Emerging Microbes & Infections
2021, VOL. 10
https://doi.org/10.1080/22221751.2021.1956373

Identification of a novel lineage bat SARS-related coronaviruses that use bat ACE2 receptor

Hua Guo[a,b]*, Ben Hu[a]*, Hao-Rui Si[a,b]*, Yan Zhu[a], Wei Zhang[a], Bei Li[a], Ang Li[a,b], Rong Geng[a,b], Hao-Feng Lin[a,b], Xing-Lou Yang[a], Peng Zhou [a] and Zheng-Li Shi [a]

[a]CAS Key Laboratory of Special Pathogens and Biosafety, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan Institute of Virology, Wuhan, People's Republic of China; [b]University of Chinese Academy of Sciences, Beijing, People's Republic of China

ABSTRACT

Severe respiratory disease coronavirus-2 (SARS-CoV-2) has been the most devastating disease COVID-19 in the century. One of the unsolved scientific questions of SARS-CoV-2 is the animal origin of this virus. Bats and pangolins are recognized as the most probable reservoir hosts that harbour highly similar SARS-CoV-2 related viruses (SARSr-CoV-2). This study identified a novel lineage of SARSr-CoVs, including RaTG15 and seven other viruses, from bats at the same location where we found RaTG13 in 2015. Although RaTG15 and the related viruses share 97.2% amino acid sequence identities with SARS-CoV-2 in the conserved ORF1b region, it only shows less than 77.6% nucleotide identity to all known SARSr-CoVs at the genome level, thus forming a distinct lineage in the *Sarbecovirus* phylogenetic tree. We found that the RaTG15 receptor-binding domain (RBD) can bind to ACE2 from *Rhinolophus affinis*, Malayan pangolin, and use it as an entry receptor, except for ACE2 from humans. However, it contains a short deletion and has different key residues responsible for ACE2 binding. In addition, we showed that none of the known viruses in bat SARSr-CoV-2 lineage discovered uses human ACE2 as efficiently as the pangolin-derived SARSr-CoV-2 or some viruses in the SARSr-CoV-1 lineage. Therefore, further systematic and longitudinal studies in bats are needed to prevent future spillover events caused by SARSr-CoVs or to understand the origin of SARS-CoV-2 better.

ARTICLE HISTORY Received 8 April 2021; Revised 12 July 2021; Accepted 13 July 2021

KEYWORDS SARS-related coronavirus; novel lineage; bat; reservoir host; ACE2

Introduction

SARS-CoV-2, a novel coronavirus that causes COVID-19, was first identified in late 2019 [1] and took just a few months to sweep the globe. As the largest pandemic in the past century in human history, it has a severe impact on human health and leads to stagnation in economics, travel, education, and many other societal functions globally.

The natural origin of SARS-CoV-2 is an unanswered scientific question. It is believed that SARS-CoV-2 is transmitted from an animal reservoir host to human society through multiple intermediate hosts [2]. The discovery of SARS-CoV-2 related viruses (SARSr-CoV-2), RaTG13 and Pangolin-CoV from horseshoe bats and pangolin, respectively, shed light on the importance of these two groups as animal reservoirs of SARSr-CoV-2 viruses [1,3,4]. However, among the six critical residues of the receptor-binding domain (RBD) in the spike to interact with the human ACE2 receptor, RaTG13 only shares one with SARS-CoV-2 [5]. The RBD of RaTG13 has a lower binding affinity and usage efficiency with human ACE2, although it shares 96% genome sequence identity with SARS-CoV-2 [6–8]. One of the viruses derived from Malayan pangolin (*Manis javanica*), Pangolin-CoV-GD, possesses six identical critical residues of RBD with SARS-CoV-2 and displays a similar binding affinity to human ACE2 as SARS-CoV-2. However, it shares lower sequence identity with SARS-CoV-2 in the genome compared with RaTG13 [4,7,8]. Another SARSr-CoV-2 detected in bat (*Rhinolophus malayanus*), RmYN02, contains a similar insertion at the S1/S2 cleavage site in the spike of SARS-CoV-2; however, it has some deletions in the RBD and fails to bind with human ACE2 [9]. In addition, more SARSr-CoV-2 viral genome sequences from bats have been reported in Eastern China, Japan, and Southeast Asian countries [10–13]. However, a progenitor virus that shares >99% identity with SARS-CoV-2 remains unknown.

Bats also carry SARSr-CoV-1 with all the genetic building blocks of SARS-CoV-1, which jumped to

CONTACT Peng Zhou peng.zhou@wh.iov.cn; Zheng-Li Shi zlshi@wh.iov.cn CAS Key Laboratory of Special Pathogens and Biosafety, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan Institute of Virology, Wuhan 430071, People's Republic of China
*These authors contributed equally.

Supplemental data for this article can be accessed https://doi.org/10.1080/22221751.2021.1956373

humans in 2002 [14]. Therefore, the investigation of bat SARSr-CoVs is important for tracing the origin and immediate progenitor viruses of SARS-CoV-2 and critical for public health measures to prevent future outbreaks caused by these viruses. This study reports the genome characterization and viral receptor analysis of a novel lineage of SARSr-CoVs in Tongguan Town, Mojiang County, Yunnan Province in China in 2015, the same location where bat RaTG13 was found in 2013 [1].

