ACE2 than the two bat SARSr-CoV RBDs. These results demonstrated that the binding affinity between spike RBD and ACE2 affects the ability of the virus to infect.

Structure modeling of the interaction between SARSr-CoV RBDs and *R. sinicus* ACE2s. The four tested spike proteins of bat SARSr-CoV are identical in size and share

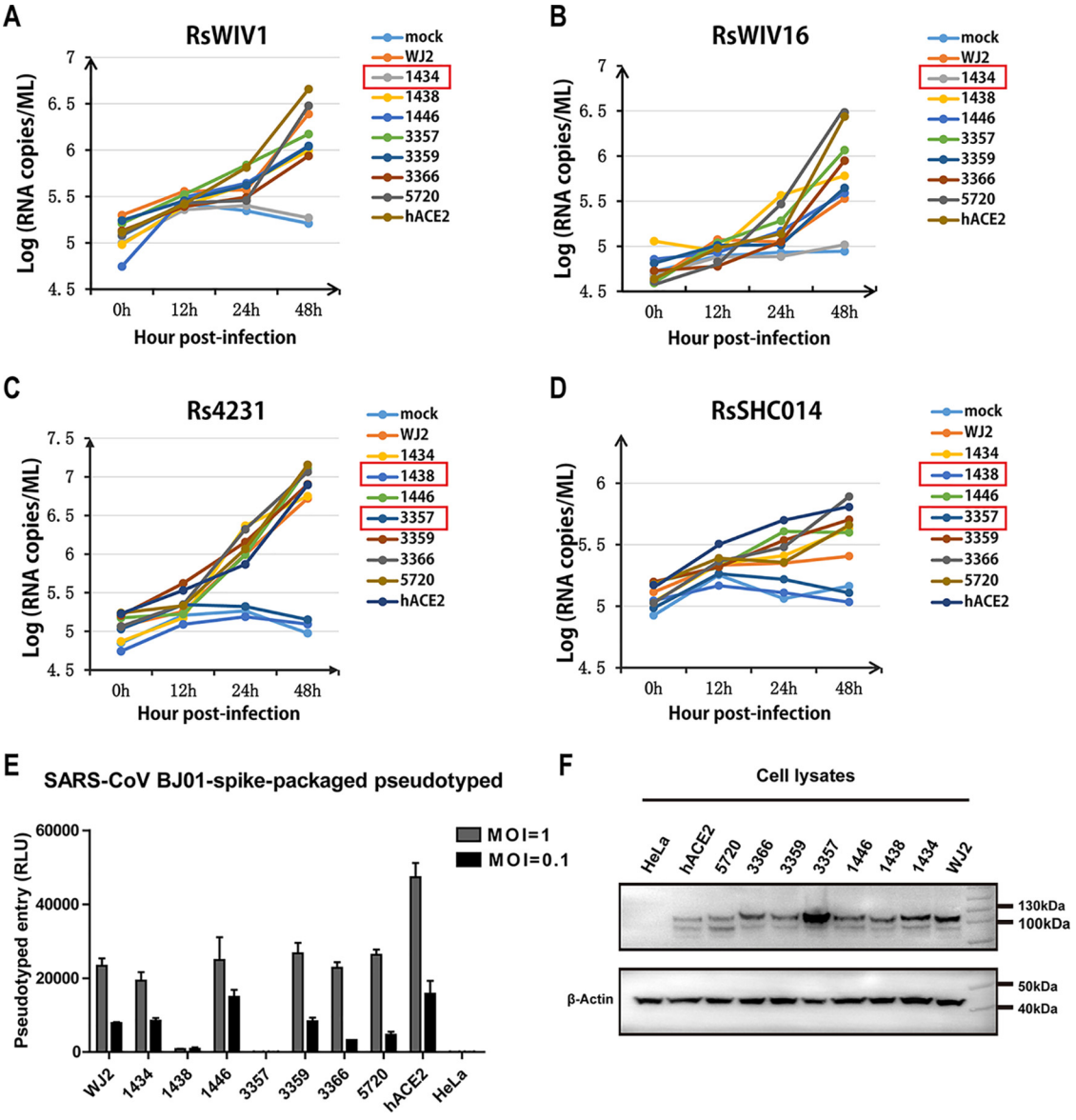


FIG 2 Infectivity of SARSr-CoV in HeLa cells expressing *R. sinicus* ACE2. (A to D) Determination of bat SARSr-CoV infectivity in HeLa cells with and without the expression of ACE2 from *R. sinicus* or human (hACE2) at an MOI of 0.01. The growth curves were determined by real-time PCR. Numbers in red square indicate samples that are not susceptible to bat SARSr-CoV. The values reported at each time point were averaged from two independent experiments. (E) The infectivity of the SARS-CoV-BJ01 pseudotyped was used for the assay at MOIs of 1 and 0.1 due to biosafety regulation in China. Error bars represent the standard error of the mean (SEM) from two independent transfections; each assay was performed in triplicate. (F) The expression of ACE2 in HeLa cells in SARS-CoV BJ01 pseudotyped infection assay. ACE2 expression was detected with mouse anti-S tag monoclonal antibody followed by horseradish peroxidase (HRP)-labeled goat anti-mouse IgG antibody. β-Actin was detected with mouse anti-β-action monoclonal antibody by HRP-labeled goat anti-mouse IgG antibody.

over 90% amino acid identity with SARS-CoV, which suggests that these proteins have a similar structure. In this study, we built structural complex models of bat SARSr-CoV-RsWIV1 RBD with *R. sinicus* ACE2 3357 (allele 7) and RsSHC014 RBD with *R. sinicus* ACE2 1434 (allele 6), in concordance with the results of the binding affinity assay between SARS-CoV RBD and human ACE2 (Fig. 4). As reported previously, there are two hot spots in ACE2 at the SARS-CoV RBD/human ACE2 binding interface. Hot spot K31 consists of a salt bridge with E35 and hot spot K353 consists of a salt bridge with D38 (Fig. 4A). The two hot spots are buried in a hydrophobic environment and provide a substantial amount of energy to the virus-receptor binding interactions, as well as filling critical voids in the hydrophobic stacking interactions at the binding interface (21, 24, 36).

