

# Which mammalian species are at risk of being infected by SARS-CoV-2: an ACE2 perspective

**Yulong Wei**

University of Ottawa

**Parisa Aris**

University of Ottawa

**Heba Farookhi**

University of Ottawa

**Xuhua Xia** (✉ [xxia@uottawa.ca](mailto:xxia@uottawa.ca))

University of Ottawa

---

## Research Article

**Keywords:** ACE2, coronavirus Spike protein, SARS-CoV-2, SARS-CoV

**Posted Date:** May 11th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-28141/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on January 18th, 2021. See the published version at <https://doi.org/10.1038/s41598-020-80573-x>.

# Abstract

SARS-CoV-2 can transmit efficiently in humans, but it is less clear what other mammalian are at high risk of being infected. SARS-CoV-2 contain a Spike (S) protein that uses mammalian *ACE2* receptors to mediate cell entry, a species with a human-like *ACE2* receptor is therefore at risk of being infected by SARS-CoV-2. We compared between 131 mammalian *ACE2* genes and 15 coronavirus S proteins. We showed that global similarity reflected by the phylogenetic relationship from *ACE2* gene alignment is a poor predictor of high-risk mammals, whereas local *ACE2* similarities at key binding sites highlight several high-risk mammals. Both SARS-CoV and SARS-CoV-2 likely have a bat origin; however, direct human transmission is unlikely due to their differences in *ACE2* receptors, and various mammals share similar or better homologies in *ACE2* receptor with humans. Furthermore, by comparing key binding sites at S protein of SARS-like coronaviruses in high-risk mammals, we found high similarities in S protein binding domains between SARS-CoV-2 and Pangolin-CoV but not Civets-CoV, and high similarities between SARS-CoV and Civets-CoV but not Pangolin-CoV. Hence, evolutionary adaptation of the bat virus in different intermediate hosts could allow it to acquire distinct high binding potential between S protein and human-like *ACE2* receptors.

## Introduction

SARS-CoV-2 poses a serious global health emergency. Since its emergence in Wuhan city, Hubei province of China in December 2019, the viral outbreak has resulted in over 2.8 million confirmed cases worldwide and 195,000 deaths<sup>1</sup>. While it is evident that SARS-CoV-2 can transmit efficiently from person to person, it is less clear what other mammalian species are at high risk of being infected. Both SARS-CoV and SARS-CoV-2 genomes encode a spike protein that uses the human *ACE2* receptor to mediate viral entry into the host cell<sup>2-5,6,7</sup>. Mechanistically, two S protein domains are involved during *Betacoronavirus* infection on mammalian cells that express *ACE2* receptors, with the S1 domain interacting with the *ACE2* receptor<sup>8,9</sup>, and the S2 domain undergoing structural rearrangements to mediate membrane fusion<sup>4</sup>. Interacting with the S1 domain, the *ACE2* receptor-binding domains are primarily located in  $\alpha$ -helix 1 and  $\beta$ -sheet 5<sup>4</sup>. Specifically, previous SARS-CoV studies have identified many *ACE2* genetic hotspots that affect viral infectivity, and the efficacy of the interaction between S proteins and *ACE2* receptors is a good predictor of severity of *Betacoronavirus* infection<sup>4</sup> and potentially plays an important role in subsequent viral replication<sup>10,11</sup>. Hence, determining the key sites at the human *ACE2* receptor involved in binding with SARS-CoV-2 S proteins and their differences in comparison to mammalian *ACE2* receptors is crucial to provide insights into viral infectivity, host range, and the patterns of viral evolutionary adaptation that have allowed interspecies transmission and infection of human *ACE2* expressing cells. Answering these questions is important to 1) improve our ability to predict and control future pandemics, 2) facilitate efforts in the development of a potential SARS-CoV-2 vaccine, and 3) manage and protect wildlife and domesticated animals.

A potential source of the 2002-2003 SARS-CoV outbreak was the masked palm civets. The palm civet *ACE2* receptors bind efficiently to Civet-CoV S proteins isolated from infected palm civet hosts (strain SZ3), from patients with the severe 2002-2003 SARS-CoV (strain TOR2), and from patients with the less severe 2003-2004 SARS-CoV (strain GD)<sup>12-15</sup>. Whereas in human cells, *ACE2* receptors only bind efficiently to S proteins from SARS-CoV TOR2, but not to that from Civet-CoV SZ3 or from the less severe GD variant<sup>4</sup>. Furthermore, early studies found that differences in binding domains explained why S protein-mediated SARS-CoV infection in human is also efficient in palm civet but is not efficient in rats<sup>4</sup>, and introducing point mutations in the *AEC2* gene variably changed the efficiency of receptor binding in palm civets, humans, and rats. For example, experimentally mutating rat *ACE2* His353 to human Lys353 changed the rat *ACE2* receptor from one that poorly binds S protein into one that is efficient for binding, introducing human *ACE2* residues 82 to 84 into rats led to an increase in S1-protein binding, and changing human *ACE2* Met82 into rat *ACE2* Asn82 interfered with S protein mediated entry. Furthermore, changing human *AEC2* Lys31, Tyr41, Asp355, and Arg357 also interfered with S protein binding. On the contrary, other *ACE2* site differences, such as Glycine 354 in human and Asparatic acid 354 in palm civet, and their differences in residues 90-93 that potentially interact to S protein residue 479, did not contribute to a difference the efficacy of S protein binding<sup>10</sup>. Hence, although major sequence differences are found between palm civet and human *ACE2* in  $\alpha$ -helix 1 (residues 30-40) and in  $\alpha$ -helix 3 (residues 90-93), they do not contribute to differences in efficacy of S protein binding. In brief, mutation experiments by Li, et al.<sup>3</sup> shows that the conservation for Lys31 and Tyr41 on  $\alpha$ -helix 1, residues 82-84 in the vicinity of  $\alpha$ -helix 3, and residues 353-357 in  $\beta$ -sheet 5 in human *ACE2* may be crucial for efficient S protein binding in SARS-CoV because their replacements lead to altered binding dynamics between the S protein and *ACE2* receptor.

To illustrate the importance of contrasting between human and other mammalian *ACE2* at key sites crucial for S protein binding, figure 1 compares *ACE2* gene similarities in a sample of mammalian species in two ways: one is global similarity reflected by the phylogenetic tree from *ACE2* alignment (Fig. 1A), and the other is local similarity in *ACE2* key sites that are crucial for S protein binding from mutation experiments by Li, et al.<sup>3</sup> (Fig. 1B). Figure 1A shows that the human *ACE2* gene shares higher global similarity to mammalian *ACE2* genes of other Primates and species of the Rodentia family than to species of the Carnivora, Artiodactyla, and Chiroptera orders. This emphasizes that global *ACE2* gene comparison is not a good predictor of which species is at high risk of being infected by SARS-CoV because we cannot infer the two aforementioned information: that the progenitor of SARS-CoV is likely a bat (from the Chiroptera order) virus, and that the probably intermediate host is masked palm civets (from the Carnivora family). In contrast, by comparing the *ACE2* gene at key residues that are involved in human SARS-CoV infection, we found that certain species in the Carnivora, Artiodactyla, and Chiroptera orders are at higher risk of being infected by SARS-CoV-2, even though they share less global *ACE2* similarity than other Primates and Rodentia species. Indeed, global similarity between palm civet *ACE2* and human *ACE2* does not identify palm civet as a high-risk creature, but key site similarity comparisons does suggest palm civets as a high-risk intermediate host.

