



## RESEARCH ARTICLE

# Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2

Zhixin Liu<sup>1,2</sup> | Xiao Xiao<sup>1</sup> | Xiuli Wei<sup>1</sup> | Jian Li<sup>1</sup> | Jing Yang<sup>1</sup> | Huabing Tan<sup>1</sup> | Jianyong Zhu<sup>1</sup> | Qiwei Zhang<sup>3,4</sup> | Jianguo Wu<sup>2,3</sup>  | Long Liu<sup>1</sup> 

<sup>1</sup>Department of Respiratory, Department of Infectious Diseases, School of Basic Medical Sciences, Renmin Hospital, Hubei University of Medicine, Shiyan, China

<sup>2</sup>State Key Laboratory of Virology and College of Life Sciences, Wuhan University, Wuhan, China

<sup>3</sup>Guangdong Provincial Key Laboratory of Virology, Institute of Medical Microbiology, Jinan University, Guangzhou, China

<sup>4</sup>School of Public Health, Southern Medical University, Guangzhou, China

## Correspondence

Jianguo Wu, Guangdong Provincial Key Laboratory of Virology, Institute of Medical Microbiology, Jinan University, 510632 Guangzhou, China.

Email: [jwu898@jnu.edu.cn](mailto:jwu898@jnu.edu.cn)

Long Liu, Department of Respiratory, Department of Infectious Diseases, School of Basic Medical Sciences, Renmin Hospital, Hubei University of Medicine, 442000 Shiyan, China.

Email: [liulong2015@outlook.com](mailto:liulong2015@outlook.com)

## Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 81471942, 81730061, 81902066; Hubei Provincial Natural Science Foundation, Grant/Award Numbers: 2018CFB093, 2018CFB185; Natural Science Foundation of Hubei Provincial Department of Education, Grant/Award Numbers: Q20192104, Q20192105; Cultivating Project for Young Scholar at Hubei University of Medicine, Grant/Award Numbers: 2016QDJZR03, 2017QDJZR08; Hubei Provincial Training Program of Innovation and Entrepreneurship for Undergraduates, Grant/Award Numbers: S201910929032, S201913249006

## Abstract

From the beginning of 2002 and 2012, severe respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) crossed the species barriers to infect humans, causing thousands of infections and hundreds of deaths, respectively. Currently, a novel coronavirus (SARS-CoV-2), which has become the cause of the outbreak of Coronavirus Disease 2019 (COVID-19), was discovered. Until 18 February 2020, there were 72 533 confirmed COVID-19 cases (including 10 644 severe cases) and 1872 deaths in China. SARS-CoV-2 is spreading among the public and causing substantial burden due to its human-to-human transmission. However, the intermediate host of SARS-CoV-2 is still unclear. Finding the possible intermediate host of SARS-CoV-2 is imperative to prevent further spread of the epidemic. In this study, we used systematic comparison and analysis to predict the interaction between the receptor-binding domain (RBD) of coronavirus spike protein and the host receptor, angiotensin-converting enzyme 2 (ACE2). The interaction between the key amino acids of S protein RBD and ACE2 indicated that, other than pangolins and snakes, as previously suggested, turtles (*Chrysemys picta bellii*, *Chelonia mydas*, and *Pelodiscus sinensis*) may act as the potential intermediate hosts transmitting SARS-CoV-2 to humans.

## KEYWORDS

angiotensin-converting enzyme 2 (ACE2), Coronavirus Disease 2019 (COVID-19), receptor-binding domain (RBD) of coronavirus, SARS-CoV-2, SARS-CoV-2 spike protein (S)

## 1 | INTRODUCTION

Severe respiratory syndrome coronavirus-2 (SARS-CoV-2), a novel coronavirus, caused Coronavirus Disease 2019 (COVID-19) outbreak in December 2019 in Wuhan, China.<sup>1,2</sup> SARS-CoV-2 is the seventh

coronavirus known to infect humans, the other six coronaviruses being HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, SARS-CoV, and Middle East respiratory syndrome coronavirus (MERS-CoV).<sup>3</sup> HCoV-229E and HCoV-NL63 belong to the alpha coronavirus, and others, including SARS-CoV-2, belong to beta coronavirus.