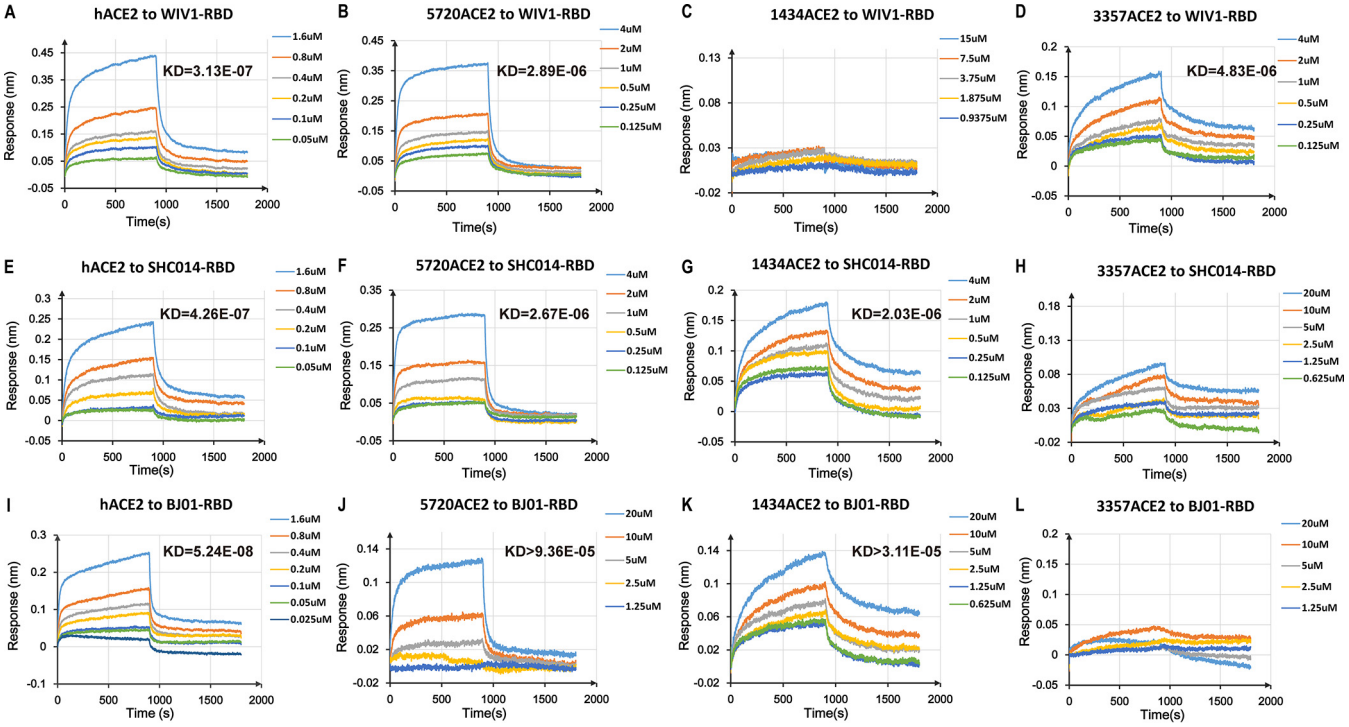


FIG 3 Binding affinity assay between different RBDs and ACE2s by biolayer interferometry. (A to D) Binding assay of human ACE2 or bat ACE2 to RsWIV1-RBD. (E to H) Binding assay of human ACE2 or bat ACE2 to RsSHC014-RBD. (I to L) Binding assay of human ACE2 or bat ACE2 to SARS-CoV BJ01-RBD. Binding assay of human DPP4 to WIV1 RBD was used as the negative control (data not shown). The parameters of the equilibrium dissociation constant (K_D) 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 human ACE2 and *R. sinicus* ACE2s. The *y* axis shows the real-time binding response. Values reported represent the global fit to the data. The coefficient of determination (R^2) for these interactions was close to 1.0. The RsSHC014-RBD/*R. sinicus* ACE2-3357 and BJ01-RBD/*R. sinicus* ACE2-3357 did not show obviously binding activity. For details of the kinetic analysis between different RBD and ACE2 proteins, see Table 1.

Compared with the contact residues at the SARS-CoV RBD/human ACE2 interface, changes in or near the two virus-binding hot spots were observed in *R. sinicus* ACE2 (Fig. 4D and E). The *R. sinicus* ACE2-3357 contains two main residue changes, E35K and Y41H (Fig. 4D). E35K breaks the salt bridge with K31 and R479 in RsSHC014 RBD, which may influence the binding affinity between them. A histidine at position 41 may weaken the support for the K353-D38 salt bridge because it is a weaker hydrophobic stacker than tyrosine, resulting in decreased binding affinity with BJ01 RBD (Fig. 4C) (37).

In *R. sinicus* ACE2-1434, we found a threonine at position 31, unlike human ACE2, which has a lysine at this position (Fig. 4A and E). Although the K31T change would fail to form a salt bridge with E35, Y442 in BJ01 RBD can form a hydrogen bond with it (Fig. 4A) (36). However, a serine at 442 in RsWIV1 RBD cannot. Besides, RsSHC014 contains an arginine at 479, the T31 cannot form a salt bridge with E35, making E35 available to form a salt bridge with R479 (Fig. 4E and F), but the RBD residue N479 in RsWIV1 may lose this ability (Fig. 4D).