Similarly, given there are key binding sites in human *ACE2* that form contacts with SARS-CoV-2 S protein, it is important to understand how well different mammalian *ACE2* feature these binding sites, in order to find species at high risk of SARS-CoV-2 infection that may act as potential intermediate hosts. *ACE2* binding by SARS-CoV and SARS-CoV-2 share overall structural similarity, likely constrained by the structure of *ACE2*<sup>6</sup>. However, the spike protein is notably different between SARS-CoV and SARS-CoV-2 with only about 75% homology<sup>5,16</sup>. Based on two recent crystal structure experiments<sup>6,7</sup>, different residues on the *ACE2* receptor are involved in SARS-CoV-2 and SARS-CoV infections, these differences are summarized in Table 1. In brief, in SARS-CoV-2 Ser19 forms a new hydrogen bond with *ACE2* Ala475 while SARS-CoV lacks this interaction<sup>7</sup>. The *ACE2* Asp30 forms salt bridge with SARS-CoV-2 S protein at Lys417 but does not interact with SARS-CoV S protein<sup>6</sup>. Other noteworthy differences are *ACE2* Glu35 and Arg393 interacting only with the SARS-CoV-2 S-proteins (Table 1). These differences may contribute to a difference in binding affinity between the two viruses and suggest distinct viral infectivity and host range. Additionally, because the S proteins of SARS-CoV-2 and SARS-CoV have unique structural differences, even when the same *ACE2* residues are involved in binding, they form different contacts. For example, human *ACE2* Gln24 interacts with SARS-CoV-2 S-proteins at Leu472 but with SARS-CoV S proteins at Asn473<sup>6</sup>, and *ACE2* Leu79 and Met82 interact with SARS-CoV-2 S proteins at Phe486 but with SARS-CoV S protein at Leu472. While binding at Gln24 was not previously documented to affect S protein binding in SARS-CoV<sup>10</sup>, the difference in its binding to SARS-CoV-2 may contribute to a difference in binding affinity between the two viruses<sup>7</sup>. Furthermore, the conservation of several *ACE2* hotspots may be critical to achieve efficient S protein binding; examples are Lys31 and Lys353<sup>7</sup>, both forming salt bridges with the S protein of SARS-CoV-2 and of SARS-CoV. In total, at the *ACE2* receptor's binding region, 15 residues are potentially crucial in binding to SARS-CoV-2 S protein, among which four are distinctly found to interact with SARS-CoV-2 S proteins (Table 1). As of the time this manuscript was submitted, studies that mutate human *ACE2* interface residues to assess their binding potential to SARS-CoV-2 are scarce. Hence, we cannot rule out the possibility that all *ACE2* interface residues listed in Table 1 are important in SARS-CoV-2 binding. A comparative gene analysis of these 15 binding residues on mammalian *ACE2* receptors may provide crucial insights into 1) the identification of mammals with high risk of SARS-CoV-2 infections, and 2) the potential intermediate hosts that facilitated the zoonotic transmission of SARS-CoV-2 infection into humans.

Besides the comparison of *ACE2* gene to understand SARS-CoV-2 and SARS-CoV *ACE2* host range, one other important question is how did bat SARS-like coronavirus transmitted to humans? Given that there are distinct sites on the S1 receptor binding domain between SARS-CoV-2 and SARS-CoV, we wish to know which mammalian-CoV have the same sites as the two human SARS coronaviruses. Answering this question facilitates the identification of 1) a progenitor that preceded the common ancestor of SARS-CoV-2 and SARS-CoV, and 2) which mammalian-CoV may evolve to exploit these similarities to infect humans. SARS-CoV-2 might have evolved from the bat RaTG13 coronavirus because they share high S protein sequence similarity. Notably, different from SARS-CoV, SARS-CoV-2 S protein contains a Gly482, Val483, Glu484, and Gly485 four-residue motif in the binding ridge that may facilitate contact with the N-terminal helix on human *ACE2* receptors. Similarly, the bat RaTG13 virus also contains the four-residue

motif<sup>7</sup>. This may explain why RaTG13 could use human *ACE2* as its receptor<sup>7</sup> and the simplest assumption is that there are no intermediate hosts. However, the S protein binding domain of the pangolin *Betacoronavirus* Pangolin-CoV also shares high sequence similarity with that of SARS-CoV-2 and it also contains the 482-485 four-residue motif<sup>7</sup>. Additionally, the *ACE2* receptor is more similar between humans and pangolin than between humans and bats. Indeed, pangolins have been proposed to be an intermediate host<sup>17</sup>.

If the intermediate host for SARS-CoV-2 is distinct from SARS-CoV, then it should meet two criteria. First, the culprit should share high similarity with humans in *ACE2* binding domains and are at high risk for SARS-CoV-2 infection. Second, the SARS-like coronaviruses that are species-specific to high risk mammals should share similarities in the S1 domains with human SARS-CoV-2 and with bat RaTG13 but not with SARS-CoV. Many residues in RaTG13 S protein are not fine-tuned for binding human *ACE2*<sup>6,7</sup>. Changing bat RaTG13 S protein at Lys486 and Tyr493 to SARS-CoV-2 S protein Phe486 and Gln493, respectively, enhanced human *ACE2* recognition<sup>7</sup>. Similarly, changing RaTG13 Lys479 to SARS-CoV Asn479 also increased human *ACE2* binding. Moreover, residues Leu455 and Asn501 are found to favor human *ACE2* binding and they are conserved between RaTG13 and SARS-CoV-2<sup>7</sup>, and residues Tyr442, Leu472, Asn479, Thr487 and Tyr505 were also identified as critical sites involved in *ACE2* binding<sup>18</sup>. Table 1 details all other S protein residues on SARS-CoV-2 and on SARS-CoV that interacts with the human *ACE2* receptor. Although S protein amino acid sites may differ, many interact with the same *ACE2* residues, among which some residues are identical amino acids (in blue, Table 1) whereas others are different amino acids but share similar chemical properties (underlined, Table 1)<sup>6</sup>. In total, there are 15 SARS-CoV-2 S protein residues and 13 SARS-CoV S protein residues that are critical to *ACE2* binding (Table 1). Determining the differences in S protein binding sites of SARS-CoV-2, SARS-CoV, and SARS-related viruses in other mammalian species may shed light on the evolution of Betacoronaviruses that mediated animal to human transmission.

Table 1. Contacting residues at *ACE2*/SARS-CoV-2 (columns 1 and 2) and at *ACE2*/SARS-CoV (columns 3 and 4) crystal structure interfaces, retrieved from Shang, et al. <sup>7</sup> and Lan, et al. <sup>6</sup>. Each row specifies the contacts formed between human *ACE2* residue and S protein residue (e.g., row 2; *ACE2* S19 – SARS-CoV-2 A475). In red are distinct residues on the *ACE2* receptor interface for binding with SARS-CoV-2 and SARS-CoV S proteins. In blue are shared S protein amino acids between the two SARS coronaviruses that interact with the same *ACE2* residues, although the position of S protein residue differs between the two viruses. Underlined are different S protein amino acids interacting with the same *ACE2* residues, but these interactions share similar biochemical properties.