Methods

Bat sampling and coronavirus detection

The sampling of bats was conducted in Mojiang County, Yunnan Province, in May 2015. Bats were released after the anal swab sampling. Samples were aliquoted and stored at −80°C until use. RNA was extracted using a High Pure Viral RNA Kit (Roche, Basel, Switzerland). Partial RdRp was amplified using the SuperScript III OneStep RT–PCR and Platinum Taq Enzyme kit (Invitrogen, Carlsbad, CA, USA) by family-specific degenerate seminested PCR. The PCR products were gel purified and sequenced using an ABI Prism 3700 DNA analyzer (Applied Biosystems, Foster City, CA, USA). The sequences were blasted against the GenBank database.

Genome sequencing

For SARSr-CoV-positive RNA extraction, next-generation sequencing (NGS) was performed using BGI MGISEQ 2000. NGS reads were first processed using Cutadapt (v.1.18) to eliminate possible contamination. Thereafter, the clean reads were assembled into genomes using Geneious (v11.0.3) and MEGAHIT (v1.2.9). PCR and Sanger sequencing were used to fill the genome gaps. To amplify the terminal ends, a SMARTer RACE 5`/3`kit (Takara) was used. Bat species identification was based on the partial sequence of the cytochrome c oxidase subunit I (COI) gene.

Phylogenetic analysis

Routine sequence management and analysis were performed using DNAStar software. Sequence alignments were created using ClustalW implemented in MEGA6 with default parameters. Maximum-Likelihood phylogenetic trees were generated using the Jukes-Cantor model with 1000 bootstrap replicates in the MEGA6 software package. Similarity plot analysis of the full-length genome sequences was conducted using the Simplot 3.5.1. The genome IDs used in the analysis were MN996528 for SARS-CoV-2, AY278488 for SARS-CoV-1, MN996532 for the bat SARSr-CoV RaTG13; MG772933 for ZC45; MW251308 for RacCS203; LC556375 for Rc-o319; KF367457 for WIV1; DQ022305 for HKU3-1; MT121216 for pangolin-CoV-GD strain; MT072864.1, pangolin-CoV-GX strain; EPI_ISL_412977 for bat SARSr-CoV RmYN02, and EPI_ISL_852604 for RshSTT182. The National Genomics Data Center of China ID for the eight novel lineage SARSr-CoVs were GWHBAUM01000000- GWHBAUT01000000. And the data of raw read generated from the sequencing of the 8 samples in the National Genomics Data Center (China) under the accession number CRA004339.

Expression constructs, protein expression, and purification

Codon-optimized RBD genes from the following viruses were used (see above genome accession number): SARS-CoV-2 (spike aa 330–583), SARS-CoV-1 (spike aa 317–569), RaTG13 (spike aa 330–583), pangolin-CoV-GD (spike aa 326–579), pangolin-CoV-GX (spike aa 330–583), and RaTG15 (spike aa 317–566). They were synthesized (Sangon Biotech, Shanghai, China) and placed into an expression vector with an N-terminal signal peptide and an S-tag, as described previously [15]. The ectodomains of human ACE2 (aa 19–615, accession number: AB046569), *R. affinis* ACE2 (aa 19–615, accession number: MT394204), and Malayan pangolin ACE2 (aa 19–615, accession number: XM_017650263.2) were amplified or synthesized, and cloned into the expression vector with an N-terminal signal peptide and C-terminal S-tag as described previously [15].

The RBD and ACE2 proteins used in the BLI binding assay were produced in HEK 293 T/17 cells. Cells were transiently transfected with expression plasmids using Lipofectamine 3000 (Life Technologies), washed twice with D-Hanks solution 6 h post-transfection, and cultured in fresh 293 T FreeStyle expression medium (Life Technologies) at 37°C in a humidified 5% CO2 incubator. The supernatants were harvested 48 h post-transfection and centrifuged at 4000 × *g* for 10 min at 4°C. Clarified supernatants were purified using S-tag agarose beads and eluted with 3 M $MgCl_2$. The purified protein was finally buffered with PBS, quantified using a Qubit 2 Fluorometer (Thermo Fisher Scientific), and stored at −80°C until use.

Bio-layer interferometry binding assays

Binding assays between RBDs and ACE2 proteins were performed using the Octet RED system (ForteBio, Menlo Park, CA, USA) in 96-well microplates at 30°C with shaking at 1000 rpm, as described previously [15]. The RBD was biotinylated using EZ-Link NHS-LC-LC-Biotin (Thermo Fisher Scientific, Waltham, MA, USA). The streptavidin biosensors

were activated for 200s before coupling with 50 μg/mL biotinylated RBD proteins for 600s. A baseline was collected in the kinetic buffer (1 M NaCl, 0.1% BSA, 0.02% Tween-20; pH 6.5) for 200s before immersing the sensors in a 1:2 or 1:3 serial diluted ACE2 protein for 900s and then dissociated in the same kinetic buffer for another 900s. Data analysis from the ForteBio Octet RED instrument included reference subtraction. Inter-step correction and Y-alignment were used to minimize the tip-dependent variability. Curve fitting was performed in a 1:1 model using data analysis software v7.1 (ForteBio, Menlo Park, CA, USA). The mean Kon and Koff values were determined using a global fit applied to all data. The coefficient of determination (R^2) for these interactions was close to 1.0.