In addition, all *R. sinicus* ACE2s in this study contain an asparagine at position 82 rather than a methionine in human ACE2 (Fig. 4B, D, and E). The M82N change introduces an unfavorable hydrophilic residue that would weaken the hydrophobic network around hot spot K31. Moreover, N82 introduces a glycosylation site, like that in rat ACE2, which cannot support SARS-CoV infection efficiently; the glycan at position 82 of ACE2 may lead to steric interference with viral RBD binding (37). Hence, M82N may have significant effects on the interaction between SARS-CoV and SARSr-CoV RBD with *R. sinicus* ACE2s. As previously described, residue 487 of the RBD interacts with hot spot K353 on ACE2. RsWIV1 contains an N487 in its RBD, and the polar side chain of N487 may have unfavorable interactions with the aliphatic portion of residue K353 in

TABLE 1 Kinetic analysis of different RBDs binding to human and *R. sinicus* ACE2s by biolayer interferometry[a]

	hACE2			5720ACE2 (allele 2)			1434ACE2 (allele 6)			3357ACE2 (allele 7)		
RBD	K_D (M)	K_{on} (1/M·s)	K_{dis} (1/s)	K_D (M)	K_{on} (1/ M·s)	K_{dis} (1/s)	K_D (M)	K_{on} (1/ M·s)	K_{dis} (1/s)	K_D (M)	K_{on} (1/ M·s)	K_{dis} (1/s)
WIV1	3.13E-07	1.03E+04 (±2.07E+02)	3.21E-03 (±2.10E-05)	2.89E-06	4.46E+03 (±1.10E+02)	1.29E-02 (±9.10E-05)	No obvious binding	No obvious binding	No obvious binding	4.83E-06	2.85E+03 (±4.67E+01)	1.38E-02 (±9.00E-06)
SHC014	4.26E-07	1.16E+04 (±2.42E+02)	4.95E-03 (±3.20E-05)	2.67E-06	6.51E+03 (±1.90E+02)	1.74E-02 (±1.56E-04)	2.03E-06	1.86E+03 (±6.17E+01)	3.78E-03 (±2.83E-05)	No obvious binding	No obvious binding	No obvious binding
BJ01	5.24E-08	5.26E+04 (±1.69E+03)	2.76E-03 (±2.18E-05)	9.36E-05	1.44E+04 (±1.85E+04)	1.35E+00 (±2.73E-01)	3.11E-05	9.95E+03 (±3.24E+04)	3.09E-01 (±3.10E-01)	No obvious binding	No obvious binding	No obvious binding

[a]Values reported representing the global fit to the data and the averages obtained from 6 or 5 replicates carried out with different ACE2 protein concentrations are shown in the upper right of Fig. 3. K_{on}, association rate constant; K_{dis}, dissociation rate constant.

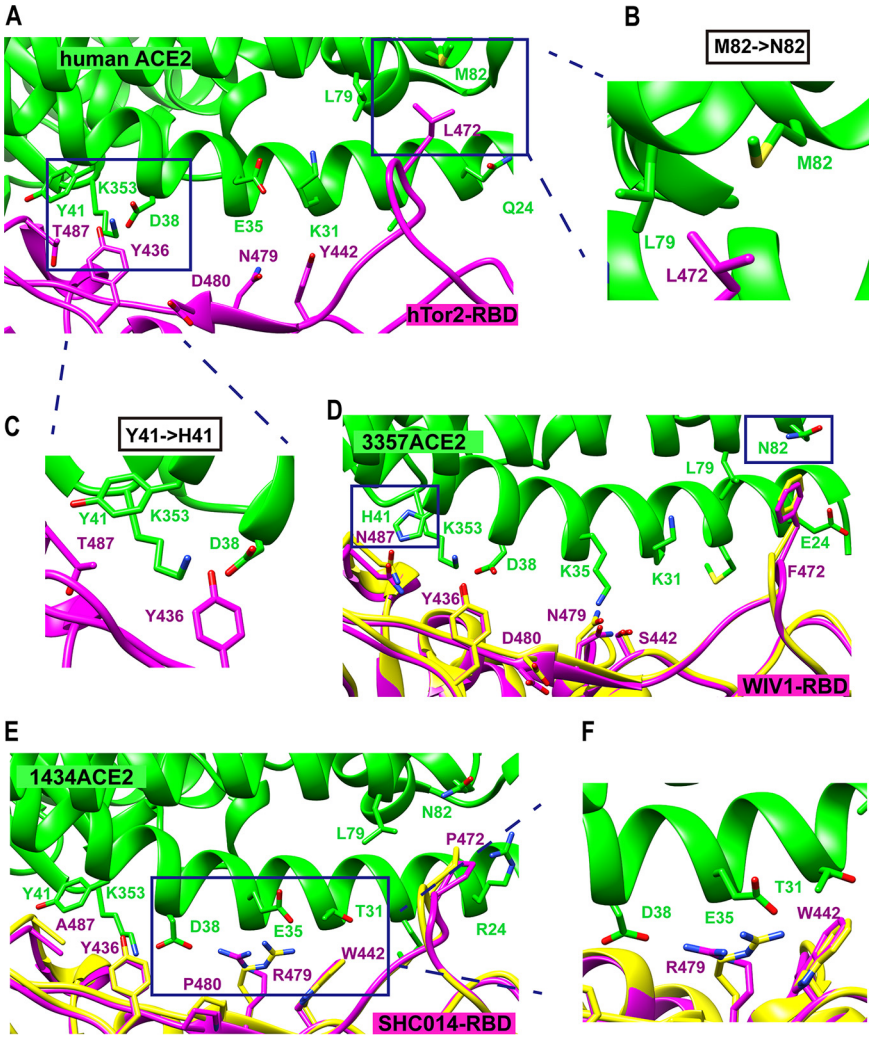


FIG 4 Structure modeling at the interface between SARSr-CoV spike and human or bat ACE2. Detailed view of the interaction between RBD and ACE2. Several important residues in RBD and ACE2 that are involved in the interactions are shown. ACE2 residues are in green and RBD residues are in magenta and yellow. (A) Structure complex of the SARS-CoV hTor2 RBD and human ACE2 (PDB code 2AJF). (B, C) Structural details of the interfaces between human ACE2 (M82, Y41) and SARS-CoV RBD. (D) Predicted structure complex of *R. sinicus* ACE2 3357 with RsWIV1 RBD. (E) Predicted structure complex of *R. sinicus* ACE2 1434 with RsSHC014 RBD. (F) Structural details of the interfaces between *R. sinicus* ACE2 1434 (T31, E35) and RsSHC014 RBD. The two predicted models (D, E) were built based on the structure of hTor2 RBD with hACE2 (yellow) and civet-optimized RBD with hACE2 (magenta) (PDB codes 2AJF and 3SCJ).