Human <i>ACE2</i>	SARS-CoV-2	Human <i>ACE2</i>	SARS-CoV
<b>S19</b> <sup>1</sup>	A475 <sup>1</sup>	Q24 <sup>2</sup>	N473 <sup>2</sup>
Q24 <sup>2</sup>	N487 <sup>2</sup>	K311,2	Y442 <sup>1,2</sup>
Q24 <sup>2</sup> , M82 <sup>1,2</sup> , L79 <sup>1,2</sup> , Y83 <sup>1,2</sup>	F486 <sup>1,2</sup>	H34 <sup>2</sup>	L443 <sup>2</sup> , N479 <sup>2</sup>
<b>D30</b> <sup>2</sup>	K417 <sup>2</sup>	E37 <sup>2</sup>	Y491 <sup>1,2</sup>
K311,2	L455 <sup>1,2</sup> , Q493 <sup>2</sup>	D381,2	Y436 <sup>2</sup> , Y484 <sup>2</sup>
H34 <sup>2</sup> , <b>E35</b> <sup>1,2</sup>	F456 <sup>2</sup> , Q493 <sup>1,2</sup>	Y41 <sup>2</sup>	Y484 <sup>2</sup> , T486 <sup>2</sup> , T487 <sup>2</sup>
E37 <sup>2</sup> , R393 <sup>2</sup>	Y505 <sup>2</sup>	Q42 <sup>2</sup>	Y436 <sup>2</sup> , Y484 <sup>2</sup>
D381,2	Y449 <sup>2</sup> , Q498 <sup>2</sup>	L79 <sup>2</sup> , M82 <sup>1,2</sup>	L472 <sup>1,2</sup> ,
Y41 <sup>2</sup>	Q498 <sup>2</sup> , T500 <sup>2</sup> , N501 <sup>2</sup>	Y83 <sup>2</sup>	Y475 <sup>2</sup> , N473 <sup>2</sup>
Q42 <sup>2</sup>	G446 <sup>2</sup> , Y449 <sup>2</sup> , Q498 <sup>2</sup>	Q325 <sup>2</sup> , E329 <sup>2</sup>	R426 <sup>2</sup>
Y83 <sup>2</sup>	Y489 <sup>2</sup> , N487 <sup>2</sup>	N330 <sup>2</sup>	T486 <sup>2</sup>
K353 <sup>1,2</sup>	N501 <sup>1</sup> , G502 <sup>2</sup>	K353 <sup>1,2</sup>	T487 <sup>1</sup> , G488 <sup>2</sup>

1 Interacting residues retrieved from Shang, et al. <sup>7</sup>

2 Interacting residues retrieved from Lan, et al. <sup>6</sup>

We performed comparative gene analyses to trace differences in *ACE2* binding regions for *ACE2* genes in 131 mammalian species across 18 orders. We found that the *ACE2* receptors are most strongly conserved among Primates. Among other mammals, in comparison to bat species in the Chiroptera order, the human *ACE2* receptor's key residues are at least similarly or more conserved for SARS-CoV-2 binding in specific species from the Artiodactyla, Rodentia, Carnivora, Perissodactyla, Pholidota, Lagomorpha, Proboscidea, Sirenia, and Tubulidentata orders. Next, we compared sequence similarity in S protein key sites from SARS-CoV-2 and SARS-CoV with that of SARS-like coronaviruses that infect 12 high risk mammalian species, these include species from the Primates, Perissodactyla, Artiodactyla, Canivora, Chiroptera, Rodentia and Pholidota orders, but the Lagomorpha, Proboscidea, Sirenia, and Tubulidentata orders are ruled out because there are no records of mammalian-specific SARS-like coronaviruses for these species. Based on sequence similarity, we found that key residues on S protein in SARS-CoV is most similar to that of palm civet Civet-CoV, whereas that in SARS-CoV-2 is most similar to Pangolin-CoV, and both human coronaviruses share S protein similarities to bat RaTG13 but not with other mammalian coronaviruses. Together, our results suggest that that the progenitor of both SARS- CoV-2 and SARS-CoV are of bat origin. The bat coronavirus infected two different distinct species from different orders, the palm civets and pangolins, because both share high similarities at key sites with human *ACE2* and with human SARS S1 binding sites. The rapid evolution of the bat virus in the intermediate hosts allowed it to distinctly adapt high binding potential between the S protein and human-like *ACE2* receptors. Indeed, the S protein binding sites are highly similar between SARS-CoV-2 and Pangolin-CoV but not Civets-CoV, and highly similar between SARS-CoV and Civets-CoV but not Pangolin-CoV.

## Methods

### Retrieving mammalian *ACE2* genes and putative S protein binding regions

The nucleotide sequences of 296 *ACE2* gene variants from 131 mammalian species were retrieved from the National Center for Biotechnology Information (NCBI) Nucleotide Database (<https://www.ncbi.nlm.nih.gov/>). The NCBI Nucleotide Database was queried for records containing “ACE2” as a gene name and “Mammalia” as a taxonomic class, excluding whole-genome and chromosome-wide results. Next, each record was searched for /product= “angiotensin-converting enzyme 2”, and all others were removed. For each entry, only the coding DNA sequence region was extracted in FASTA format. The nucleotide CDS was translated into amino acids using DAMBE7<sup>19</sup> and verified with annotated amino acid sequences from NCBI GenBank files. Lastly, for amino acid sequence files, gene IDs were renamed as follow: ACE2\_NCBI gene accession ID\_Species name (Supplemental File S1). Similarly, the amino acid sequences of 15 coronavirus Spike proteins from 13 mammalian species were retrieved from NCBI and extracted in FASTA format.

For each *ACE2* variant and for each coronavirus spike protein, we extracted the amino acid at the *ACE2* receptor hotspots and S protein’s binding hotspots listed in Table 1. Both BLAST and site-specific amino acid identities show that most variants within species are well conserved. To compare amino acid residues at the *ACE2* hotspots between human and other mammalian species, *ACE2* amino acid sequences were first aligned with MAFFT<sup>20</sup> with the slow but accurate G-INS-i option, then a match-mismatch heat-map was generated for each site and a total similarity score (the number of matching amino acid sites between human and mammalian species) was calculated. Similarly, to compare S protein binding sites between SARS-CoV-2 and SARS-CoV to other coronaviruses, S protein amino acid sequences were first aligned with MAFFT<sup>20</sup> with the slow but accurate G-INS-i option, then site-specific amino acids were retrieved to generate the heat-map.

## Phylogenetic reconstruction

Only one *ACE2* isoform was used for each species for phylogenetic reconstruction. Many isoforms differ only in the 5' and 3' UTRs. For humans, *ACE2* transcript variant 2 (NM\_021804.3) contains 19 exons whereas transcript variant 1 (NM\_001371415.1) contains 18 exons, although both variants encode the same proteins. The same information was not documented for *ACE2* variants in other mammalian species. Furthermore, some isoforms are experimentally validated while others are predictions by automated computational approaches. We chose experimentally verified variants when available, but when only predicted variants are available, we picked one among those whose amino acid identities at *ACE2* receptors are conserved (picked *ACE2* variants are listed in Supplemental file S1). Similarly, one representative coronavirus was picked out of the several available strains (e.g., SARS-CoV Urbani). Then, the amino acid sequences for these *ACE2* variants and coronavirus Spike proteins were aligned with MAFFT<sup>20</sup> with the slow but accurate G-INS-i option.