Pseudovirus entry assays

Pseudotyped VSV-ΔG particles were generated as previously described, with minor adjustments [16]. HEK 293 T/17 cells were seeded in a 6-well-plate and transfected with plasmids containing codon-optimized SARSr-CoV-2 spike at 70% confluency using Lipofectamine 3000. At 6 h post-transfection, the medium was replaced with fresh DMEM + 10% FBS medium. At 24 h after transfection, cells were incubated with VSV-G-pseudotyped VSV△G/Fluc at 37°C for 1 h. Subsequently, the cells were washed five times and supplied with fresh DMEM + 10% FBS medium + anti-VSV-G antibody (Kerafast). Cell-free supernatants were harvested at 48 h after transduction and then centrifuged at 4000 × *g* for 10 min at 4°C. The virus particles were used directly for infection.

The 48-well-plate was treated with poly L-lysine solution (Sigma) before seeding HEK293 T/17 cells. Cells were transiently transfected with equal amounts of human ACE2, *R. affinis* ACE2, Malayan pangolin ACE2, or empty vector plasmids at 70% confluency. At 24 h post-transfection, the cells were incubated with the same amounts of S-pseudotyped virions for 2 h at 37°C, washed twice with PBS 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. Infection experiments were performed independently in triplicate, with three technical replicates each time.

Quantification of pseudotyped virus particles using RT–PCR

Viral RNA of all VSV-spike pseudovirus particles was extracted from 200 μl supernatant using the High Pure Viral RNA Kit (Roche, Cat. No. 11858882001), following the manufacturer's instructions. Quantification of pseudovirus by real-time PCR was performed using the HiScript® II One Step qRT-PCR SYBR Green Kit (Vazyme, Cat. No. Q221-01). The VSV P protein gene was amplified and synthesized *in vitro* using the mMESSAGE mMACHINE Kit (Life Technologies, Cat. No. AM1344) as the standard. Viral copy numbers were calculated using standard curves. Primers used for transcription *in vitro* were: VSV (P protein)-F1: GTTCGTGAGTATCTCAAGTCCT, VSV (P protein)-R2-T7: TAATACGACTCACTA-TA*GGGAGA*GCCTTGATTGTCTTCAATTTCTGG; primers used for real-time PCR were as described previously [17].

Results

Identification of a novel lineage of SARSr-CoVs

In tracing the origin of SARS-CoV-2 from bats, we identified RaTG13, which shares 96.2% genome identity with SARS-CoV-2 and is the closest genome to date [1]. Following the investigation, we identified eight SARSr-CoV sequences that share 93.5% sequence identity with SARS-CoV-2 in the 402-nt partial RdRp gene from bat samples collected in the same place in 2015. Seven samples were obtained from *Rhinolophus stheno*, and the other one was from *Rhinolophus affinis* (Table S1). Thus, we performed next-generation sequencing (NGS) to analyze these viruses further. Whole-genome sequences were obtained from all eight samples. The eight SARSr-CoV genomes are almost identical, sharing more than 99.7% sequence identity. One strain from *R. affinis*, designated RaTG15, was used as a representative in the subsequent analysis.

In the seven conserved replicase domains used for coronavirus species classification, RaTG15 was 95.3% or 92.5% identical to SARS-CoV-2 and SARS-CoV-1, respectively, suggesting that it remains a member of the *severe acute respiratory syndrome-related coronavirus* species in the *Sarbecovirus* subgenus within the *Betacoronavirus* genus, *Coronaviridae* family [18]. Furthermore, RaTG15 is genetically close to SARS-CoV-2 in open reading frame 1b (ORF1b). In the complete ORF1b region, RaTG15 showed 84.6~89.0% nucleotide identities and 95.6~97.3% amino acid sequence identities to bat SARSr-CoV-2 from wildlife in China and Southeast Asia, including bat CoVs RaTG13 and RmYN02 from Yunnan, Rc-o319 from Japan, RshSTT182 from Cambodia, RacCS203 from Thailand, and two different strains of pangolin-CoVs (Table S2). Phylogenetic analysis using full-length RdRp gene sequences also suggested that RaTG15 clustered with SARSr-CoV-2 (Figure S1A).

