



OPEN

## Predicting mammalian species at risk of being infected by SARS-CoV-2 from an ACE2 perspective

Yulong Wei<sup>1</sup>, Parisa Aris<sup>1</sup>, Heba Farookhi<sup>1</sup> & Xuhua Xia<sup>1,2</sup>✉

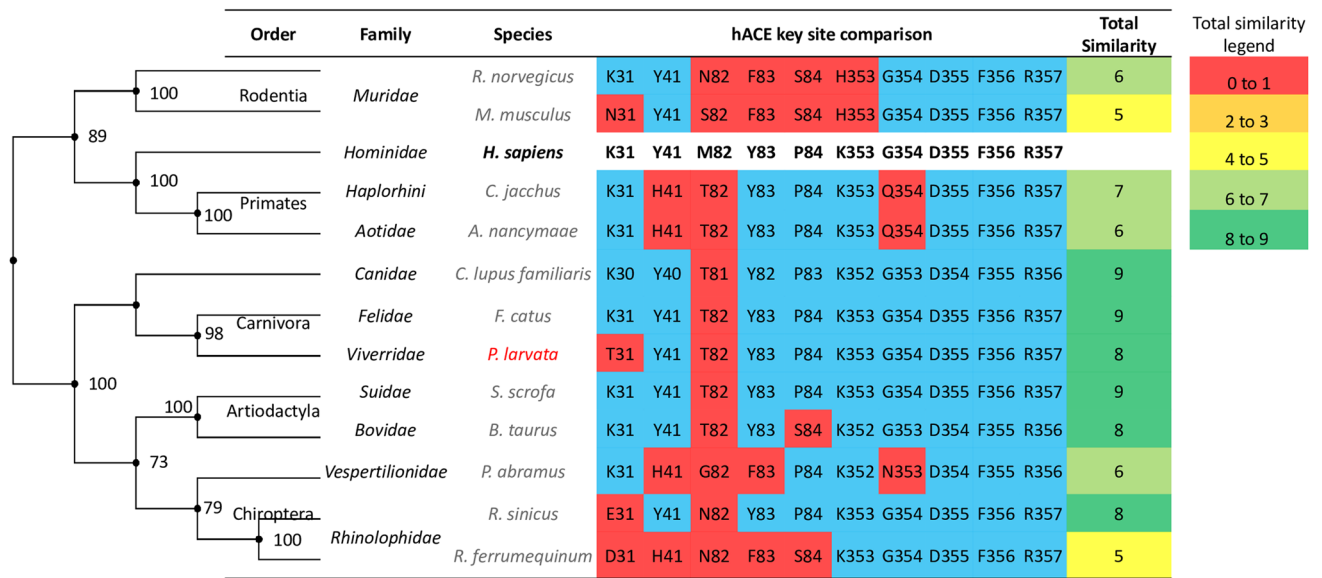
SARS-CoV-2 can transmit efficiently in humans, but it is less clear which other mammals are at risk of being infected. SARS-CoV-2 encodes a Spike (S) protein that binds to human ACE2 receptor to mediate cell entry. A species with a human-like ACE2 receptor could therefore be at risk of being infected by SARS-CoV-2. We compared between 132 mammalian ACE2 genes and between 17 coronavirus S proteins. We showed that while global similarities reflected by whole ACE2 gene alignments are poor predictors of high-risk mammals, local similarities at key S protein-binding sites highlight several high-risk mammals that share good ACE2 homology with human. Bats are likely reservoirs of SARS-CoV-2, but there are other high-risk mammals that share better ACE2 homologies with human. Both SARS-CoV-2 and SARS-CoV are closely related to bat coronavirus. Yet, among host-specific coronaviruses infecting high-risk mammals, key ACE2-binding sites on S proteins share highest similarities between SARS-CoV-2 and Pangolin-CoV and between SARS-CoV and Civet-CoV. These results suggest that direct coronavirus transmission from bat to human is unlikely, and that rapid adaptation of a bat SARS-like coronavirus in different high-risk intermediate hosts could have allowed it to acquire distinct high binding potential between S protein and human-like ACE2 receptors.