Three phylogenetic trees were constructed using MAFFT aligned amino acids with the maximum-likelihood-based PHYML approach. One tree for aligned *ACE2* gene for 131 mammalian species (bootstrap = 500, model = JTT + G + I + F), another tree for *ACE2* gene for 13 sample mammalian species (bootstrap =100, model = JTT + G + I + F), and a third tree for 15 mammalian SARS-like coronaviruses

(bootstrap = 100, model = WAG + G + I + F). All were constructed using the PHYML model<sup>21</sup> implemented in DAMBE. The tree improvement option “-s” was set to “BEST” (best of NNI and SPR search). The “-o” option was set to “tlr” which optimizes the topology, the branch lengths and rate parameters.

## Results

### **The Human *ACE2* receptor is most conserved among Primates, but variably more conserved in species of other orders in comparison to bat species of the Chiroptera order.**

The majority of the 131 mammalian species with available *ACE2* gene records belongs to five major mammalian orders (Artiodactyla, Chiroptera, Rodentia, Carnivora, and Primates) (Fig. 2). Among which the key residues at *ACE2* receptor for binding with SARS-CoV-2 share perfect similarity between humans and closely related Primates (Fig. 2A). However, the *ACE2* key residues of species from three Primates families, *Cheirogaleidae*, *Galagidae* and *Tarsiidae*, share less similarity with that of humans.

Between human and mammalian species from other orders, the key residues on the *ACE2* receptor are variably conserved for SARS-CoV-2 S protein binding. In the Artiodactyla order (Fig. 2B), species belonging to the *Bovidae*, *Monodontidae*, *Phocoenidae*, and *Physeteridae* families share very high similarity with humans at the *ACE2* receptor whereas others do not. In the Chiroptera order (Fig. 2C), only species belonging to the *Pteropodidae* and *Rhinolophidae* share high similarity with humans at the *ACE2* receptor. In the Carnivora order (Fig. 2D), species that share high similarity with human in *ACE2* receptors are from the *Canidae*, *Felidae*, and *Ursidae* families. In the Rodentia order (Fig. 2E), most species share high similarity with humans at the *ACE2* receptor except species in the *Muridae* and *Rattus* families. Lastly, for minor orders with fewer documented *ACE2* genes (Fig. 2F), species *ACE2* receptor in all orders share poor similarities to that of humans except species in the Dasyuromorphia, Lagomorpha, Perissodactyla, Proboscidea, Pholidota, Tubulidentata, and Sirenia orders. A phylogenetic analysis by the maximum-likelihood-based PHYML approach using aligned *ACE2* amino acid sequences (Supplemental Fig. S2) showed the relationships among these mammalian species. They showed that each of the five major orders and some minor orders host distinct lineages having high similarity to humans in *ACE2* receptor's pairing potential with SARS-CoV-2 (Supplemental Fig. S2). This indicates that the risk of infection vary among mammalian species, and that species with high risk for SARS-CoV-2 infection belong to specific families of the Primates, Perissodactyla, Artiodactyla, Carnivora, Chiroptera, Rodentia, Lagomorpha, Perissodactyla, Proboscidea, Sirenia, and Tubulidentata orders.

### **The S protein receptor binding sites is distinctly conserved between SARS-CoV-2, bat RaTG13 and Pangolin-CoV, and between SARS-CoV, bat RaTG13, and Civet-CoV.**

Above results shows that mammalian lineages are subjected to different degrees of risks for SARS-CoV-2 infections. Potentially, the intermediate host for SARS-CoV-2 could be located within high risk lineages since it would be evolutionarily advantageous for the progenitor virus to first infect and adapt within an intermediate host having a *ACE2* receptor that is similar to humans, and then infect humans through

zoonotic transmission. We compared key sites in the SARS-CoV-2 and SARS-CoV S proteins (Table 1) that are involved in binding to human *ACE2* receptor in 12 high risk mammalian species from the Artiodactyla, Canivora, Chiroptera, Perissodactyla, Primates, Pholidota and Rodentia orders. However, species of the Lagomorpha, Proboscidea, Sirenia, and Tubulidentata orders are ruled out because there are no records of mammalian-specific SARS-like coronaviruses that infect species in these orders.

Indeed, figure 3 shows differences in the conservation of key sites at S protein's binding domains on SARS-CoV-2 and SARS-CoV (SARS Urbani) in comparison to other mammalian SARS-like coronaviruses. Between the two human viruses, the S protein's binding domain for SARS-CoV-2 is weakly shared by SARS-CoV (Fig. 3A), and vice versa (Fig. 3B). This suggests that the two viruses are evolutionarily distinct. Furthermore, the S protein's binding sites on SARS-CoV-2 shares highest similarities with that of Pangolin-CoV strain Guangdong (GD) that was sequenced with high coverage (Fig. 3A), but shares lower similarity with Pangolin-CoV strain Guangxi (GX) flagged with poor quality by GISAID; whereas the S protein's binding domain of SARS-CoV shares highest similarity with that Civets-CoV (Fig. 3B). Additionally, both viral S protein binding domains share a medium degree of similarity with that of the bat coronavirus RaTG13 from the species *Rhinolophus affinis*, but not with other bat SARS-like viruses from *Rhinolopus sinicus* and *Rhinolopus ferrumequinum*. Among other SARS-like coronaviruses, their S proteins share little similarities with that of both SARS-CoV-2 and SARS-CoV, including the human MERS-CoV. Together, these results suggest that both SARS-CoV-2 and SARS-CoV likely originated from the bat (Chroptera order, *Rhinolophus* family) but they have distinct intermediate hosts.

In terms of similarities in S protein binding domains, *Rhinolophus affinis* is less similar to humans than two specific species, palm civets and pangolins. While the S protein binding domains of Civets-CoV is more similar to that of SARS-CoV, the S protein binding domain of Pangolin-CoV is more similar to that of SARS-CoV-2. One additional evidence of this distinct zoonotic transmission for SARS-CoV-2 and SARS-CoV is the presence of a unique 482-485 domain that is also present in bat RaTG13, SARS-CoV-2, and Pangolin-CoV GD, but not in SARS-CoV and palm civets (Fig. 3A). Furthermore, both groups are more closely related with bat SARS-like coronaviruses than with the human MERS coronavirus with a presumed camel origin, and both human SARS coronaviruses are more distant than SARS-like coronaviruses of other mammals such as those infecting canine, bovine, camel, equine, rat, and feline (Fig. 4). More importantly, figure 4A shows that while SARS-CoV-2 is more similar to bat RaTG13 than to Pangolin-CoV in overall S protein sequence similarity, the S protein's binding domains are more similar between SARS-CoV-2 and Pangolin-CoV than to RaTG13. Hence, the S protein of Pangolin-CoV is better adapted to infect humans than is the S protein of RaTG13 even though globally SARS-CoV-2 is more similar to RaTG13. The notion that local similarities are more important than global similarities is therefore consistent for the identification of high-risk mammals (Fig. 1) and coronavirus infectivity (Fig. 4).