In contrast, similarity plot analysis revealed that beyond ORF1b, RaTG15 was significantly distinct from SARSr-CoV-2 and SARSr-CoV-1 in the majority

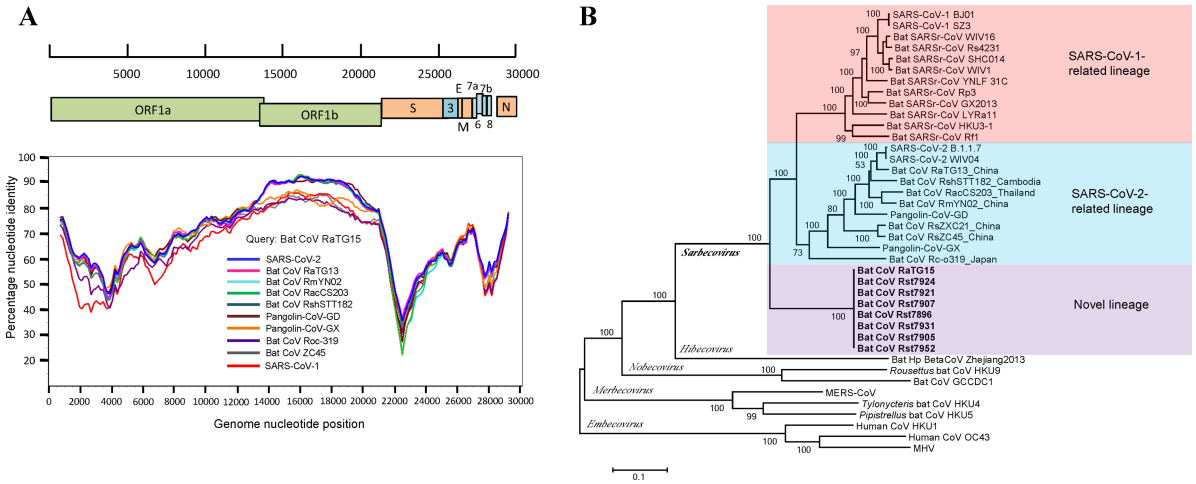


Figure 1. Discovery of a novel lineage of bat SARSr-CoVs. (A) Similarity plot analysis based on the full-length genome sequence of bat SARSr-CoV RaTG15. Full-length genome sequences of SARS-CoV-1, SARS-CoV-2, bat SARSr-CoVs, and pangolin CoVs related to SARS-CoV-2 were used as reference sequences. The analysis was performed with the Kimura model, a window size of 1500 base pairs and a step size of 150 base pairs. (B) Phylogenetic tree based on complete genome sequences of betacoronaviruses. The trees were constructed by the Maximum-Likelihood method using the Jukes-Cantor model with bootstrap values determined by 1000 replicates. Bootstraps > 50% are shown. The scale bars represent 0.1 substitutions per nucleotide position. The novel SARSr-CoVs characterized in this study are shown in bold. Ra, *Rhinolophus affinis*; Rst, *Rhinolophus stheno*; Rsh, *Rhinolophus shameli*; Rs, *Rhinolophus sinicus*; Rac, *Rhinolophus acuminatus*; Rm, *Rhinolophus malayanus*; Rc, *Rhinolophus cornutus*; MHV, murine hepatitis virus.

of the genome (Figure 1A). It exhibits less than 80% nucleotide identities in ORF1a, M, and N genes and lower than 70% identity in S, ORF3, 6, and 7a/7b to all other SARSr-CoVs (Table S2). Overall, the full genome of RaTG15 showed 74.4% sequence identity to SARS-CoV-1 and 77.6% to SARS-CoV-2. RaTG15 shows a higher sequence identity to SARS-CoV-1 than to SARS-CoV-2 in the spike, E, M, N, and ORF6 proteins. It also has nearly equivalent homology to any other known SARSr-CoVs from bat or pangolin CoVs (Table S2). This mosaic profile suggests that these novel lineage viruses may be a result of the recombination of different SARSr-CoVs.