The *Betacoronavirus* 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 20 million confirmed cases of COVID-19 worldwide (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>, last accessed August 25, 2020). While it is evident that SARS-CoV-2 can transmit efficiently from person to person, it is less clear which other mammalian species are at high risk of being infected. The answer to this question is important to (1) improve our ability to predict and control future pandemics, and (2) manage and protect wildlife and domesticated animals.

**Mammals at high risk of SARS-CoV-2 infection should have human-like ACE2 receptors at key S protein-binding sites.** Both SARS-CoV and SARS-CoV-2 genomes encode a Spike (S) protein that binds to human Angiotensin-converting enzyme 2 (ACE2) receptor to mediate viral entry into the host cell<sup>1–6</sup>. Mechanistically, two S protein domains are involved during coronavirus infection in mammalian cells that express ACE2: the S1 domain interacts with the ACE2 receptor<sup>7,8</sup>, and the S2 domain undergoes structural rearrangements to mediate membrane fusion<sup>4</sup>. Interacting with the S1 domain, the S protein-binding sites on ACE2 receptor are primarily located in the  $\alpha$ -helix 1 and  $\beta$ -sheet 5 domains<sup>4</sup>. The efficacy of the interaction between the S protein and the ACE2 receptor is a good predictor of the severity of coronavirus infection<sup>4,9,10</sup>. For example, a potential source of the SARS-CoV outbreak was the masked palm civets<sup>11,12</sup>. Palm civet ACE2 receptors bind efficiently to S proteins of Civet-CoV strain SZ3 isolated from infected palm civets, to S proteins of SARS-CoV strain TOR2 that caused the severe 2002–2003 outbreak, and to S proteins of SARS-CoV strain GD that caused the mild 2003–2004 outbreak<sup>11–14</sup>. Whereas in human cells, ACE2 receptors bind efficiently to S proteins of the severe SARS-CoV strain TOR2 but do not bind efficiently to S proteins of Civet-CoV strain SZ3 or of the less severe SARS-CoV strain GD<sup>4</sup>. Indeed, the binding potential between viral S protein and host ACE2 receptor is a key determinant of viral infectivity.

The binding potential between viral S protein and host ACE2 receptor is attributed to several key binding sites. Differences at key S protein-binding sites between mammalian ACE2 receptors explained why SARS-CoV

<sup>1</sup>Department of Biology, University of Ottawa, 30 Marie Curie, Station A, P.O. Box 450, Ottawa, ON K1N 6N5, Canada. <sup>2</sup>Ottawa Institute of Systems Biology, Ottawa, ON K1H 8M5, Canada. ✉email: Xuhua.Xia@uottawa.ca



**Figure 1.** A sample comparison between mammalian ACE2 genes. Global similarities represented by whole ACE2 gene comparisons (the phylogenetic tree on the left) poorly predict mammals at high risk of being infected by SARS-CoV; whereas local similarities at key human ACE2 sites (the table on the right) show that species in the Carnivora, Artiodactyla and Chiroptera orders are at high risk of SARS-CoV infection. The tips of the phylogenetic tree are aligned with the corresponding species listed in the table. Species in red is the masked palm civet, a possible intermediate host of SARS-CoV. Species in black bold is human, and all other species are in grey. Amino acids highlighted in blue and red are matching and mis-matching sites, respectively, between mammalian ACE2 and human ACE2 at key SARS-CoV S protein binding sites. Total similarity in the table is calculated as the sum of matching amino acids between mammalian and human ACE2 at key binding sites. A comparison of key ACE2 sites involved in SARS-CoV binding between human and 131 mammals is shown in Supplementary Figure S3.

efficiently infected humans and palm civets but not rats, and introducing point mutations in the ACE2 gene variably affected the binding potential between ACE2 receptor and SARS-CoV S protein in these species<sup>4</sup>. For example, experimentally mutating rat His353 into human Lys353 turned the rat ACE2 receptor from one that poorly binds S protein into one that is efficient for binding, introducing amino acid residues 82–84 from human ACE2 into rat ACE2 also led to an increase in S protein-binding potential, but replacing human ACE2 Met82 to rat ACE2 Asn82 partially inhibited S protein-binding. Furthermore, mutating human ACE2 at Lys31, Tyr41, Asp355, and Arg357 also interfered with S protein-binding. On the contrary, changes to other human ACE2 sites such as Gly354 (corresponding to Asp354 in palm civet) and amino acid residues 90–93 that potentially interact with S protein residue 479 did not affect the efficacy of S protein-binding<sup>10</sup>. In brief, gene mutation experiments<sup>3,4</sup> showed that the conservation of Lys31 and Tyr41 on α-helix 1, residues 82–84 in the vicinity of α-helix 3, and residues 353–357 on β-sheet 5 in human ACE2 receptor are crucial for SARS-CoV infectivity because their replacements weakened the binding potential between SARS-CoV S protein and ACE2 receptor.