ACE2, which may affect the hot spot interaction between K353 and D38. Moreover, the RsSHC014 RBD contains an alanine at position 487, and the small side chain of A487 does not provide support to the structure of hot spot K353 (19). Therefore, these residue changes in SARSr-CoV RBD and *R. sinicus* ACE2 may account for the different binding affinities between them, which correlates well with the results from virus infection and binding assays.

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 hypothesis testing using phylogenies (HyPhy) package to analyze the ratio of nonsynonymous to synonymous mutations (dN/dS ratio) at individual codons (38, 39).

By using four test models (fixed effects likelihood [FEL], mixed-effects model of evolution [MEME], fast-unconstrained Bayesian approximation [FUBAR], and single-likelihood ancestor counting [SLAC]), we analyzed the 25 *R. sinicus* ACE2 gene se-

TABLE 2 Positive selection analysis of *R. sinicus* ACE2 and SARSr-CoV spike sequences by HyPhy

Test models	Dataset [a] (Codons under positive selection [b])	
	R.sinicus ACE2 [e]	SARSr-CoVs Spike [f]
FEL (p-value <0.2) [c]	19* (0.1777), 24* (0.0345), 27* (0.1416), 31* (0.1356), 34* (0.099), 38* (0.0483), 84* (0.1364), 139* (0.1459), 159* (0.0234)	418* (0.0607), 480* (0.1735), 488* (0.1586)
MEME (p-value <0.2) [c]	19* (0.1994), 24* (0.0145), 27* (0.1787), 31* (0.1614), 34* (0.1044), 38* (0.0989), 84* (0.1621), 139* (0.1737), 159* (0.0706)	14 (0.1334), 26 (0.1691), 38 (0.1220), 85 (0.1327), 146 (0.1787), 152 (0.1264), 170 (0.0697), 199 (0.0648), 245 (0.1752), 355 (0.0454), 418* (0.0359), 427 (0.1473), 433 (0.0532), 440 (0.1199), 447 (0.0467), 450 (0.0972), 459 (0.0817), 480* (0.1158), 488* (0.1039), 557 (0.0097), 623 (0.0117), 1167 (0.0105), 1244 (0.0118)
FUBAR (Prob >0.9) [d]	5 (0.949), 24* (0.989), 27* (0.971), 31* (0.967), 34* (0.975), 35 (0.941), 38* (0.983), 41 (0.949), 42 (0.946), 84* (0.978), 139* (0.925), 159* (0.982)	418* (0.9343), 488* (0.9076)

[a]Data set consisted of 25 *R. sinicus* ACE2 and 9 SARSr-CoV spike sequences.
[b]Codons with an asterisk (*) are those identified by more than one model.
[c]Likelihood ratio tests were used to ascertain the ratio of nonsynonymous to synonymous mutations (dN/dS) in the fixed effects likelihood (FEL) and mixed-effects model of evolution (MEME) models. The *P* value given in parentheses behind the codons indicates that the evidence of selection with confidence in the model can be rejected. *P* values of <0.2 are shown.
[d]The posterior probability of positive selection at a site ($P > 0.9$) in the fast-unconstrained Bayesian approximation (FUBAR) model.
[e]Codons in orange are those identified in a given model which interact with spike directly.
[f]Codons in blue are those identified in a given model which are located at the RBD region of SARSr-CoV spike.

quences obtained in this study and downloaded from GenBank. We found 13 codons (Table 2) that were under positive selection in at least one model, including nine codons in models FEL and MEME ($P < 0.2$) and 12 in model FUBAR (posterior probabilities of >0.9), while the SLAC model found no evidence for sites under positive selection with a *P* value of <0.1 (data not shown). Among the 13 codons, 8 of them (codons 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). We also analyzed the ACE2 gene of *Rhinolophus affinis*, which has been reported to carry SARSr-CoV occasionally (15). Due to the more diverse *R. sinicus* ACE2 alleles found in Yunnan, we amplified 22 *R. affinis* ACE2 sequences from Yunnan and one from Hubei. Used an alignment of the 23 *R. affinis* ACE2 gene sequences, we found that *R. affinis* ACE2 was more conserved between different individuals in the entire coding region than *R. sinicus* ACE2, 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 throughout 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 (frequencies of 0.00001 and 0.00002, respectively).

We next used the four test models to analyze the complete gene encoding the SARSr-CoV spike protein by aligning nine SARSr-CoV spike sequences from *R. sinicus* samples. We found that two codons, 480 and 488, which correspond to residues 479 and 487 in the spike protein of SARS-CoV, were under positive selection in models FEL and MEME ($P < 0.2$) (Table 2). These two residues were identified previously to be critical for receptor binding. We also found that codon 418, which in the RBD region, was under positive selection in three models, and that there are also some codons in the RBD region or the other regions of the spike gene under positive selection in the

MEME model, but none in the SLAC model (Table 2). These results indicate that positive selection has happened both on *R. sinicus* ACE2 and bat SARSr-CoV spike protein and that the positively selected sites that localize at the interface between them imply they may have experienced selection from each other.

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 the 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 that of 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* specimens collected from four provinces and one city, and we 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 coevolved 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 the 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 physical function (40, 41). In this study, our analysis of the SARSr-CoV spike protein and *R. sinicus* ACE2 gene sequences showed that some codons related to the interaction interface between them were under positive selection. Similar rapid adaptation for SARS-CoV occurred during the SARS outbreak in 2002 and 2003 (21, 42). When SARS-CoV was transmitted from market civet to human, the spike gene experienced positive selection, in which mutations in two critical residues (amino acids 479 and 487) of the spike protein changed the binding affinity of the virus to human ACE2 from low to high, subsequently turning it into a pandemic strain (42).