SARS-CoV and MERS-CoV are considered highly pathogenic and are known to be transmitted from bats to humans via intermediate host palm civets<sup>4</sup> and dromedary camels<sup>5</sup>. SARS-CoV, which causes severe acute respiratory syndrome (SARS), infected 8422 humans and resulted in 916 deaths in 37 countries between 2002 and 2003.<sup>6,7</sup> MERS-CoV was first identified in the Middle East in 2012. A report confirmed 1791 MERS-CoV infection cases, including at least 640 deaths in 27 countries, as of July 2016.<sup>8</sup> Because of the confirmed human-to-human transmission route, up to 18 February 2020, a total of 72 533 patients of COVID-19, caused by SARS-CoV-2, including 1872 deaths, were reported in China (<http://2019ncov.chinacdc.cn/2019-nCoV/index.html>). In addition, 505 COVID-19 cases have now been transmitted across other 24 countries (<http://2019ncov.chinacdc.cn/2019-nCoV/global.html>), a part of them because of contact or residence history with Wuhan.

Confirmation of intermediate hosts is essential to prevent further spread of the epidemic. This study focuses on comparisons of the spike sequences between SARS-CoV-2 with SARS-CoV, bat SARS-like CoV, and other coronaviruses, which are helpful for evolutionary analysis and finding the possible virus reservoirs. In addition, analysis of the ACE2 structures and binding motif alignment facilitates obtaining clues to differentiate the potential hosts.

## 2 | MATERIALS AND METHODS

### 2.1 | Sequences used in the study

Full-length protein sequences of spike glycoprotein and ACE2 were downloaded from the NCBI GenBank Database, including SARS-CoV-2 spike proteins (accession number: QHU79173, QHR84449, QHQ71963, QHO62107, QHO60594, QHN73795, QHD43416, and BBW89517) SARS-CoV spike proteins (accession number: ACU31051, ACU31032, NP\_828851, ABF65836, AAR91586, and AAP37017), bat SARS-like CoV (RaTG13, AVP788042, AVP78031, ATO98231, AGZ48828, AKZ19087, and AID16716). The bat SARS-like CoV RaTG13 sequence was downloaded from the GISAID (<http://www.GISAID.org>). The pangolin metagenome was downloaded from the NCBI BioProject database (PRJNA5732983), and the coronavirus genomes sequences were analyzed by VirMAP.<sup>9</sup>

### 2.2 | Protein sequences alignment and phylogenetic analysis

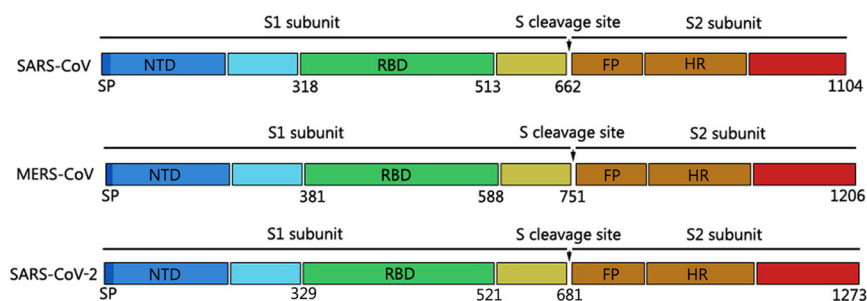
Alignment of spike protein sequences from different sources and residue comparison of ACE2 among different species were accomplished by MGEA-X (version 10.0.5). The phylogenetic analysis was accomplished through multiple comparisons using the neighbor-joining algorithm in the MGEA-X (version 10.0.5). Multiple comparisons were done by ClustalW multiple sequence alignment, the neighbor-joining phylogenies were estimated, and the number of bootstraps was 1000. The Poisson correction model and gamma-distributed pattern were used.

### 2.3 | Structure and binding model of spike receptor

The full-length structure of SARS-CoV-2 spike glycoprotein was simulated by the I-TASSER server online tool.<sup>10</sup> The spike-ACE2 binding model was predicted using PRISM 2.0.<sup>11</sup> The spike protein and ACE2 structure files were analyzed using PyMOL software (PyMOL v1.0).

## 3 | RESULTS

SARS-CoV-2 encodes at least 27 proteins, including 15 nonstructural proteins, 4 structural proteins, and 8 auxiliary proteins.<sup>12</sup> Spike glycoprotein (S), a structural protein located on the outer envelope of the virion, binds to the host-receptor angiotensin-converting enzyme 2 (ACE2). The S glycoprotein of SARS-CoV, MERS-CoV, and SARS-CoV-2 has 1104 to 1273 amino acids and contains an amino (N)-terminal S1 subunit and a carboxyl (C)-terminal S2 subunit<sup>13</sup> (Figure 1). In the S1 subunit, the receptor-binding domain (RBD), spanning about 200 residues, consists of two subdomains: the core and external subdomains.<sup>14,15</sup> The RBD core subdomain is responsible for the formation of S trimer particles.<sup>16</sup> The external subdomain contains two exposed loops on the surface, which bind with ACE2.<sup>17</sup> Investigating the evolutionary relationship of the RBD sequence in spike protein is helpful for understanding the virus origin trends.