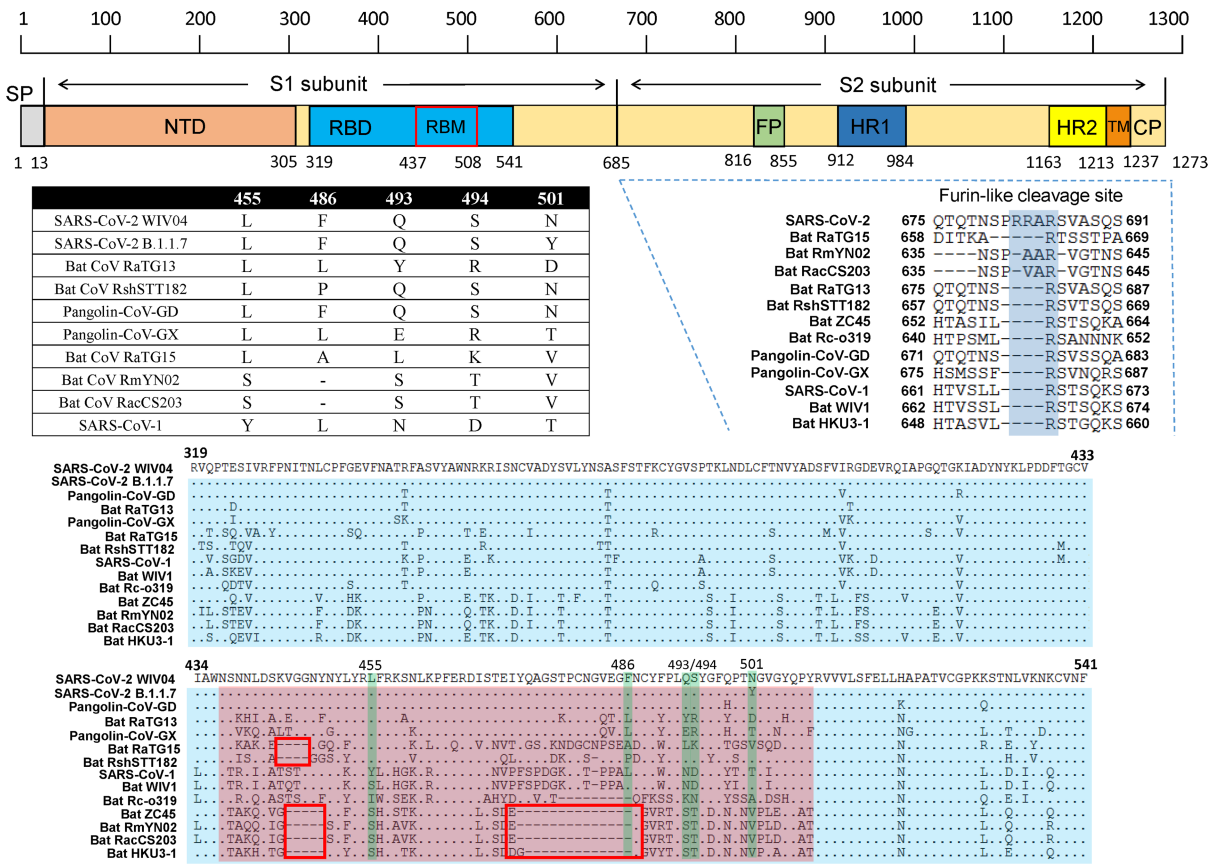
The results of the phylogenetic analysis were in accordance with the similarity plot. SARSr-CoVs mainly consist of two sub-lineages, SARSr-CoV-1 and SARSr-CoV-2 (Figure 1B). The latter includes SARSr-CoV-2 from pangolins and different *Rhinolophus* bat species recently reported in a wide range of areas in Asia. In the full-length genome tree and S gene tree, RaTG15 and the related viruses were distant from both existing sub-lineages and formed a well-supported novel lineage with the *sarbecoviruses* (Figure 1B and Figure S1B).

In silico analysis of receptor-binding domain (RBDs) of SARSr-CoVs

We further examined the spike protein sequence of RaTG15 in comparison with the other SARSr-CoV-2 strains. The receptor-binding domain (RBD) of the RaTG15 spike is highly divergent from other *sarbecoviruses*, with 72.6% amino acid sequence identity to SARS-CoV-2 and 68.6%–73.3% identity to related bat and pangolin CoVs. Unlike RmYN02 and RacCS203, the RaTG15 RBD does not contain the deletion corresponding to aa 473–486 (deletion 2) of the SARS-CoV-2 spike, which determines ACE2 usage based on previous reports [19]. However, aligned with SARS-CoV-2 and RaTG13, a short deletion was noted at the position corresponding to aa 444–447 (deletion 1). The location of this deletion is similar to that in the spike of RshSTT182, a SARSr-CoV-2 identified in *Rhinolophus shameli* from Cambodia. Within the receptor-binding motif (RBM), four of the five amino acid residues critical for the binding of SARS-CoV-2 to the ACE2 receptor (486, 493, 494, and 501) are varied in RaTG15. Similar to most bat SARSr-CoVs, the polybasic furin cleavage site is absent at the S1–S2 junction of RaTG15 (Figure 2).

Functional comparison of RBD from three lineages of SARSr-CoVs

Sequence analysis indicated that the RaTG15 virus possibly uses ACE2 as an entry receptor, which was then experimentally confirmed by RBD-ACE2 binding studies using purified recombinant proteins. RBD proteins from SARS-CoV-2, SARS-CoV-1, RaTG13, pangolin-CoV-GD, pangolin-CoV-GX, and RaTG15, and ectodomains of human (hACE2), *R. affinis* (RaACE2), and Malayan pangolin ACE2 (MpACE2) proteins were used (Figure S2A). We found that two *R. affinis*-derived RaTG13 and RaTG15 RBD proteins



	455	486	493	494	501
SARS-CoV-2 WIV04	L	F	Q	S	N
SARS-CoV-2 B.1.1.7	L	F	Q	S	Y
Bat CoV RaTG13	L	L	Y	R	D
Bat CoV RshSTT182	L	P	Q	S	N
Pangolin-CoV-GD	L	F	Q	S	N
Pangolin-CoV-GX	L	L	E	R	T
Bat CoV RaTG15	L	A	L	K	V
Bat CoV RmYN02	S	-	S	T	V
Bat CoV RacCS203	S	-	S	T	V
SARS-CoV-1	Y	L	N	D	T