## Discussion

We performed comparative gene analyses to trace differences at key SARS-CoV-2 binding sites on the *ACE2* receptors in 131 mammalian species across 18 orders. Similarities in mammalian *ACE2* genes are

measured in two ways, one is by global similarity reflected by the phylogenetic tree from *ACE2* gene alignment, and the other is key site similarities to human *ACE2*. While global similarities is not a good predictor of which species is at high risk of being infected by SARS-CoV-2, local *ACE2* similarities highlight which mammals that are at high risk of being infected by SARS-CoV-2 (Fig. 1). Specifically, the human *ACE2* receptor key sites involved in S protein binding are perfectly conserved among Primates. Among other mammals, the human *ACE2* receptor sites are at least, if not more, conserved than bat (Chiroptera) *ACE2* receptor by various mammalian species from the Perissodactyla, Artiodactyla, Canivora, Rodentia, Pholidota, Lagomorpha, Proboscidea, Sirenia, and Tubulidentata orders. Because their *ACE2* receptors share high similarities to humans, these species are potentially at high risk of SARS-CoV-2 infection. Furthermore, the intermediate host for SARS-CoV-2 could be a species within these high risk lineages because it would be evolutionarily advantageous for the presumed progenitor bat SARS-like coronavirus to first infect and adapt within an intermediate host having a *ACE2* receptor that is similar to humans, and then infect humans through zoonotic transmission.

We next compared the SARS-CoV-2 and SARS-CoV S proteins with SARS-like coronaviruses in 12 high risk mammalian species from the Perissodactyla, Artiodactyla, Canivora, Rodentia, Pholidota orders. However, species from the Lagomorpha, Proboscidea, Sirenia, and Tubulidentata orders are ruled out because there are no documented SARS-like coronaviruses that infect species from these orders. Given that distinct key sites on the S proteins of SARS-CoV-2 and SARS-CoV are important for human *ACE2* binding, we wished to know which high risk mammals host SARS-like coronaviruses have shared key sites with the two human SARS coronaviruses. We found that key residues on the S protein receptor in SARS-CoV is most shared by Civet-CoV, similar to Pangolin-CoV, weakly shared by bat RaTG13 and by SARS-CoV-2, but not shared by other mammalian-specific viruses. Whereas key residues on S protein receptor in SARS-CoV-2 are most shared by Pangolin-CoV GD, less shared by Pangolin-CoV GX and bat RaTG13, weakly shared by Civet-CoV and SARS-CoV, but not shared by other mammalian-specific viruses (Fig. 3). These findings corroborate a recent study<sup>22</sup> that found that the entire SARS-CoV-2 S1 receptor-binding domain from residues 451 to 509 share higher similarity to Pangolin-CoV GD than to bat RaTG13 and SARS-CoV.

Because the S proteins of human SARS coronaviruses and of bat SARS-like coronavirus share close phylogenetic relationships (Fig. 4), this suggest that both SARS-CoV-2 and SARS-CoV are indeed likely to have a bat origin<sup>2,23</sup>, possibly one from the *Rhinolophidae* family. However, direct transmission from bats to humans is unlikely because 1) key residues on bats *ACE2* receptors are less similar to humans (Fig. 2C) than those on *ACE2* receptors of mammalian species from other orders (Fig. 2), and 2) key residues on the bat SARS-like coronavirus S protein are less similar to those on human SARS coronaviruses than those on other mammalian-specific SARS-like coronaviruses (Fig. 3). Indeed, our results suggests that the intermediate host for both viruses are likely distinct for SARS-CoV-2 and SARS-CoV, namely the previously proposed palm civets for SARS-CoV<sup>4,12</sup> and pangolins for SARS-CoV-2<sup>17,24,25</sup>, respectively, due to their high similarities in both coronavirus S proteins and *ACE2* receptors to that of humans. Importantly, based on global similarities, the S protein of SARS-CoV-2 is more closely related to bat RaTG13 than it is to Pangolin-CoV (Fig. 4), similar to what recent studies have shown<sup>6,25</sup>. However, here we showed that key

binding sites at the S protein that are involved in *ACE2* receptor binding are more alike between SARS-CoV-2 and Pangolin-CoV than RaTG13. Hence, the S protein of Pangolin-CoV is better adapted to human *ACE2* receptor than that of RaTG13 even though SARS-CoV-2 is more similar to RaTG13 globally. The notion that key site differences are more important than global similarities is therefore important in the determination of both high-risk mammals (Fig. 1) and S protein infectivity (Fig. 4). Hence, rapid evolution of the bat virus in the intermediate hosts allowed it to adapt higher binding potential between the S protein on the virus and mammalian *ACE2* receptors that are similar to human *ACE2* receptor, leading to evolutionary differences between SARS-CoV-2 and SARS-CoV S proteins. The corroborative evidence for this proposition is the observed similarities in S protein binding receptors between SARS-CoV-2 and Pangolin-CoV but not Civets-CoV, and similarities between SARS-CoV and Civets-CoV but not Pangolin-CoV. One additional evidence of this differential adaptation is the presence of the S protein 482-485 motif that allows favorable human *ACE2* recognition<sup>7</sup> in SARS-CoV-2, bat RaTG13, and Pangolin-CoV, but not in SARS-CoV or Civets-CoV (Fig. 3A).

## Declarations

### ACKNOWLEDGEMENTS

This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant to X.X. [RGPIN/2018-03878], and NSERC Doctoral Scholarship to Y.W. [CGSD/2019-535291].

### DATA AVAILABILITY

Supplementary file S1 contains *ACE2* and S protein gene accessions, species used for phylogenetic reconstruction, and amino acid site-specific comparisons between mammalian species at the *ACE2* and S protein binding domains. Supplemental Fig. S1. contains key site comparisons between human *ACE2* and *ACE2* from 131 mammalian species in binding to SARS-CoV. Supplemental Fig. S2 contains the phylogenetic relationship detailing global similarities between 131 mammalian *ACE2* genes.

Supplemental File S2 contain Supplemental figure legends.

### AUTHOR CONTRIBUTIONS

Y.W. and X.X. designed the study and wrote the manuscript. Y.W., P.A., and H.F. collected the data and

Y.W. and P.A. analyzed the data. P.A. prepared all figures. X.X. supervised the study. All authors reviewed the manuscript. X.X. supervised the project.