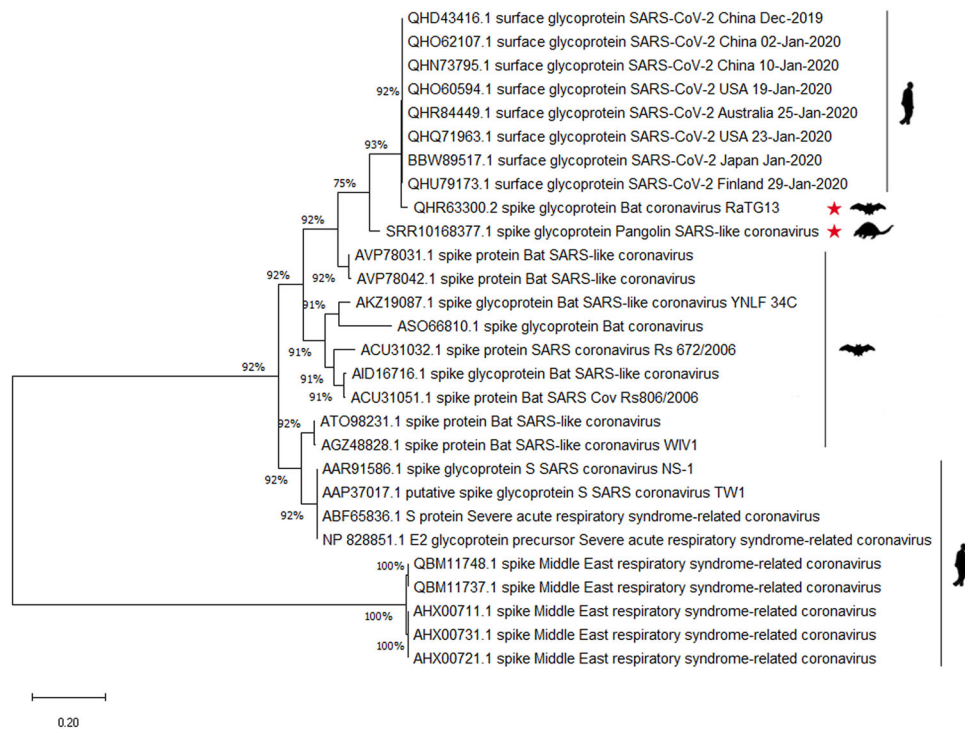
**FIGURE 1** Structural diagrams of spike glycoproteins of SARS-CoV, MERS-CoV, and SARS-CoV-2. All spike proteins of coronaviruses contain S1 subunit and S2 subunit, which were divided by the S cleavage sites. FP, fusion peptide; HR, heptad repeat 1 and heptad repeat 2; RBD, receptor-binding domain, contains core binding motif in the external subdomain; SP, signal peptide

Phylogenetic reconstruction determines the evolutionary relationship and host selection between spike glycoproteins in the human-close beta coronaviruses. To better understand the host selection of beta coronaviruses, the relationship of spike glycoprotein between SARS-CoV-2 and other closely related beta coronaviruses has been analyzed. The result showed that bat SARS-like CoV RaTG13, with 96.2% overall genome sequence identity,<sup>18</sup> is an inner joint neighbor of SARS-CoV-2 (Figure 2). In October 2019, it was reported that Sendai virus and coronavirus were the dominant viruses in the virome data of Malayan pangolins.<sup>19</sup> It is noteworthy that SARS-CoV was the most widely distributed coronavirus in the pangolin samples. We blasted the SARS-CoV-2 reads and found that the spike protein sequence was present in pangolin SARS-like CoV (numbered SRR10168377 here). The S protein in pangolin SARS-like CoV SRR10168377 (marked as red star) possesses only 75% similarity with SARS-CoV-2, partly because there were more than 220 residues in the S2 subunit that had not been read in the virome data. Removing the lost sequences, there was still 88% similarity with SARS-CoV-2 (data not shown); the full-length sequence of spike protein between pangolin SARS-like CoV and SARS-CoV-2 seems to be a little different. The temporal tree showed that the divergence time of spike sequence between bat SARS-like RaTG13 and SARS-CoV-2 is 0.18, while it is 1.50 in bat SARS-like RaTG13 to SARS-CoV-2 cluster (Figure S1).