To showcase the importance of contrasting between mammalian ACE2 genes at key S protein-binding sites, Fig. 1 compares ACE2 genes in a sample of mammalian species in two ways: the global similarities among whole ACE2 gene alignments as reflected by the mammalian phylogenetic relationships on the left, and the local similarities at key ACE2 sites that are involved in SARS-CoV S protein-binding<sup>3,4</sup> in the table on the right. Figure 1 shows that human ACE2 shares higher global similarity with ACE2 of species from Primates and Rodentia orders than with ACE2 of species from Carnivora, Artiodactyla, and Chiroptera orders. This emphasizes that a phylogenetic relationship based on whole ACE2 gene comparisons is not a good predictor of mammals at high risk of being infected by SARS-CoV for the following three reasons. First, it does not identify the *Rhinolophus* bats as a potential reservoir for the progenitor virus of SARS-CoV as previously postulated<sup>15</sup>. Second, it does not reveal masked palm civets as a potential intermediate host<sup>11,12</sup>. Third, it misidentifies the rat as a high-risk mammal while experimental evidence suggests SARS-CoV poorly infects the rat<sup>4</sup>. In contrast, comparisons of ACE2 genes at key SARS-CoV S protein-binding sites show that species in the Carnivora, Artiodactyla, and Chiroptera orders share highest local similarities with human. These high-risk mammals include the *Rhinolophus* bat and masked palm civets, even though they share less global ACE2 similarity with human than Primates and Rodentia species. Hence, comparing key binding sites on ACE2 may provide insights into the host range of viral infection.

The above observations allow for the prediction that high-risk mammals of the newly emerging SARS-CoV-2 should have human-like ACE2 receptors at key SARS-CoV-2 S protein-binding sites. Current research on SARS-CoV-2 infection in animals is limited. Nonetheless, SARS-CoV-2 infection has been detected in the cat<sup>16</sup>, dog<sup>17</sup>, mink<sup>18</sup>, and tiger<sup>19</sup>. In addition, ferret<sup>20,21</sup>, macaque<sup>22,23</sup>, and grivet<sup>24</sup> could be experimentally infected by SARS-CoV-2. We therefore expect the ACE2 genes of these known high-risk mammals to bear a high resemblance to human ACE2 at key S protein-binding sites. Moreover, two Rodentia species have been experimentally infected by SARS-CoV-2: the golden Syrian hamster (*Mesocricetus auratus*) belonging to the *Cricetidae* family, and the