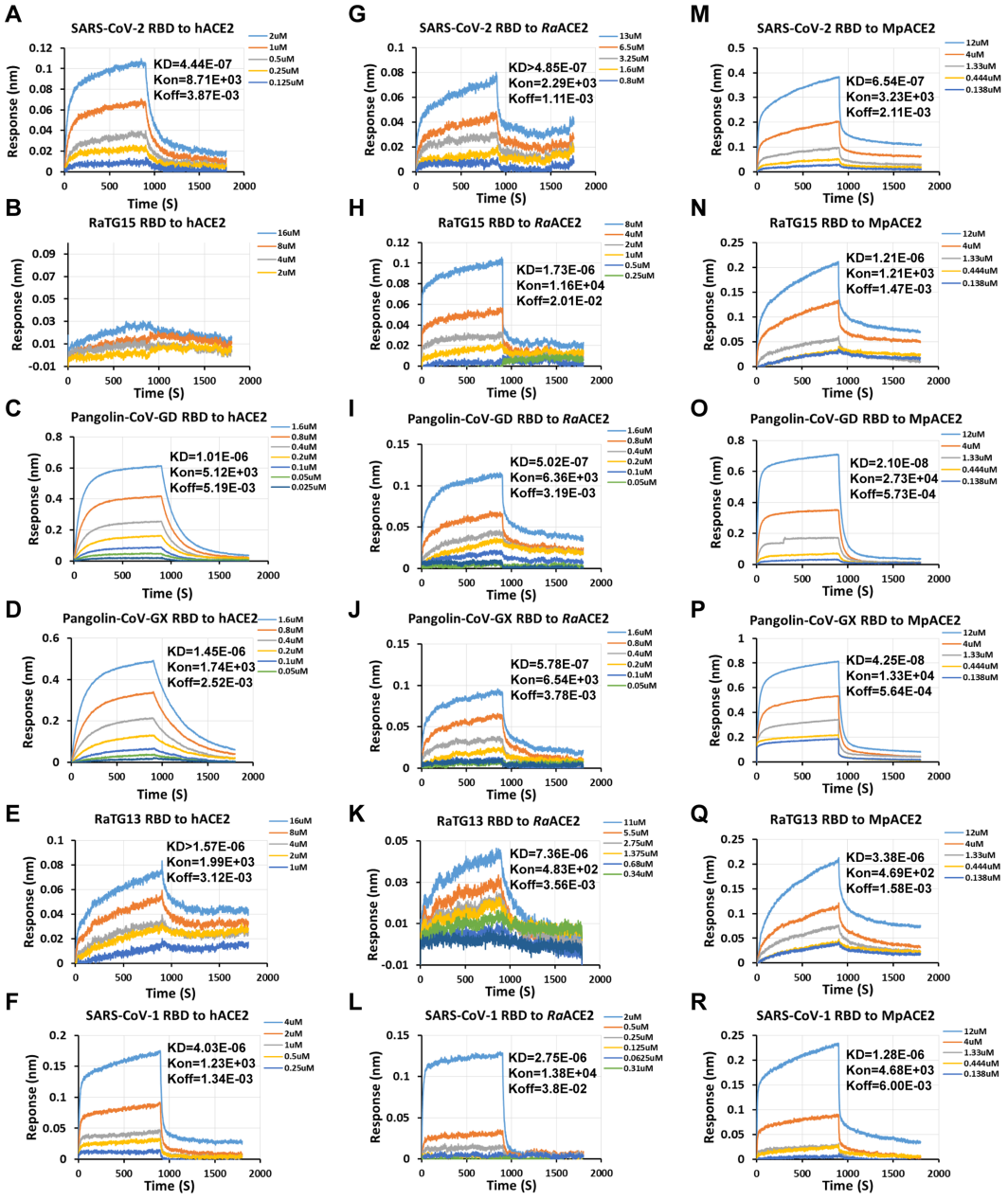
Figure 2. Comparison of the receptor-binding domain (RBDs) of SARSr-CoVs. The RBM is shown in pink and the five key residues that contact ACE2 directly are highlighted in green. A comparison of the five critical residues of these SARSr-CoVs is listed in the table. Two deletions in the RBM, aa 444–447 (deletion 1) and aa 473–486 (deletion 2) are indicated by red boxes. GenBank or GISAID entries for each virus can be found in Methods.

either showed very weak or no obvious binding affinity to human ACE2 (hACE2). In contrast, RBD proteins from the two pangolin SARSr-CoVs displayed much higher binding affinity to human ACE2, only slightly weaker than SARS-CoV-2 RBD, although remained higher than SARS-CoV-1 (Figure 3A–F and S). Furthermore, the binding affinity between human ACE2 and pangolin-CoV-GD RBD is comparable to that between SARS-CoV-2 and human ACE2. Thereafter, we needed to determine whether bat CoVs RaTG13 and RaTG15 can use *R. affinis* ACE2 or Malayan pangolin ACE2 more efficiently than human ACE2. Detectable binding was observed between RaTG15 RBD to *R.affinis* ACE2 (RaACE2) and Malayan pangolin ACE2, though the affinity remained weaker than SARS-CoV-2 and pangolin-CoV-GD/GX to RaACE2 and MpACE2. RaTG13 RBD showed very weak binding to RaACE2, MpACE2, and human ACE2 (Figure 3G-S and Figure S2B).