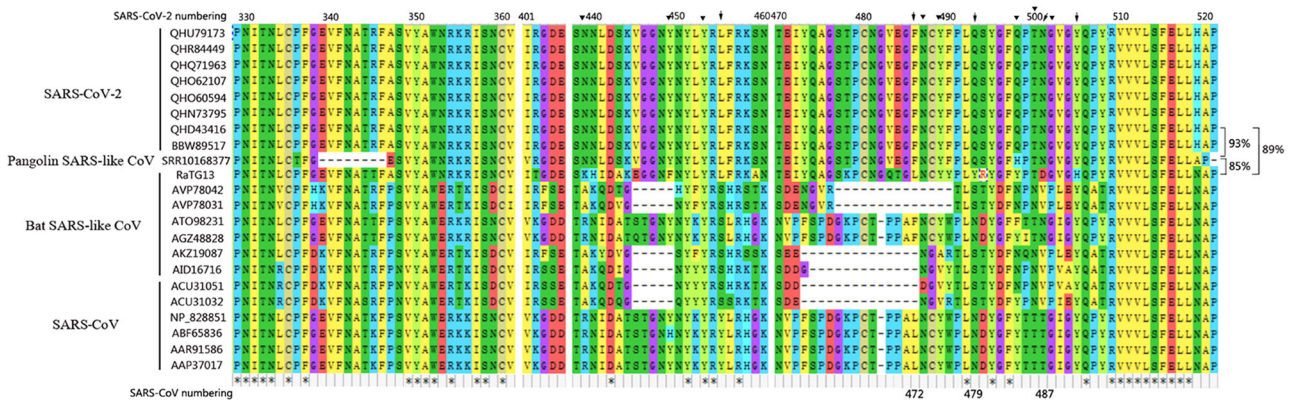
Global expansion and deep sequencing work have led to an increased amount of the SARS-CoV-2 genotype. Studying selective

stress may be helpful for assessing the variability and potential for host changes of SARS-CoV-2. Based on a report,<sup>20</sup> the selective pressure analysis showed that genes (*ORF10* and *ORF7a*) have a greater selective pressure, and *Spike* has an average pressure relative to the whole genome (Figure S2). The spike RBD-receptor interaction is a key factor determining the host range of coronaviruses. The RBD sequences of spike protein from SARS-CoV, bat, or pangolin SARS-like CoV and SARS-CoV-2 were aligned (Figure 3). There are some deletions from 473 to 490 residues in bat SARS-like CoV, which are located in the external subdomain, and it seems that these viruses do not infect humans naturally, given that there is no direct evidence showing that bat SARS-like CoV has ever infected humans. Interestingly, the SARS-CoV-2 RBD sequence from 329 to 521 possesses 93% identity with pangolin SARS-like CoV SRR10168377, which has higher than 89% similarity with bat SARS-like CoV RaTG13 (Figure 3). Pangolin SARS-like CoV SRR10168377 has only 85% similarity with bat SARS-like CoV RaTG13, indicating that if focusing on only the spike RBD, pangolin SARS-like CoV SRR10168377 has a higher probability to cross host barriers and infect humans.

SARS-CoV enters the respiratory tract by the receptor ACE2.<sup>14</sup> It was reported that SARS-CoV-2 uses the same cell entry receptor.<sup>18</sup> To understand the possibility of SARS-CoV-2 RBD binding with ACE2, the structure of spike protein of SARS-CoV-2 was simulated (Figure 4A). The difference in the whole structure of spike protein between SARS-CoV-2 (red) and SARS-CoV (green) was aligned, the coverage of the threading alignment was 0.83 (the number of aligned