Human ACE2	SARS-CoV-2	Human ACE2	SARS-CoV
<i>S19</i> <sup>a</sup>	A475 <sup>a</sup>	<b>Q24</b> <sup>b</sup>	N473 <sup>b</sup>
<b>Q24</b> <sup>b</sup>	N487 <sup>b</sup>	<b>K31</b> <sup>a,b</sup>	Y442 <sup>a,b</sup>
<b>Q24</b> <sup>b</sup> , <b>M82</b> <sup>a,b</sup> , <b>L79</b> <sup>a,b</sup> , <b>Y83</b> <sup>a,b</sup>	F486 <sup>a,b</sup>	<b>H34</b> <sup>b</sup>	L443 <sup>b</sup> , N479 <sup>b</sup>
<i>D30</i> <sup>b</sup>	K417 <sup>b</sup>	<b>E37</b> <sup>b</sup>	Y491 <sup>a,b</sup>
<b>K31</b> <sup>a,b</sup>	L455 <sup>a,b</sup> , Q493 <sup>b</sup>	<b>D38</b> <sup>a,b</sup>	Y436 <sup>b</sup> , Y484 <sup>b</sup>
<b>H34</b> <sup>b</sup> , <i>E35</i> <sup>a,b</sup>	F456 <sup>b</sup> , Q493 <sup>a,b</sup>	<b>Y41</b> <sup>b</sup>	Y484 <sup>b</sup> , T486 <sup>b</sup> , T487 <sup>b</sup>
<i>E37</i> <sup>b</sup> , <i>R393</i> <sup>b</sup>	Y505 <sup>b</sup>	<b>Q42</b> <sup>b</sup>	Y436 <sup>b</sup> , Y484 <sup>b</sup>
<b>D38</b> <sup>a,b</sup>	Y449 <sup>b</sup> , Q498 <sup>b</sup>	<b>L79</b> <sup>b</sup> , <b>M82</b> <sup>a,b</sup>	L472 <sup>a,b</sup> ,
<b>Y41</b> <sup>b</sup>	Q498 <sup>b</sup> , T500 <sup>b</sup> , N501 <sup>b</sup>	<b>Y83</b> <sup>b</sup>	Y475 <sup>b</sup> , N473 <sup>b</sup>
<b>Q42</b> <sup>b</sup>	G446 <sup>b</sup> , Y449 <sup>b</sup> , Q498 <sup>b</sup>	Q325 <sup>b</sup> , E329 <sup>b</sup>	R426 <sup>b</sup>
<b>Y83</b> <sup>b</sup>	Y489 <sup>b</sup> , N487 <sup>b</sup>	N330 <sup>b</sup>	T486 <sup>b</sup>
<b>K353</b> <sup>a,b</sup>	N501 <sup>a</sup> , G502 <sup>b</sup>	<b>K353</b> <sup>a,b</sup>	T487 <sup>a</sup> , G488 <sup>b</sup>

**Table 1.** Contacting sites at the interface of ACE2-S protein of SARS-CoV-2 (column 1, 2) and of SARS-CoV (column 3, 4). Data are retrieved from X-ray crystallography experiments by Shang et al.<sup>5</sup> and Lan et al.<sup>2</sup>. Each row specifies the site-specific contacts formed between human ACE2 and S protein (e.g., first row: ACE2 S19 binds SARS-CoV-2 S protein A475). In italics are key ACE2 sites that distinctly interact with SARS-CoV-2 or SARS-CoV S protein sites. In bold are ACE2 sites that interact with both SARS-CoV-2 and SARS-CoV S proteins at different ACE2-binding sites. The structural and chemical bond evidence for all listed interactions were described in Shang et al.<sup>5</sup> and Lan et al.<sup>2</sup> <sup>a</sup>Interacting residues retrieved from Shang et al.<sup>5</sup> <sup>b</sup>Interacting residues retrieved from Lan et al.<sup>2</sup>

house mouse (*Mus musculus*) belonging to the *Muridae* family. While hamsters could be consistently infected by SARS-CoV-2<sup>25</sup>, wild-type mice were not susceptible to SARS-CoV-2 infection<sup>26</sup>. It follows that these two Rodentia species should have distinct differences at the ACE2 receptor, and we expect key ACE2 sites in human to be well conserved by hamster ACE2 but not by mouse ACE2. Sequence comparisons between human ACE2 receptor and mammalian ACE2 receptors at key S protein-binding sites may provide crucial insights into the identification of mammals at high risk of SARS-CoV-2 infection.