An increasing number of studies have suggested that coevolution is driven by specific interactions between high levels of virus sequence divergence and polymorphic host receptors (40, 41, 43–46). The first example was observed for the avian leucosis virus infection, in which receptor polymorphism in chicken could alter the sensitivity to virus entry (44). Arms races can also occur between the glycoproteins of two different viruses and the same receptor, as seen for the receptor transferrin receptor 1 (TfR1) of rodent arenavirus and the mouse mammary tumor virus (47).

The factors in a virus successfully infecting a host will depend on its abilities to bind to the receptor of host cells, to replicate in suitable tissues, and to avoid or suppress the host immune response (48). Poor receptor binding sometimes leads to inefficient transmission for a virus in the host. This situation recently has been shown by the ongoing COVID-19 pandemic. The pathogen that causes COVID-19 is SARS-CoV-2; some studies showed that it uses the same ACE2 receptor as SARS-CoV (7, 49). However, the RBD protein of SARS-CoV-2 has higher human ACE2 binding affinity than that of SARS-CoV, suggesting its efficient cell entry, which could further explain the rapid transmission of SARS-CoV-2 in humans (49, 50). In this study, we found that bat SARSr-CoV RBDs showed a higher binding affinity to human ACE2 than to *R. sinicus* ACE2s but did not display obvious differences among the *R. sinicus* ACE2s. Our limited data on *R. sinicus* ACE2 sequences cannot exclude the possibility that some ACE2s in *R. sinicus* population have a higher binding affinity to SARSr-CoV or SARS-CoV. The higher binding affinity of bat SARSr-CoV RBD to human ACE2 suggests they have the potential to infect humans cross-species. Similar examples can be found in other cases of

coevolution between host receptor and virus (44, 47, 51, 52). Previous studies have shown the adaptive evolution of the coronavirus spike protein to host receptors *in vivo* and *in vitro* (30, 53–56). Recently, the MERS-CoV spike protein was shown to rapidly adapt to the *Desmodus rotundus* DPP4 (*dr*DPP4) receptor from a semipermissive to a 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 (57). 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 (58–61). 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 potential spillover from viruses circulating in animal populations through 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 (26, 62). 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, 63). 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, 64). 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. The 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 cDNA was synthesized from total RNA by reverse transcription with random hexamers; full-length bat ACE2 fragments were amplified by reverse transcription0nested PCR (RT-PCR). Primers were designed based on available ACE2 sequences from NCBI. The first-round primers wereF-ACE2-out-AATGGGGTTTTGGCGCTCAG and R-ACE2-out-CATACAATGAAATCACCTCAAGAG, and the second-round primers were F-ACE2-in-CAGGGAAAGATG TCAGGCTC and R-ACE2-in-TTCTAAAABGAVGTYTGAAC. The ACE2 gene was cloned into the pcDNA3.1 vector with XhoI and BamHI. N-terminal mouse Igk or human interferon alpha-1 (IFN-α1) signal peptide and an 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 to 569; GenBank accession number AY278488.2), RsWIV1 RBD (spike residues: aa 318 to 570; accession number KF367457.1), RsSHC014 RBD (spike residues: aa 318 to 570; accession number KC881005.1), *R. sinicus* ACE2 (aa 19 to 615), human ACE2 (aa 19 to 615, accession number BAJ21180), and human DPP4 (aa 39 to 766, accession number NP_001926) into the expression vector, as described previously (65). For each protein, an N-terminal signal peptide and an S tag were added to facilitate protein secretion and purification. The proteins used for the Octet RED binding assay were expressed in HEK 293T/17 cells. Cells were washed twice with D-Hanks solution at 6 h posttransfection, then cultured in fresh 293T FreeStyle expression medium. The cells were collected at 48 h posttransfection and centrifuged at 4,000 $\times$ *g* for 10 min at 4°C. The supernatant was incubated with S tag agarose beads overnight at 4°C and passed through a column, and the protein was eluted from the column with 3 M $MgCl_2$. The protein was finally buffered with phosphate-buffered saline (PBS) and stored at −80°C until further use.

Immunofluorescence assay. HeLa cells were seeded in 24-well plates and transfected with equal amounts of human or *R. sinicus* ACE2 plasmids. After 24 h, the cells were incubated with different SARSr-CoV strains at a multiplicity of infection (MOI) of 1 for 1 h. Thereafter, the cells were washed with D-Hanks solution thrice and supplemented with the medium. At 24 h after infection, cells were washed

with PBS and fixed with 4% formaldehyde for 30 min at room temperature (about 25°C). ACE2 expression was detected using mouse anti-S tag monoclonal antibody (Kyab Biotech Co., Ltd., Wuhan, China), followed by fluorescein isothiocyanate (FITC)-labeled goat anti-mouse IgG H&L (Abcam, 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).

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 of 0.01 with 1 h of adsorption. Thereafter, the cells were washed with D-Hanks solution three times and cultured in 1 ml DMEM plus 2% fetal bovine serum (FBS). The viral supernatants were harvested at 0, 12, 24, and 48 h postinfection—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 minikit (Qiagen, Waltham, MA) and then quantified on the ABI Step One system (Thermo Fisher Scientific), with the AgPath-ID one-step RT-PCR kit (Applied Biosystems/Life Technologies, Waltham, MA). RNA dilutions from purified SARSr-CoV-RsWIV1 stock were used as a standard (with a titer of 6.5×10^6 PFU/ml; about 2.34×10^9 genome RNA copies/ml). Every sample was analyzed in triplicate on two independent occasions.