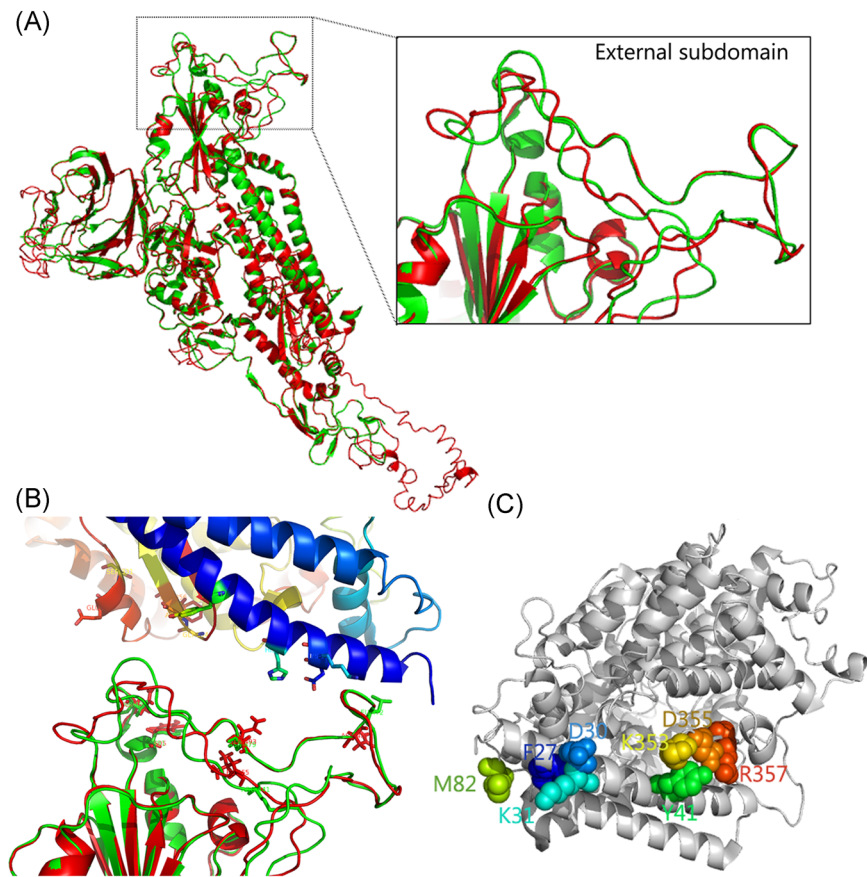
**FIGURE 2** Phylogenetic analysis of sequences of coronavirus spike glycoproteins. The sequences of spike glycoproteins of SARS-CoV-2, bat SARS-like CoV, pangolin SARS-like CoV, and SARS-CoV were analyzed. The red stars indicate pangolin SARS-like CoV and bat SARS-like CoV. Host flags are marked after the clusters. SARS-CoV-2, severe respiratory syndrome coronavirus-2



**FIGURE 3** Sequence alignments of the receptor-binding domain (RBD) of coronavirus spike glycoproteins. The amino acid sequences of RBD of severe respiratory syndrome coronavirus-2 (SARS-CoV-2), pangolin SARS-like CoV, bat SARS-like CoV, and SARS-CoV were aligned. The sequence identities among SARS-CoV-2, pangolin SARS-like CoV SRR10168377, and bat SARS-like CoV RaTG13 are marked. Arrows show the key residues in the interface binding to ACE2. Arrows indicate the critical binding residues for interaction between coronaviruses spike and receptor, triangle marks the ACE2-contacting residues on spike RBD. \*The unified amino acids in the same sites of all aligned sequences

residues divided by the full-length protein), and the normalized Z score of the threading alignment was 6.27 (>1 means a good alignment). The external subdomain in RBD has also been identified. The interaction model of SARS-CoV-2 with human ACE2 was then calculated by the PRISM 2.0 database (Figure 4B), which was generally similar to the structure of SARS-CoV S protein binding with human

receptor ACE2 (Figure 4C). Based on our results and two recent reports,<sup>21,22</sup> we noticed that SARS-CoV-2 numbering S protein may bind to ACE2 through Leu455, Phe486, Gln493, Asn501, and Tyr505, only quarter is same as SARS-CoV. In the external subdomain of SARS-CoV, residues L472, N479, and T487 (SARS-CoV numbering) are the main interfacing positions; mutations of these sites support S



**FIGURE 4** The binding of receptor-binding domain (RBD) external subdomain of spike glycoprotein with receptor ACE2. A, Structure comparison of stimulated SARS-CoV-2 spike and SARS-CoV spike glycoprotein (PDB: 5 × 58), SARS-CoV-2 spike is shown in red and SARS-CoV spike is shown in green. RBD contained external subdomain marked in the box. B, The binding model of SARS-CoV-2 spike (lower) with human receptor (upper). The possible residues in the interface of SARS-CoV-2 spike are shown as sticks. C, Human ACE2 critical for the binding with SARS-CoV-2 spike RBD is shown. The key residues located in the interface of ACE2 possibly in combination with spike RBD are shown as a sphere and colored