Human ACE2 receptors bind to S proteins on SARS-CoV and on SARS-CoV-2 with overall structural similarity; this is likely constrained by the structure of ACE2<sup>2</sup>. However, the S protein-coding genes are notably different between SARS-CoV and SARS-CoV-2 with about 75% homology<sup>6,27</sup>. This dissimilarity may contribute to a difference in viral infectivity and host range between the two SARS viruses. Indeed, based on two recent X-ray crystallography experiments<sup>2,5</sup>, different key residues on human ACE2 receptor are structurally involved in binding to S proteins of SARS-CoV-2 and SARS-CoV. Table 1 summarizes all experimentally identified site-specific interactions at the ACE2-S protein interface<sup>2,5</sup>. Four key human ACE2 sites bind to S protein of SARS-CoV-2 but they do not interact with S protein of SARS-CoV: ACE2 Ser19 forms a hydrogen bond with S protein Ala475, ACE2 Asp30 forms a hydrogen bond with S protein at Lys417, ACE2 Glu35 forms a hydrogen bond with S protein Gln493, and ACE2 Arg393 forms a hydrogen bond with S protein Tyr505. Furthermore, there are 11 other key human ACE2 sites (Gln24, Lys31, His34, Glu37, Asp38, Tyr41, Gln42, Leu79, Met82, Tyr83, and Lys353) that form either hydrogen bonds or salt bridges with S proteins of both SARS-CoV-2 and SARS-CoV, but they interact with different S protein residues. For example, ACE2 Gln24 forms a hydrogen bond with SARS-CoV-2 S protein at Leu472 but with SARS-CoV S protein at Asn473, and ACE2 Tyr41 forms a hydrogen bond with SARS-CoV-2 S protein at Thr500 and Asn501 but with SARS-CoV S protein at Thr486 and Thr487. The extended data in Lan, et al.<sup>2</sup> also highlights ACE2 sites Thr27, Phe28, Asn330, Gly354 and Asp355 to be in proximity with SARS-CoV-2 S protein. However, these residues were excluded from analysis because chemical bond evidence was not shown at these sites. In brief, we analyzed all key interacting residues (listed in Table 1) whose structures were shown and whose chemical bond evidences were reported, by one or both X-ray crystallography experiments<sup>2,5</sup> with structure resolution cut-off < 3 Å and interface distance cut-off < 5 Å. Nonetheless, it is worth noting that future gene mutation studies should be performed to verify whether all 15 key ACE2 sites are essential for SARS-CoV-2 infection.

### Species-specific coronaviruses infecting high-risk mammals may express SARS-like S proteins at key ACE2-binding sites.

Next, we wish to know which mammalian-specific coronaviruses are similar to SARS-CoV-2 and to SARS-CoV in infectivity. We discussed above that a high-risk mammal that is potentially capable of carrying SARS viruses should express a human-like ACE2 receptor, it follows that its host-specific coronavirus may be SARS-like in infectivity. Just as mammals at high risk of SARS-CoV-2 infection should have human-like ACE2 receptors at key S protein-binding sites, we expect SARS-like mammalian coronaviruses to express SARS-like S proteins at key ACE2-binding sites. Table 1 details 15 key SARS-CoV-2 S protein sites and 13 key SARS-CoV S protein sites that were deemed as essential in human ACE2-binding<sup>2,5</sup>. Determining similarities at key ACE2-binding sites among S proteins may help identify mammalian coronaviruses that have SARS-like infectivity and shed light on the zoonosis of SARS-CoV-2.

To expand on the above point, many recent studies point to the bat *Betacoronavirus* RaTG13 as a close relative of SARS-CoV-2<sup>5,28,29</sup>. SARS-CoV-2 S protein contains a unique Gly482, Val483, Glu484, and Gly485 four-residue

motif in the binding ridge that may facilitate contact with human ACE2 N-terminal helix. Indeed, bat RaTG13 also contains a similar four-residue motif<sup>5</sup>. Moreover, residues Leu455 and Asn501 on SARS-CoV-2 S protein are conserved by RaTG13 S protein<sup>3</sup>; these sites contribute favorably to human ACE2-binding because their mutations reduced binding potential. These observations may explain why RaTG13 could use human ACE2 as its receptor<sup>5</sup>. Nonetheless, many residues in RaTG13 S protein are not fine-tuned for binding with human ACE2<sup>2,5</sup>. Changing bat RaTG13 S protein at residues Lys486 and Tyr493 to SARS-CoV-2 S protein residues Phe486 and Gln493, respectively, enhanced human ACE2 recognition<sup>5</sup>. Besides bat RaTG13, the pangolin *Betacoronavirus* Pangolin-CoV also shares high sequence similarity with SARS-CoV-2<sup>28</sup> and contains the 482–485 four-residue motif in its S protein<sup>5</sup>. Furthermore, key ACE2-binding sites such as Leu455, Phe486, Gln493, and Asn501 are conserved between SARS-CoV-2 and Pangolin-CoV S proteins. Indeed, both bats and pangolins have been proposed to be potential intermediate hosts for SARS-CoV-2, and host-specific coronaviruses of these two mammals bear a resemblance to SARS-CoV-2 at key S protein sites<sup>30</sup>.