To exclude the possibility that the ACE2 binding of RBD may not represent the functionality of the full-length S protein, we also constructed a VSV-based pseudovirus using a previously published method [16]. We produced a list of SARSr-CoV pseudoviruses or MERS-CoV pseudovirus as a negative control. HEK293 T/17 cells overexpressing human ACE2, *R. affinis* ACE2, Malayan pangolin ACE2, or empty vector were infected with VSV-based pseudoviruses, and the infection efficiency was determined 48 h after infection. Consistent with the RBD-ACE2 protein binding assays, *R. affinis* and Malayan pangolin ACE2 supported all SARSr-CoV entries. The human ACE2 mediated entry of SARS-CoV-2, SARS-CoV-1, and pangolin-CoV-GD/GX efficiency, although less efficiency in RaTG15 and RaTG13 if it is positive, compared to other groups. As a control, MERS-CoV pseudovirus failed to infect ACE2-expression cells, confirming ACE2-independent infectivity of the VSV backbone (Figure S3). Collectively, none of the SARSr-CoV-2 lineage or the novel lineage virus from bats could efficiently bind to human ACE2 [10,11], and it appears that deletion at the RBD region greatly affects the binding capacity (Figure 3T).

Discussion

In general, this study analyzed the discovery of a novel lineage of SARSr-CoVs from bats that are closely related to SARS-CoV-2 in the RdRp region, although genetically distant from any known SARSr-CoVs at the genome level. Although several SARS-CoV-2



S

ACE2 / RBD	Human	*R.affinis*	Pangolin
SARS-CoV-2	1.00	>1.09	1.47
RaTG15	Not detectable	3.90	2.72
Pangolin-CoV-GD	2.28	1.13	0.05
Pangolin-CoV-GX	3.26	1.30	0.10
RaTG13	Low affinity, >3.53	16.57	7.61
SARS-CoV-1	9.07	6.19	2.88

T

		Human ACE2	Bat ACE2	Pangolin ACE2	Deletion 1	Deletion 2
SARS-CoV-1 related lineage	SARS-CoV-1	√	√	√		
	WIV16-CoV	√	√	ND		
	Rp3-CoV	×	×	ND	Y	Y
SARS-CoV-2 related lineage	SARS-CoV-2	√	√	√		
	RaTG13	weak	weak	weak		
	Rc-o0319	×	√	ND		Y, partial
	RmYN02	×	ND	ND	Y	Y
	RacCS213	×	ND	ND	Y	Y
	Pangolin-CoV-GX	√	√	√		
	Pangolin-CoV-GD	√	√	√		
Novel lineage	RaTG15	×	√	√	Y	

Figure 3. Binding affinity of SARSr-CoV RBDs to ACE2 from human, *R.affinis* and pangolin. (A-F) Binding of different RBD proteins to human ACE2. (G-L) Binding of different RBD proteins to *R.affinis* ACE2. (M-R) Binding of different RBD proteins to Malayan pangolin ACE2. (S) Comparison of dissociation constants (KD) between different RBD to human, *R.affinis,* and Malayan pangolin ACE2. Relative binding is analyzed by comparing with SARS-CoV-2 RBD to human ACE2. (T) Summary of the binding efficiency of different RBD to human, bat or pangolin ACE2. Y, yes; ND, not determined. Evidences for WIV16-CoV, Rc-o0319, RmYN02, and RacCS213 were from previous reports [10,11,22]. The presence of deletion in RBM (related to Figure 2) is indicated. Binding assay of human, *R.affinis* or pangolin ACE2 to different RBD proteins was measured by Bio-layer interferometry. The parameters of KD value (M), Kon (1/M.s), Koff (1/s) are shown on the upper right side of the picture. Different RBD proteins were immobilized on the sensors and tested for affinity with graded concentrations of human, *R. affinis,* or pangolin ACE2s. The Y-axis shows the real-time binding response. Values reported representing the global fit to all data. The coefficient of determination (R^2) for these interactions was close to 1.0 (Figure S2B).

related coronaviruses have been detected in wildlife, none of them share >99% genetic identity with SARS-CoV-2 at the genome level. Recombination events occur commonly in coronaviruses and can be referred to as the potential origin of the progenitor of SARS-CoV-1, as SARSr-CoVs discovered in a bat colony carried all the genomic fragments of SARS-CoV-1 [14,20]. The high sequence similarity to SARS-CoV-2 in some genomic regions detected from different wildlife species implies that recombination may occur during virus evolution in cross-species or inter-species transmission. The RBD protein of RaTG15 in this study showed weak binding affinity to its own and pangolin ACE2; however, no detectable binding to human ACE2 was observed, although it possesses one deletion in the RBD of the spike, which is different from the previously reported SARSr-CoVs in bats (Figure 2). Because of the failure to amplify the ACE2 gene from the feces of *R. stheno*, we could not test the binding affinity between RaTG15 RBD and *R. stheno* ACE2. Given that the other viruses in the novel lineage we found in this study were detected in *R. stheno*. We cannot rule out the possibility that RaTG15 may have originated from *R. stheno*. These results suggest that the SARSr-CoVs we discovered from bats now may be just the tip of the iceberg. These viruses may have experienced selection or recombination events in the animal hosts and render viral adaption to a new host, then spread to the new species before they jumped into human society. Therefore, surveillance of these new lineage viruses should be conducted to prevent future outbreaks, as viruses from the other two lineages of SARSr-CoV caused SARS and COVID-19, respectively [1,21]. Furthermore, to date, none of the bat SARSr-CoV-2 lineage or the novel lineage viruses discovered could be isolated or appear to be capable of efficiently using human ACE2, which suggests that without further adaptation, there is a limited zoonotic potential for these bat-derived SARSr-CoV-2 to spill and infect humans directly [22]. In contrast, pangolin-CoV has a high spillover potential in the context of cell receptor usage. Moreover, the ACE2 usage viruses in bat SARSr-CoV-1 lineage appear to be more dangerous in cross-species transmission, as demonstrated in animal studies [23,24].