**TABLE 1** Residues comparison of ACE2 among different species interacting with SARS-CoV-2 spike glycoprotein RBD

Position		27	28	30	31	34	38	41	42	82	329	353	354	355	357
Species	Accession ID														
Human	NM_001371415	T	F	D	K	H	D	Y	Q	M	E	K	G	D	R
Gorilla	XM_019019204	T	F	D	K	H	D	Y	Q	M	E	K	G	D	R
Macaca	XM_011735203	T	F	D	K	H	D	Y	Q	M	E	K	G	D	R
Hamster	XM_003503235	T	F	D	K	Q	D	Y	Q	N	G	K	G	D	R
Cat	AY957464	T	F	E	K	H	E	Y	Q	T	E	K	G	D	R
Dog	GQ262793	T	F	E	K	Y	E	Y	Q	T	G	K	G	D	R
Pangolin ( <i>Manis javanica</i> )	XM_017650257	T	F	E	K	S	E	Y	Q	N	E	K	H	D	R
Snake ( <i>Ophiophagus hannah</i> )	AZIM01003601	...	F	E	Q	A	D	Y	N	M	V	K	E	D	R
Turtle ( <i>Chrysemys picta bellii</i> )	XM_024108749	...	F	S	Q	V	D	Y	A	K	N	K	K	D	R
Turtle ( <i>Chelonia mydas</i> )	XM_007070499	...	F	S	Q	V	D	Y	A	K	N	K	K	D	R
Turtle ( <i>Pelodiscus sinensis</i> )	XM_006122829	...	F	S	E	V	D	Y	A	K	K	K	K	D	R
Bat ( <i>Rhinolophus sinicus</i> )	KC881004	M	F	D	K	T	D	H	Q	N	N	K	G	D	R
Bat ( <i>Rhinolophus pearsonii</i> )	EF569964	T	F	D	K	H	D	H	E	D	N	K	D	D	R
Mouse	NM_027286	T	F	N	N	Q	D	Y	Q	S	A	H	G	D	R

Note: Mutated residues are highlighted as red at the corresponding positions based on human ACE2 numbering.

Abbreviations: RBD, receptor-binding domain; SARS-CoV-2, severe respiratory syndrome coronavirus-2.

adaptation for binding to the human ACE2 better than civet ACE2.<sup>23</sup> At the residue position of 82 in the mouse receptor, asparagine of bat (*Rhinolophus pearsonii*) ACE2 may disrupt the interaction with Leu472 of SARS-CoV RBD by steric interference hydrophobic contact, while Met82 in human receptor may interact with position 472 of RBD. Residues K479 and S487 in civet SARS-CoV can effectively recognize civet ACE2, but bind with human ACE2 much less efficiently.<sup>23</sup> The residue Thr487 in RBD binds to Tyr41 and Lys353 in human ACE2 by van der Waals contacts,<sup>25</sup> and threonine at position 487 enhanced the affinity of most RBDs for civet and human ACE2.<sup>24</sup> Residues in positions 479 and 487 in SARS-CoV correspond to the sites 493 and 501 in SARS-CoV-2, respectively. Residues Tyr493 and Asp501 in bat SARS-like CoV RaTG13 were replaced with Gln493 and Asn501 in SARS-CoV-2, respectively, same as pangolin SARS-like CoV SRR10168377.

Consistent with SARS-CoV binding to human ACE2, the residues 31, 41, 82, 353, 355, and 357 on the receptor locate in the interface when interacting with SARS-CoV-2 spike protein (Figure 4B,C). More than five substitutions were observed in turtle (*Chrysemys picta bellii*, *Chelonia mydas*, and *Pelodiscus sinensis*), pangolin (*Manis javanica*), and bat (both *Rhinolophus sinicus* and *R. pearsonii*) receptors in contrast to the theoretical infection-permissive species, such as human, gorilla, and macaca (Table 1). Mouse and dog also have multiple ( $\geq 5$ ) substitutions while cat and hamster only contain three mutations in the region. It was showed that Lys31, Tyr41, and Lys353 mutations substantially interfere with the S1-Ig association.<sup>24</sup> It is noteworthy