Biolayer interferometry. 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) in 96-well microplates at 30°C, as described previously (66, 67). Briefly, the RBD was biotinylated using EZ-Link NHS-LC-LC-biotin [succinimidyl-6-(biotinamido)-6-hexanamido hexanoate; Thermo Fisher Scientific]. The assay was carried out by placing the streptavidin biosensors (ForteBio) 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% bovine serum albumin [BSA], and 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 proteins 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 1,000 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. Interstep correction and *y* alignment were used to minimize tip-dependent variability. Data were globally fitted in a 1:1 model using Data Analysis Software v7.1 (ForteBio). The mean association rate constant (K_{on}) and dissociation rate constant (K_{dis}) values were determined with a global fit applied to all data. The coefficient of determination (R^2) for these interactions was close to 1.0 ($>$0.9).

Pseudotyped production and infection of ACE2-transfected HeLa cells. SARS-CoV-BJ01 pseudotyped particles were produced by HEK 293T/17 cells. The plasmids pcDNA3.1-BJ01-S and pNL4-3.luc.R-E- were cotransfected into HEK 293T/17 cells. At 6 h posttransfection, the medium was replaced with fresh 293T FreeStyle expression medium. The supernatant, which contained the pseudotyped particles, was harvested 48 h posttransfection and filtered using 0.45-μm filters. Thereafter, the pseudotyped particles were aliquoted and stored at −80°C until further use. The viral titer was determined by the HIV p24 quantitation enzyme-limited immunosorbent assay (ELISA) kits (Kyab Biotech Co., Ltd.) before the viruses were used for the infection assay. HeLa cells were transfected with the same amount of human or *R. sinicus* ACE2 plasmids. After 24 h, the cells were incubated with SARS-CoV-BJ01 pseudotyped at an MOI of 1 or 0.1 for 1 h at 37°C, washed twice with D-Hanks solution, and supplemented with DMEM containing 10% FBS. Luciferase activity was determined using a GloMax luminometer (Promega Biotech Co., Ltd., Beijing, China) 48 h after infection. Each sample was analyzed in triplicate on two independent occasions.

Structure modeling. The structure complex of SARS-CoV/SARSr-CoVs RBD and *R. sinicus* ACE2 was homology modeled using SWISS-MODEL (https://swissmodel.expasy.org/), based on the structure of human SARS-CoV RBD (hTor2 RBD) complexed with human ACE2 (hACE2) and civet-optimized RBD complexed with hACE2 (PDB codes 2AJF and 3SCJ) (36). Molecular graphics visualization and analyses were performed using the UCSF Chimera software (http://www.rbvi.ucsf.edu/chimera).

Codon-based analysis of molecular evolution. Bat ACE2 and SARSr-CoV spike sequences were analyzed for positive selection. In this study, bat ACE2 sequences were either amplified or downloaded from the NCBI database, and SARSr-CoV spike sequences were downloaded from NCBI. Sequences were aligned in Clustal X. Phylogenetic trees were built by the maximum likelihood method implemented in RAxML program in CIPRES Science Gateway (https://www.phylo.org/). Codon-based analysis of positive selection was performed using the hypothesis testing using phylogenies (HyPhy) package version 2.5.14 (MP) (38, 39). In brief, four test models, namely fixed effects likelihood (FEL), fast-unconstrained Bayesian approximation (FUBAR), mixed-effects model of evolution (MEME), and single-likelihood ancestor counting (SLAC) were used. For SARSr-CoV spike genes, we run the genetic algorithm for recombination detection (GARD) model to detect the potential recombination before selection analyses. The output data set from GARD was used as the input for the models in subsequent positive selection analyses (68).

The models FEL and MEME use likelihood ratio tests (LRTs) to assess a better fit of codons that allowed positive selection ($P < 0.2$), while the SLAC model used an extended binomial distribution to ascertain the positive selection at each site ($P < 0.1$), and the FUBAR model employs a Bayesian algorithm to infer rates when posterior probabilities > 0.9 are generally suggestive of positive selection (69–71).

Data availability. The complete ACE2 sequences of *R. sinicus* and *R. affinis* ACE2 obtained in this study have been deposited in the GenBank database under accession numbers MT394181 to MT394201 and MT394203 to MT394225, respectively. ACE2 sequences of Rs-411 (GenBank accession number GQ999933), Rs-832 (GQ999936), Rs-3357 (KC881004), and Rs-ACT66275 have been published previously. The sequences of SARSr-CoVs spike detected from *R. sinicus* have been published previously as SARSr-CoV RsWIV16 (KT444582), RsWIV1 (KF367457), Rs3367 (KC881006), Rs4084 (KY417144), RsSHC014 (KC881005), Rs7327 (KY417151), Rs4231 (KY417146), Rs4874 (KY417150), and Rs9401 (KY417152).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

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

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

We declare no competing interests.

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.O., and Z.-L.S. wrote the paper.

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. https://doi.org/10.1128/JVI.06540-11.
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. https://doi.org/10.1016/j.tim.2016.03.003.
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 JSM, Guan Y. 2003. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet 362:1353–1358. https://doi.org/10.1016/S0140-6736(03)14630-2.
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 348:1967–1976. https://doi.org/10.1056/NEJMoa030747.
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, SARS Working Group. 2003. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 348:1953–1966. https://doi.org/10.1056/NEJMoa030781.
6. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus A, Fouchier R. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 367:1814–1820. https://doi.org/10.1056/NEJMoa1211721.
7. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng X-S, Zhao K, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G-F, Shi Z-L. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273. https://doi.org/10.1038/s41586-020-2012-7.
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 302:276–278. https://doi.org/10.1126/science.1087139.
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 102:14040–14045. https://doi.org/10.1073/pnas.0506735102.
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 310:676–679. https://doi.org/10.1126/science.1118391.
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 91:1058–1062. https://doi.org/10.1099/vir.0.016378-0.
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 11:1690–1702. https://doi.org/10.1016/j.meegid.2011.06.021.
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, Herrler G, Drosten C. 2010. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. J Virol 84:11336–11349. https://doi.org/10.1128/JVI.00650-10.
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 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. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503:535–538. https://doi.org/10.1038/nature12711.
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. 2014. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory

syndrome-like coronavirus from bats in China. J Virol 88:7070–7082. https://doi.org/10.1128/JVI.00631-14.
16. Yang XL, Hu B, Wang B, Wang MN, Zhang Q, Zhang W, Wu LJ, Ge XY, Zhang YZ, Daszak P, Wang LF, Shi ZL. 2015. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. J Virol 90:3253–3256. https://doi.org/10.1128/JVI.02582-15.
17. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ, Wang N, Luo DS, Zheng XS, Wang MN, Daszak P, Wang LF, Cui J, Shi ZL. 2017. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog 13:e1006698. https://doi.org/10.1371/journal.ppat.1006698.
18. Wacharapluesadee S, Duengkae P, Rodpan A, Kaewpom T, Maneeorn P, Kanchanasaka B, Yingsakmongkon S, Sittidetboripat N, Chareesaen C, Khlangsap N, Pidthong A, Leadprathom K, Ghai S, Epstein JH, Daszak P, Olival KJ, Blair PJ, Callahan MV, Hemachudha T. 2015. Diversity of coronavirus in bats from Eastern Thailand. Virol J 12:57. https://doi.org/10.1186/s12985-015-0289-1.
19. Cui J, Li F, Shi ZL. 2019. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 17:181–192. https://doi.org/10.1038/s41579-018-0118-9.
20. Li F. 2012. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. J Virol 86:2856–2858. https://doi.org/10.1128/JVI.06882-11.
21. Song H-D, Tu C-C, Zhang G-W, Wang S-Y, Zheng K, Lei L-C, Chen Q-X, Gao Y-W, Zhou H-Q, Xiang H, Zheng H-J, Chern S-WW, Cheng F, Pan C-M, Xuan H, Chen S-J, Luo H-M, Zhou D-H, Liu Y-F, He J-F, Qin P-Z, Li L-H, Ren Y-Q, Liang W-J, Yu Y-D, Anderson L, Wang M, Xu R-H, Wu X-W, Zheng H-Y, Chen J-D, Liang G, Gao Y, Liao M, Fang L, Jiang L-Y, Li H, Chen F, Di B, He L-J, Lin J-Y, Tong S, Kong X, Du L, Hao P, Tang H, Bernini A, Yu X-J, Spiga O, Guo Z-M, Pan H-Y, He W-Z, Manuguerra J-C, Fontanet A, Danchin A, Niccolai N, Li Y-X, Wu C-I, Zhao G-P. 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc Natl Acad Sci U S A 102:2430–2435. https://doi.org/10.1073/pnas.0409608102.
22. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong SK, Huang IC, Xu K, Vasilieva N, Murakami A, He Y, Marasco WA, Guan Y, Choe H, Farzan M. 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J 24:1634–1643. https://doi.org/10.1038/sj.emboj.7600640.
23. Li F, Berardi M, Li W, Farzan M, Dormitzer PR, Harrison SC. 2006. Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. J Virol 80:6794–6800. https://doi.org/10.1128/JVI.02744-05.
24. Li F, Li W, Farzan M, Harrison SC. 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309:1864–1868. https://doi.org/10.1126/science.1116480.
25. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 367:1444–1448. https://doi.org/10.1126/science.abb2762.
26. 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 417:822–828. https://doi.org/10.1038/nature00786.
27. Turner AJ. 2015. ACE2 cell biology, regulation, and physiological functions, p 185–189. *In* Unger T, Steckelings UM, dos Santos RAS (ed), The protective arm of the renin angiotensin system (RAS). Academic Press, Boston, MA.
28. Donoghue M, Hsieh F, Baronas E, Godbout K, Gosselin M, Stagliano N, Donovan M, Woolf B, Robison K, Jeyaseelan R, Breitbart RE, Acton S. 2000. A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9. Circ Res 87:e1–e9. https://doi.org/10.1161/01.res.87.5.e1.
29. Camargo SMR, Singer D, Makrides V, Huggel K, Pos KM, Wagner CA, Kuba K, Danilczyk U, Skovby F, Kleta R, Penninger JM, Verrey F. 2009. Tissue-specific amino acid transporter partners ACE2 and collectrin differentially interact with hartnup mutations. Gastroenterology 136:872–882. https://doi.org/10.1053/j.gastro.2008.10.055.
30. Kuba K, Imai Y, Penninger JM. 2013. Multiple functions of angiotensin-converting enzyme 2 and its relevance in cardiovascular diseases. Circ J 77:301–308. https://doi.org/10.1253/circj.cj-12-1544.
31. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426:450–454. https://doi.org/10.1038/nature02145.
32. Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pöhlmann S. 2005. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. Proc Natl Acad Sci U S A 102:7988–7993. https://doi.org/10.1073/pnas.0409465102.
33. Demogines A, Farzan M, Sawyer SL. 2012. Evidence for ACE2-utilizing coronaviruses (CoVs) related to severe acute respiratory syndrome CoV in bats. J Virol 86:6350–6353. https://doi.org/10.1128/JVI.00311-12.
34. Hou Y, Peng C, Yu M, Li Y, Han Z, Li F, Wang LF, Shi Z. 2010. Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. Arch Virol 155:1563–1569. https://doi.org/10.1007/s00705-010-0729-6.
35. Yu M, Tachedjian M, Crameri G, Shi Z, Wang LF. 2010. Identification of key amino acid residues required for horseshoe bat angiotensin-I converting enzyme 2 to function as a receptor for severe acute respiratory syndrome coronavirus. J Gen Virol 91:1708–1712. https://doi.org/10.1099/vir.0.020172-0.
36. Wu K, G P, Wilken M, Geraghty RJ, Li F. 2012. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. J Biol Chem 287:8904–8911. https://doi.org/10.1074/jbc.M111.325803.
37. Li F. 2013. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Res 100:246–254. https://doi.org/10.1016/j.antiviral.2013.08.014.
38. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, Wisotsky S, Spielman SJ, Frost SDW, Muse SV. 2020. HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol 37:295–299. https://doi.org/10.1093/molbev/msz197.
39. Pond SLK, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies, p 125–181, Statistical methods in molecular evolution. Springer-Verlag, New York, NY.
40. Daugherty MD, Malik HS. 2012. Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet 46:677–700. https://doi.org/10.1146/annurev-genet-110711-155522.
41. Meyerson NR, Sawyer SL. 2011. Two-stepping through time: mammals and viruses. Trends Microbiol 19:286–294. https://doi.org/10.1016/j.tim.2011.03.006.
42. Chinese S. 2004. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 303:1666–1669.
43. Gupta RK, Hue S, Schaller T, Verschoor E, Pillay D, Towers GJ. 2009. Mutation of a single residue renders human tetherin resistant to HIV-1 Vpu-mediated depletion. PLoS Pathog 5:e1000443. https://doi.org/10.1371/journal.ppat.1000443.
44. Crittenden LB, Stone HA, Reamer RH, Okazaki W. 1967. Two loci controlling genetic cellular resistance to avian leukosis-sarcoma viruses. J Virol 1:898–904. https://doi.org/10.1128/JVI.1.5.898-904.1967.
45. Warren CJ, Meyerson NR, Stabell AC, Fattor WT, Wilkerson GK, Sawyer SL. 2019. A glycan shield on chimpanzee CD4 protects against infection by primate lentiviruses (HIV/SIV). Proc Natl Acad Sci U S A 116:11460–11469. https://doi.org/10.1073/pnas.1813909116.
46. Bibollet-Ruche F, Russell RM, Liu W, Stewart-Jones GBE, Sherrill-Mix S, Li Y, Learn GH, Smith AG, Gondim MVP, Plenderleith LJ, Decker JM, Easlick JL, Wetzel KS, Collman RG, Ding S, Finzi A, Ayouba A, Peeters M, Leendertz FH, van Schijndel J, Goedmakers A, Ton E, Boesch C, Kuehl H, Arandjelovic M, Dieguez P, Murai M, Colin C, Koops K, Speede S, Gonder MK, Muller MN, Sanz CM, Morgan DB, Atencia R, Cox D, Piel AK, Stewart FA, Ndjango J-B, Mjungu D, Lonsdorf EV, Pusey AE, Kwong PD, Sharp PM, Shaw GM, Hahn BH. 2019. CD4 receptor diversity in chimpanzees protects against SIV infection. Proc Natl Acad Sci U S A 116:3229–3238. https://doi.org/10.1073/pnas.1821197116.
47. Demogines A, Abraham J, Choe H, Farzan M, Sawyer SL. 2013. Dual host-virus arms races shape an essential housekeeping protein. PLoS Biol 11:e1001571. https://doi.org/10.1371/journal.pbio.1001571.
48. Longdon B, Hadfield JD, Day JP, Smith SCL, McGonigle JE, Cogni R, Cao C, Jiggins FM. 2015. The causes and consequences of changes in virulence following pathogen host shifts. PLoS Pathog 11:e1004728. https://doi.org/10.1371/journal.ppat.1004728.
49. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 181:281–292.e6. https://doi.org/10.1016/j.cell.2020.02.058.
50. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. 2020. Structural basis of receptor recognition by SARS-CoV-2. Nature 581:221–224. https://doi.org/10.1038/s41586-020-2179-y.