### COMPETING INTERESTS

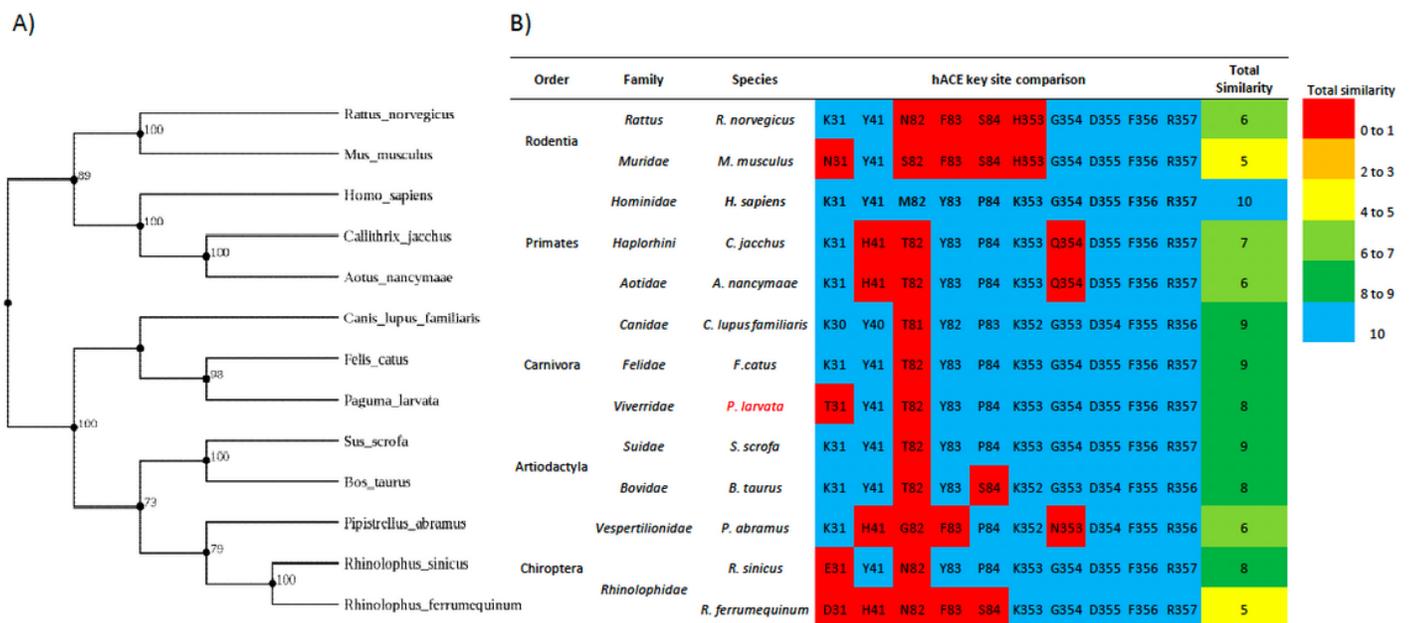
The authors declare no competing interests.

# References

1. Dong, E., Du, H. & Gardner, L. *An interactive web-based dashboard to track COVID-19 in real time.* (Lancet Infect Dis. 2020 Feb 19:S1473-3099(20)30120-1. doi: 10.1016/S1473- 3099(20)30120-1.).
2. Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* **17**, 181-192 (2019).
3. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864-1868 (2005).
4. Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human *Embo J* **24**, 1634-1643 (2005).
5. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).
6. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, doi:10.1038/s41586-020-2180-5 (2020).
7. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **30**, 020-2179 (2020).
8. Tian, X. *et al.* *Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus- specific human monoclonal antibody.* (Emerg Microbes Infect. 2020 Feb 17;9(1):382-385. doi: 10.1080/22221751.2020.1729069. eCollection ).
9. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike *Cell* **181**, 281-292 (2020).
10. Hamming, I. *et al.* Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol* **203**, 631-637 (2004).
11. Nie, Y. *et al.* Highly infectious SARS-CoV pseudotyped virus reveals the cell tropism and its correlation with receptor expression. *Biochem Biophys Res Commun* **321**, 994-1000 (2004).
12. Guan, Y. *et al.* Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276-278 (2003).
13. Marra, M. A. *et al.* The Genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399-1404 (2003).
14. Rota, P. A. *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394-1399 (2003).
15. Song, H. D. *et al.* Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci U S A* **102**, 2430-2435 (2005).
16. Gralinski, L. E. & Menachery, V. D. Return of the Coronavirus: 2019-nCoV. *Viruses* **12** (2020).
17. Xiao, K. *et al.* Isolation and Characterization of 2019-nCoV-like Coronavirus from Malayan Pangolins. *bioRxiv*, 2020.2002.2017.951335, doi:10.1101/2020.02.17.951335 (2020).
18. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *Journal of*

19. Xia, X. DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and *Mol Biol Evol* **35**, 1550-1552 (2018).
20. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol* **537**, 39-64 (2009).
21. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321 (2010).
22. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nature Medicine* **26**, 450-452, doi:10.1038/s41591-020-0820-9 (2020).
23. Perlman, S. *Another Decade, Another Coronavirus.* (N Engl J Med. 2020 Feb 20;382(8):760-762. doi: 10.1056/NEJMe2001126. Epub 2020 Jan ).
24. Lam, T.-Y. *et al.* Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. *bioRxiv*, 2020.2002.2013.945485, doi:10.1101/2020.02.13.945485 (2020).
25. Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current Biology* **30**, 1346-1351.e1342, doi:https://doi.org/10.1016/j.cub.2020.03.022 (2020).

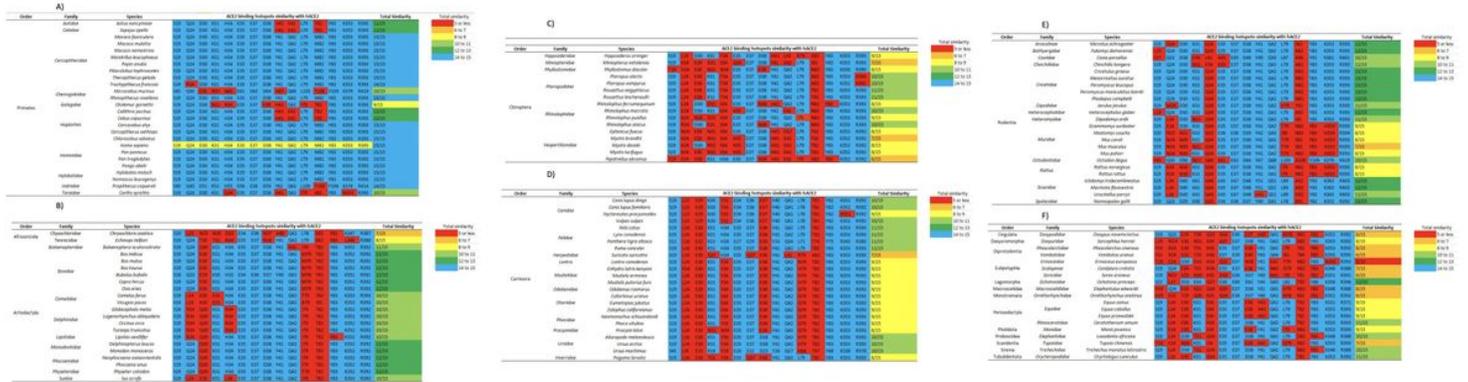
## Figures



**Figure 1**

A sample comparison between mammalian ACE2 genes. (A) ACE2 global similarities is not a good predictor of which species are at high risk of being infected by SARS-CoV. Whereas (B) local similarities at key ACE2 binding sites show that species in the Carnivora, Artiodactyla and Chiroptera orders are at more

risk of SARS-CoV infection than species in Primates and Rodentia order that shared higher global similarity with human ACE2 gene. In red is masked palm civet, the presumed intermediate host of SARS-CoV. Total similarity in (B) is calculated by the number of matching amino acids between mammalian and human ACE2 at key binding sites. The complete key site comparisons between 131 mammalian ACE2 genes for SARS-CoV binding are available in Supplemental Fig. S1.