that at position 41, all maintain tyrosine in all observed species, except bats (both *R. sinicus* and *R. pearsonii*) substituted by histidine; at position 353, all contain lysine except mouse, which contains His353. The turtle receptor at position 31 contains glutamine or glutamic acid and mouse receptor contains asparagine; other observed species all contain lysine, same as humans. Substitution of His353 with Lys353 in mouse ACE2 allows SARS-CoV to infect murine cells,<sup>14</sup> indicating the importance of residue Lys353 binding with receptor. In addition, considering the function of position 41 in the interaction with receptor, tyrosine may possess higher affinity with Asn501 of RBD than histidine. From this perspective of Asn501 in RBD domain with the sites 41 and 353 of ACE2 receptor, it seems that turtles and pangolins are closer to humans than bat, indicating that turtles and pangolins may be the potential expanded hosts of SARS-CoV-2. The structural basis underlying this observed difference remains to be illustrated.

## 4 | DISCUSSION

Knowing the structural binding mechanism will support finding the accurate virus reservoirs and facilitate prevention countermeasure development. The crystallization of spike protein of SARS-CoV-2 was urgently analyzed by cryoelectron microscopy;<sup>13</sup> it helps to verify the interfacing sites in the interaction between SARS-CoV-2 spike RBD and human ACE2 or other hosts' receptors. It has been reported that

the binding capacity of SARS-CoV-2 S protein to ACE2 is much stronger than that of SARS-CoV,<sup>13</sup> which indicates that there are more intermediate hosts for SARS-CoV-2.

Pangolin SARS-like CoV SRR10168377 seems to point to the host range of SARS-CoV-2 expanded from pangolin. It is currently known that turtle also acts as a virus reservoir, carrying more than dozens of viruses, including ranavirus (RV), nidovirus (NV), papillomavirus (PV), soft-shelled turtle iridovirus (STIV), soft-shelled turtle systemic septicemia spherical virus (STSSSV), tortoise picornavirus (ToPV), and *Trionyx sinensis* hemorrhagic syndrome virus (TSHSV).<sup>26-30</sup> It does not exclude that bat SARS-like CoV RaTG13 or other SARS-like CoV infect turtle and expand to humans after evolution.

From SARS-CoV and MERS-CoV to SARS-CoV-2, all point to bats as the natural host of coronaviruses, but the intermediate hosts are all different. Previous findings suggested that the snake is a probable wildlife animal reservoir for SARS-CoV-2 based on its relative synonymous codon usage bias resembling snake compared with other animals.<sup>31</sup> Civets and dromedary, as the intermediate hosts of zoonotic SARS-CoV and MERS-CoV, live in different time and space environments. They nevertheless transmitted the coronaviruses to humans, which makes it difficult to find the intermediate hosts of SARS-CoV-2 in the short term. Analysis of the viral receptor-binding sequences and host receptors helps to quickly target the possible intermediate hosts of SARS-CoV-2. Compared with the illegally traded pangolins, turtles in the markets are more common and popular. This study provides information and possibilities that like snakes and pangolins, turtles (*C. picta bellii*, *C. mydas*, and *P. sinensis*) may also act as the potential intermediate hosts transmitting SARS-CoV-2 to human, although much more needs to be confirmed.

## ACKNOWLEDGMENTS

This study was supported by research grants from the National Natural Science Foundation of China (81902066, 81730061, and 81471942), the Natural Science Foundation of Hubei Provincial Department of Education (Q20192104 and Q20192105), the Hubei Provincial Natural Science Foundation (2018CFB093 and 2018CFB185), Cultivating Project for Young Scholar at Hubei University of Medicine (2017QDJZR08 and 2016QDJZR03), and the Hubei Provincial Training Program of Innovation and Entrepreneurship for Undergraduates (S201910929032 and S201913249006).

## CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

## AUTHOR CONTRIBUTIONS

ZL, JW, and LL contributed to data analysis, technical graphics, and writing the paper. ZL, XX, XW, JL, JY, HT, JZ, and QZ contributed to data collection and graphics processing. QZ, JW, and LL contributed to editing the paper.