The bat CoVs closest to SARS-CoV-2 at this stage, RaTG13, only showed very weak binding affinity to human ACE2. In contrast, pangolin-CoVs displayed a much higher affinity to ACE2 from human, bat and pangolin, posing high cross-species potential to humans or other species. The high binding affinity between pangolin ACE2s and different SARSr-CoVs RBD proteins implies that pangolin may be an ideal intermediate host for the transmission of SARSr-CoVs. In the context of SARS-CoV-2 animal origin, there could be a bat SARSr-CoV closer than RaTG13 capable of using human ACE2 efficiently, or a pangolin-CoV with higher similarity to SARS-CoV-2 in genome regions outside the S gene. In the future, a more systematic and longitudinal sampling of bats, pangolins, or other possible intermediate animals is required to understand the origin of SARS-CoV-2 better.

Acknowledgements

We thank Yun-Zhi Zhang and Ji-Hua Zhou from Yunnan CDC for their assistance with bat sampling. We would also like to thank Dr. Ding Gao of the WIV Core Facility and Technical Support for his help with the Octet RED technology.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was jointly supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB29010101 to Z-L.S) and the China National Science Foundation (81822028 to P.Z and 81290341 to Z-L.S).

ORCID

Peng Zhou http://orcid.org/0000-0001-9863-4201
Zheng-Li Shi http://orcid.org/0000-0001-8089-163X

References

[1] Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270–273.
[2] Zhou P, Shi ZL. SARS-CoV-2 spillover events. Science. 2021;371(6525):120–122.
[3] Xiao K, Zhai J, Feng Y, et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature. 2020;583(7815):286–289.
[4] Lam TT-Y, Jia N, Zhang Y-W, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature. 2020;583(7815):282–285.
[5] Hu B, Guo H, Zhou P, et al. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2021;19(3):141–154.
[6] Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020;581(7807):221–224.
[7] Wrobel AG, Benton DJ, Xu P, et al. Structure and binding properties of pangolin-CoV spike glycoprotein inform the evolution of SARS-CoV-2. Nat Commun. 2021;12(1):837.
[8] Starr TN, Greaney AJ, Hilton SK, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. Cell. 2020;182(5):1295–1310.e20.
[9] Zhou H, Chen X, Hu T, et al. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. Curr Biol. 2020;30(11):2196–2203.e3.

[10] Murakami S, Kitamura T, Suzuki J, et al. Detection and characterization of bat sarbecovirus phylogenetically related to SARS-CoV-2, Japan. Emerging Infectious Disease Journal. 2020;26 (12):3025–3029.
[11] Wacharapluesadee S, Tan CW, Maneeorn P, et al. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. Nat Commun. 2021;12(1):972.
[12] Hu D, Zhu C, Ai L, et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. Emerg Microbes Infect. 2018;7(1):154.
[13] Hul V, Delaune D, Karlsson EA, et al. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. bioRxiv. 2021. doi:10.1101/2021.01.26.428212
[14] Hu B, Zeng LP, Yang XL, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. Plos Pathog. 2017;13(11):e1006698.
[15] Guo H, Hu B-J, Yang X-L, et al. Evolutionary arms race between virus and host drives genetic diversity in bat severe acute respiratory syndrome-related coronavirus spike genes. J Virol. 2020;94(20): e00902–20.
[16] Johnson MC, Lyddon TD, Suarez R, et al. Optimized pseudotyping conditions for the SARS-COV-2 spike glycoprotein. J Virol. 2020;94(21):e01062–20.
[17] Li Q, Wu J, Nie J, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell. 2020;182(5):1284–1294.e9.
[18] King AMQ AM, Carstens EB, et al. Virus taxanomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses San Diego, 2012.
[19] Ren W, Qu X, Li W, et al. Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. J Virol. 2008;82(4):1899–1907.
[20] Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17(3):181–192.
[21] Drosten C, Gunther S, Preiser W, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med. 2003;348 (20):1967–1976.
[22] Yang XL, Hu B, Wang B, et al. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. J Virol. 2016;90(6):3253–3256.
[23] Menachery VD, Yount BL, Jr., Debbink K, et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nat Med. 2015;21(12):1508–1513.
[24] Menachery VD, Yount BL, Jr., Sims AC, et al. SARS-like WIV1-CoV poised for human emergence. Proc Natl Acad Sci U S A. 2016;113(11):3048–3053.