**Figure 2**

The 15 key sites on human ACE2 receptor involved in SARS-CoV-2 S protein binding (see Table 1) are compared with the ACE2 receptors of 131 mammalian species belonging to A) Primates, B) Artiodactyla, C) Chiroptera, D) Canivora, E) Rodentia, and F) 13 other orders. All ACE2 genes are aligned in MAFFT. At the site-specific residues (ACE2 binding hotspots similarity with hACE2), in blue and red are matching and mis-matching amino acids, respectively, between mammalian ACE2 and human ACE2. At total similarity, highlighted in blue are mammalian ACE2 with perfect site similarity (14 to 15 matching sites), in green high similarity (12 to 13 matching sites), in light green medium similarity (10 to 11 matching sites), in yellow medium low similarity (8 to 9 matching sites), in orange medium low similarity (6 to 7 matching sites), and in red low similarity (5 or less matching sites) when compared to human ACE2 receptor.

A)

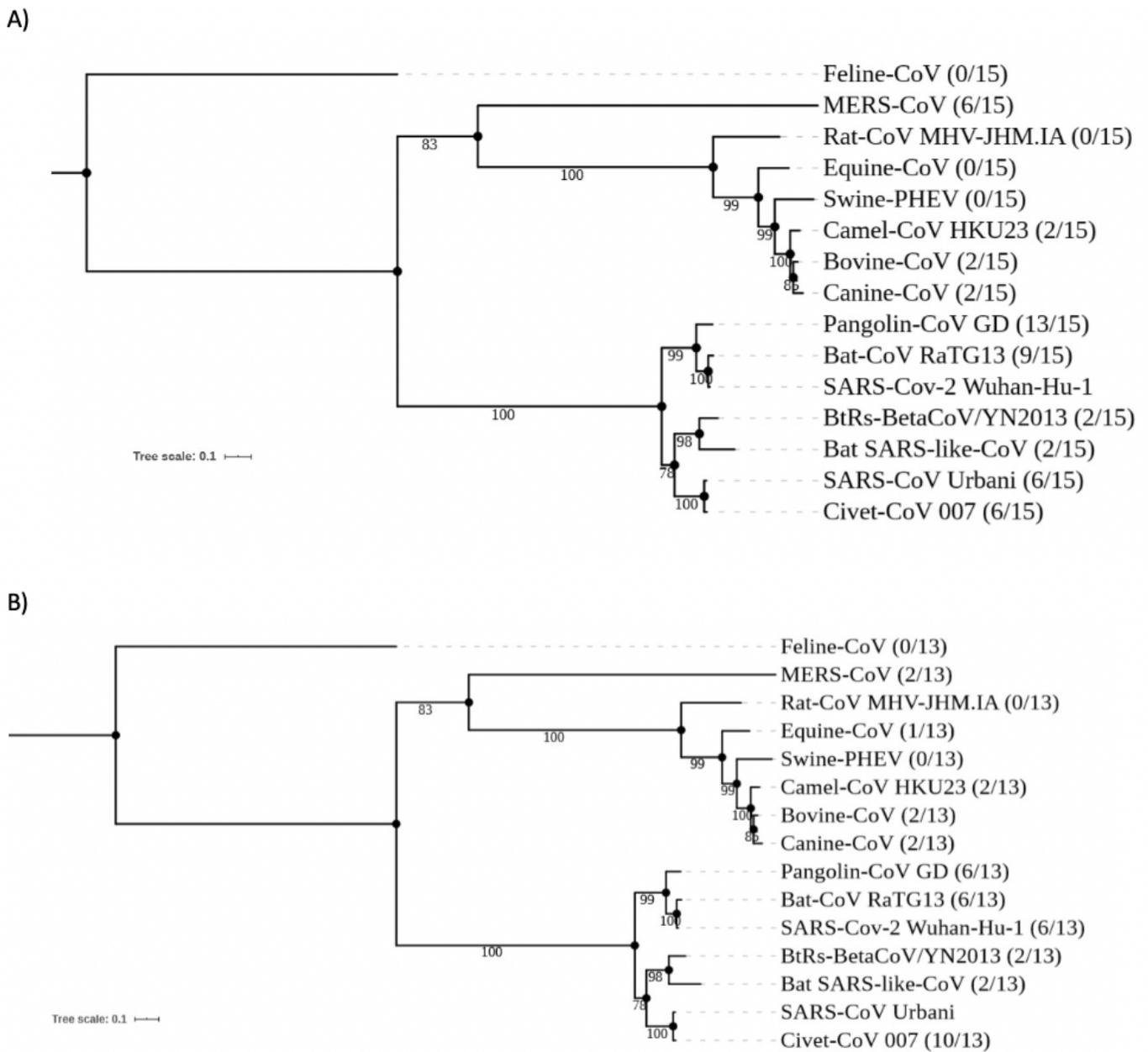
Genus	S protein accession	Species	Aligned Spike protein amino acid sites																Total Similarity	total similarity			
Betacoronavirus	YP_009724390 (SARS-Cov-2 Wuhan-Hu-1)	Human (Homo sapiens)	R417	G446	Y449	L455	F456	A475	G482	V483	E484	G485	F486	N487	Y489	Q493	Q498	T500	N501	G502	Y505	14/15	14 to 15
Betacoronavirus	EPLI5L_410721 (Pangolin-CoV GD)	Pangolin (Manis javanica) Guangdong	R418	G442	Y445	L451	F452	A477	G478	V479	E480	G481	F482	N483	Y485	Q489	Q494	T496	N497	G498	Y501	10/13	10 to 11
Betacoronavirus	QIG54048.1 (PangolinCoV GX_P2V)	Pangolin (Manis javanica) Guangxi	R418	G446	Y449	L455	F456	A477	G482	Q483	V484	G485	F486	N487	Y489	Q493	Q498	T500	N501	G502	Y505	9/15	9 to 9
Betacoronavirus	QIFR63300.2 (RaTG13)	Bat (Rhinolophus affinis)	R415	G446	Y449	L455	F456	A477	G482	Q483	T484	G485	L486	N487	Y489	Q493	Q498	T500	D501	G502	H505	9/15	8 to 9
Betacoronavirus	AAP13441.1 (SARS Urbani)	Human (Homo sapiens)	R426	Y439	Y449	L455	F456	A477	G482	Q483	T484	G485	L486	N487	Y489	Q493	Q498	T500	N501	G502	Y505	6/13	6 to 7
Betacoronavirus	AAU04646.1 (Civet SARS CoV 007)	Palm civet (Paguma larvata)	R426	Y439	Y449	L455	F456	A477	G482	Q483	T484	G485	L486	N487	Y489	Q493	Q498	T500	N501	G502	Y505	6/13	6 to 7
Betacoronavirus	QBM11748.1 (MERS)	Human (Homo sapiens)	R412	S441	Y448	F454	T455	F480	L491	T519	N520	Y521	L522	T523	Y525	T540	Q580	Q582	A583	F584	W587	2/15	5 or less
Betacoronavirus	AT098145.1 (Bat SARS-like-CoV)	Bat (Rhinolophus ferrumequinum)	R410	S441	Y448	F454	T455	F480	L491	T519	N520	Y521	L522	T523	Y525	T540	Q580	Q582	A583	F584	W587	2/15	5 or less
Betacoronavirus	AIA62330.1 (BtRS-BetaCoV/YN2013)	Bat (Rhinolophus sinicus)	R410	S441	Y448	F454	T455	F480	L491	T519	N520	Y521	L522	T523	Y525	T540	Q580	Q582	A583	F584	W587	2/15	5 or less
Betacoronavirus	ALAS0080.1 (HKU23)	Camel (Camelus dromedarius)	R412	S441	Y448	F454	T455	F480	L491	T519	N520	Y521	L522	T523	Y525	T540	Q580	Q582	A583	F584	W587	2/15	5 or less
Betacoronavirus	AV261123.1 (Bovine CoV)	Cattle (Bos taurus)	R412	S441	Y448	F454	T455	F480	L491	T519	N520	Y521	L522	T523	Y525	T540	Q580	Q582	A583	F584	W587	2/15	5 or less
Betacoronavirus	BAJ52885.1 (Equine CoV)	Horse (Equus caballus)	R412	S441	Y448	F454	T455	F480	L491	T519	N520	Y521	L522	T523	Y525	T540	Q580	Q582	A583	F584	W587	2/15	5 or less
Betacoronavirus	ACN89763.1 (murineCoV MHV-JHM.1A)	Rat (Mus musculus)	R410	S441	Y448	F454	T455	F480	L491	T519	N520	Y521	L522	T523	Y525	T540	Q580	Q582	A583	F584	W587	0/15	3 or less
Betacoronavirus	AVV64341.1 (PHEV)	Swine (Sus scrofa)	R432	W446	F452	N453	R508	E509	S511	D540	A581	D583	S584	F585	W588	0/13	3 or less						
Betacoronavirus	AFW97360.1 (CRCoV)	Canine (Lupus familiaris)	R432	W446	F454	T455	F480	L491	T519	N520	Y521	L522	T523	Y525	T540	Q580	Q582	A583	F584	W587	2/15	5 or less	
Alphacoronavirus	AFH58021.1 (Feline CoV)	Feline	R426	Y439	Y449	L455	F456	A477	G482	Q483	T484	G485	L486	N487	Y489	Q493	Q498	T500	N501	G502	Y505	9/15	6 to 7