## ORCID

Jianguo Wu  <http://orcid.org/0000-0002-8326-2895>

Long Liu  <http://orcid.org/0000-0002-1265-7603>

## REFERENCES

- Gorbalenya AE, Baker SC, Baric RS, et al. Severe acute respiratory syndrome-related coronavirus: the species and its viruses—A statement of the Coronavirus Study Group. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.02.07.937862>
- Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382:727-733. <https://doi.org/10.1056/NEJMoa2001017>
- Tang Q, Song Y, Shi M, Cheng Y, Zhang W, Xia XQ. Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. *Sci Rep*. 2015;5:17155.
- Guan Y. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science*. 2003;302(5643):276-278.
- Drosten C, Kellam P, Memish ZA. Evidence for camel-to-human transmission of MERS coronavirus. *N Engl J Med*. 2014;371(14):1359-1360.
- Chan-Yeung M, Xu RH. SARS: epidemiology. *Respirology*. 2003;8(suppl):S9-S14.
- WHO. *Coronavirus Infections: Disease Outbreak News WHO*; 2016. <http://www.who.int/csr/don/25-july-2016-mers-saudi-arabia/en/>
- WHO. *Middle East Respiratory Syndrome Coronavirus (MERS-CoV)—Saudi Arabia*; 2016. <http://www.who.int/csr/don/25-july-2016-mers-saudi-arabia/en/>
- Ajami NJ, Wong MC, Ross MC, Lloyd RE, Petrosino JF. Maximal viral information recovery from sequence data using VirMAP. *Nat Commun*. 2018;9(1):3205.
- Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res*. 2015;43(W1):W174-W181.
- Tuncbag N, Gursoy A, Nussinov R, Keskin O. Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat Protocol*. 2011;6(9):1341-1354.
- Wu A, Peng Y, Huang B, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe*. 2020. <https://doi.org/10.1016/j.chom.2020.02.001>
- Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.02.11.944462>
- Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*. 2005;309(5742):1864-1868.
- Lu G, Hu Y, Wang Q, et al. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature*. 2013;500(7461):227-231.
- Yuan Y, Cao D, Zhang Y, et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun*. 2017;8:15092.
- Letko M, Munster V. Functional assessment of cell entry and receptor usage for lineage B  $\beta$  coronaviruses, including 2019-nCoV. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.01.22.915660>
- Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020. <https://doi.org/10.1038/s41586-020-2012-7>
- Liu P, Chen W, Chen JP. Viral metagenomics revealed Sendai virus and coronavirus infection of Malayan pangolins (*Manis javanica*). *Viruses*. 2019;11(11):979.
- Yu WB, Tang GD, Zhang L, et al. Decoding the evolution and transmissions of the novel pneumonia coronavirus (SARS-CoV-2) using whole genomic data. *ChinaXiv*. 2020.02.00033v2.
- Lam TT, Shum MH, Zhu HC, et al. Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. *bioRxiv*. 2020. 02.13.945485.

22. Matthew CW, Sara JC, Nadim JA, et al. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. *bioRxiv*. 2020.02.07.939207.
23. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol*. 2010;84(7):3134-3146.
24. Li W, Zhang C, Sui J, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J*. 2005;24(8):1634-1643.
25. Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining host 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol*. 2015;23(8):468-478.
26. Zhang L, Li D, Liu L, Fang J, Xu R, Zhang G. Development of a colloidal gold immunochromatographic strip for the rapid detection of soft-shelled turtle systemic septicemia spherical virus. *J Virol Meth*. 2015;221:39-45.
27. Ng TF, Manire C, Borrowman K, Langer T, Ehrhart L, Breitbart M. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *J Virol*. 2009;83(6):2500-2509.
28. Lyu S, Yuan X, Zhang H, et al. Transcriptome profiling analysis of lung tissue of Chinese soft-shell turtle infected by *Trionyx sinensis* Hemorrhagic syndrome virus. *Fish Shellfish Immunol*. 2020;98:653-660.
29. Huang Y, Huang X, Wang S, Yu Y, Ni S, Qin Q. Soft-shelled turtle iridovirus enters cells via cholesterol-dependent, clathrin-mediated endocytosis as well as macropinocytosis. *Arch Virol*. 2018;163(11):3023-3033.
30. Zhang J, Finlaison DS, Frost MJ, et al. Identification of a novel nidovirus as a potential cause of large scale mortalities in the endangered Bellinger River snapping turtle (*Myuchelys georgesi*). *PLOS One*. 2018;13(10):e0205209.
31. Ji W, Wang W, Zhao XF, Zai JJ, Li XG. Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. *J Med Virol*. 2020;92:433-440. <https://doi.org/10.1002/jmv.25682>

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Liu Z, Xiao X, Wei X, et al. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J Med Virol*. 2020;92:595-601. <https://doi.org/10.1002/jmv.25726>