51. Adkins HB, Brojatsch J, Naughton J, Rolls MM, Pesola JM, Young JA. 1997. Identification of a cellular receptor for subgroup E avian leukosis virus. Proc Natl Acad Sci U S A 94:11617–11622. https://doi.org/10.1073/pnas.94.21.11617.
52. O'Brien SJ, Nelson GW. 2004. Human genes that limit AIDS. Nat Genet 36:565–574. https://doi.org/10.1038/ng1369.
53. Baric RS, Sullivan E, Hensley L, Yount B, Chen W. 1999. Persistent infection promotes cross-species transmissibility of mouse hepatitis virus. J Virol 73:638–649. https://doi.org/10.1128/JVI.73.1.638-649.1999.
54. 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. https://doi.org/10.1371/journal.ppat.0030005.
55. 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 82:1414–1424. https://doi.org/10.1128/JVI.01674-07.
56. 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 82:2274–2285. https://doi.org/10.1128/JVI.02041-07.
57. 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. https://doi.org/10.1016/j.celrep.2018.07.045.
58. Liu C, Tang J, Ma Y, Liang X, Yang Y, Peng G, Qi Q, Jiang S, Li J, Du L, Li F. 2015. Receptor usage and cell entry of porcine epidemic diarrhea coronavirus. J Virol 89:6121–6125. https://doi.org/10.1128/JVI.00430-15.
59. 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. https://doi.org/10.1016/j.cell.2016.10.013.
60. 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. https://doi.org/10.1016/j.cell.2016.10.014.
61. 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. https://doi.org/10.1038/s41564-018-0328-y.
62. 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. https://doi.org/10.1016/j.pharmthera.2010.06.003.
63. 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 78:10628–10635. https://doi.org/10.1128/JVI.78.19.10628-10635.2004.
64. 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. https://doi.org/10.1128/JVI.03079-15.
65. 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 16:328–337. https://doi.org/10.1016/j.chom.2014.08.009.
66. 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. https://doi.org/10.1038/srep27996.
67. 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. https://doi.org/10.1371/journal.pone.0159704.
68. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost S. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891–1901. https://doi.org/10.1093/molbev/msl051.
69. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: a Fast, Unconstrained Bayesian AppRoximation for inferring selection. Mol Biol Evol 30:1196–1205. https://doi.org/10.1093/molbev/mst030.
70. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8:e1002764. https://doi.org/10.1371/journal.pgen.1002764.
71. Kosakovsky Pond SL, Frost S. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208–1222. https://doi.org/10.1093/molbev/msi105.
72. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197.