B)

Genus	S protein accession	Species	Aligned Spike protein amino acid sites																Total Similarity	total similarity
Betacoronavirus	AAP13441.1 (SARS Urbani)	Human (Homo sapiens)	R426	Y436	Y442	L443	L472	N473	Y475	N479	Y484	T486	T487	G488	Y491	10/13	10 to 11			
Betacoronavirus	AAU04646.1 (Civet SARS CoV 007)	Palm civet (Paguma larvata)	R426	Y436	Y442	L443	R472	N473	Y475	R479	Y484	T486	S487	G488	Y491	8/13	8 to 9			
Betacoronavirus	QIQ54048.1 (PangolinCoV GX_P2V)	Pangolin (Manis javanica) Guangxi	Y439	Y449	L455	F456	L486	N487	Y489	E493	H498	T500	T501	G502	Y505	6/13	6 to 7			
Betacoronavirus	YP_009724390 (SARS-Cov-2 Wuhan-Hu-1)	Human (Homo sapiens)	R439	Y449	L455	F456	F486	N487	Y489	Q493	Q498	T500	N501	G502	Y505	6/13	6 to 7			
Betacoronavirus	EPI_ISL_410721 (Pangolin-CoV GD)	Pangolin (Manis javanica) Guangdong	R435	Y445	L451	F452	F482	N483	Y485	Q489	H494	T496	N497	G498	Y501	6/13	6 to 7			
Betacoronavirus	QHR63300.2 (RaTG13)	Bat (Rhinolophus affinis)	R439	F449	L455	F456	L486	N487	Y489	Y493	Y498	T500	D501	G502	H505	6/13	6 to 7			
Betacoronavirus	AIA62330.1 (BtRS-BetaCoV/YN2013)	Bat (Rhinolophus sinicus)	A422	S433	H434	N451	V453	S457	N462	N464	V465	P466	Y469	2/13	3 or less					
Betacoronavirus	AT098145.1 (Bat SARS-like-CoV)	Bat (Rhinolophus ferrumequinum)	A423	S434	H435	N452	V454	S458	N463	N465	V466	P467	Y470	2/13	3 or less					
Betacoronavirus	AVZ61129.1 (Bovine CoV)	Cattle (Bos taurus)	R434	W448	F454	Y455	L522	T523	Y525	T540	Q580	Q582	A583	F584	W587	2/13	3 or less			
Betacoronavirus	QBM11748.1 (MERS)	Human (Homo sapiens)	R485	P494	L495	R542	K543	L545	E549	L554	A556	S557	G558	V561	2/13	3 or less				
Betacoronavirus	ALAS0080.1 (HKU23)	Camel (Camelus dromedarius)	R434	W448	F454	F455	L522	T523	Y525	T540	Q583	Q585	A586	F587	W590	2/13	3 or less			
Alphacoronavirus	AFW97360.1 (CRCoV)	Canine (Lupus familiaris)	R434	W448	F454	T455	L522	T523	Y525	L540	Q580	K582	A583	F584	W587	2/13	3 or less			
Betacoronavirus	AVV64341.1 (PHEV)	Swine (Sus scrofa)	R434	W448	F454	N455	R504	K505	F507	T526	K566	Q568	A569	F570	W573	0/13	3 or less			
Betacoronavirus	ACN89763.1 (murineCoV MHV-JHM.1A)	Rat (Mus musculus)	R432	W446	F452	N453	R508	E509	S511	D540	A581	D583	S584	F585	W588	0/13	3 or less			
Betacoronavirus	BAJ52885.1 (Equine CoV)	Horse (Equus caballus)	G434	W448	F454	N455	R521	T522	F524	N539	P580	N582	A583	F584	W587	1/13	3 or less			
Alphacoronavirus	AFH58021.1 (Feline CoV)	Feline	R488	N502	M540	K541	Q594	E595	T597	D601	K639	D641	V642	A643	T646	0/13	3 or less			

Figure 3

MAFFT aligned amino acid comparisons at key binding residues (see Table 1) between 14 SARS-like coronavirus S proteins and A) SARS-CoV-2 S protein and B) SARS-CoV (SARS Urbani) S protein (16 virus in total). In blue and red are matching and mis-matching amino acid residues, respectively. Total similarity designates the total number of matching amino acid residues, highlighted in blue are S protein binding domains with perfect site similarity (14 to 15, and 12 to 13 matching sites), in green high similarity (12 to 13, and 10 to 11 matching sites), in light green medium similarity (10 to 11, and 8 to 9 matching sites), in yellow medium low similarity (8 to 9, and 6 to 7 matching sites), in orange medium low similarity (6 to 7, and 4 to 5 matching sites), and in red low similarity (5 or less, and 3 or less matching sites) when other SARS-like coronavirus S proteins are compared to the S protein of SARS-CoV-2 and SARS-CoV, respectively. In yellow is the 482-485 GVEG motif that is distinctly found in SARS-CoV-2.



**Figure 4**

Phylogenetic reconstruction using 15 MAFFT aligned SARS and SARS-like coronaviruses infecting mammalian species (of the Artiodactyla, Canivora, Chiroptera, Perissodactyla, Primates, Pholidota and Rodentia orders), using the maximum-likelihood-based PHYML approach, with best model = WAG + G + I + F, and Bootstrap = 100. Pangolin-CoV GX strain was excluded in favor of Pangolin-CoV GD strain having better coverage. The ratio appended to each species at the end indicates the total similarity (shown in Fig. 3) to A) 15 key S protein sites in SARS-CoV-2 and B) 13 key S protein sites in SARS-CoV.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformation.pdf](#)