Our investigation considered 132 *ACE2* genes from mammals across 19 orders and 17 mammalian-specific coronaviruses infecting high-risk species. Key *ACE2* sites are strongly conserved among Primates. Among other mammals, key binding sites on human *ACE2* are conserved by selected species of the Chiroptera, Artiodactyla, Rodentia, Carnivora, Perissodactyla, Pholidota, Lagomorpha, Proboscidea, and Sirenia orders. Among these high-risk mammals, 12 species are known to be infected by host-specific coronaviruses. We found that key S protein sites in SARS-CoV are most conserved by Civet-CoV, whereas key S protein sites in SARS-CoV-2 are most conserved by Pangolin-CoV. Both SARS viruses also share several key S protein sites with bat RaTG13 but not with other mammalian-specific coronaviruses. Together, our results reinforce the current hypothesis that the progenitor of both SARS-CoV-2 and SARS-CoV are likely of bat origin. The palm civet and pangolin may have served as distinct intermediate hosts that facilitated the adaptation of SARS viruses to bind human ACE2 receptors prior to zoonosis because both mammals express human-like ACE2 receptors at key S protein-binding sites and their host-specific coronaviruses encode S proteins that share highest similarities with S proteins of SARS viruses at key ACE2-binding sites.

## Results

**Key binding sites on the human ACE2 receptor are most conserved by primates species and variably conserved in selected species belonging to eight other mammalian orders.** We investigated 132 species belongs to 19 mammalian orders with available *ACE2* gene records. Among Primates, key sites on human *ACE2* are perfectly conserved by species belonging to the *Hominidae* and *Cercopithecoidea* families and highly conserved by other species (Supplementary Fig. S1). Among the other 18 orders, key *ACE2* sites are highly conserved in selected species belonging to eight orders: Artiodactyla, Chiroptera, Carnivora, Rodentia, Lagomorpha, Perissodactyla, Proboscidea, and Sirenia (Supplementary Fig. S1). More specifically, in the Artiodactyla order, species belonging to the *Bovidae*, *Monodontidae*, *Phocoenidae*, and *Physeteridae* families share very high similarities with human at key *ACE2* sites. In the Chiroptera order, species belonging to the *Pteropodidae* and *Rhinolophidae* share high similarities with human at key *ACE2* sites. In the Carnivora order, species belonging to the *Canidae*, *Felidae*, and *Ursidae* families share high similarities with human at key *ACE2* sites. In the Rodentia order, all species (e.g., belonging to *Cricetidae*, *Sciuridae*) share high similarities with human at key *ACE2* sites except species in the *Muridae* family. In the other four orders, the specific species that share high similarities with human at key *ACE2* sites are *Ochotona princeps* and *Oryctolagus cuniculus* (order Lagomorpha), *Ceratotherium simum* (order Perissodactyla), *Loxodonta africana* (order Proboscidea), and *Trichechus manatus latirostris* (order Sirenia). Similar to the SARS-CoV example displayed in Fig. 1, while local *ACE2* site comparisons suggest that select species among nine orders could be at high risk to SARS-CoV-2 infection (Supplementary Fig. S1), global similarities among whole *ACE2* gene alignments (Supplementary Fig. S2) grouped 132 species by order and did not point out any species as high-risk.

Figure 2 showcases key *ACE2* site comparisons among a sample of mammals. These species were selected based on the following three criteria: (1) they all belong to the nine orders that contain high-risk mammals (Primates, Artiodactyla, Chiroptera, Carnivora, Rodentia, Lagomorpha, Perissodactyla, Proboscidea, and Sirenia), (2) all species that have been experimentally identified as susceptible to SARS-CoV-2 infection were selected (underlined), and (3) some species were selected because they are known to be infected by their own host-specific coronaviruses (in bold). In addition, the pangolin (*Manis javanica*) belonging to the Pholidota order shares medium similarity with human at key *ACE2* sites; the species was added to Fig. 2 because it has been proposed as a possible intermediate host of SARS-CoV-2<sup>30</sup>.

All mammals currently known to be susceptible to SARS-CoV-2 infection share similarity with human at key *ACE2* sites (Fig. 2). These include cat (*Felis catus*)<sup>16</sup>, dog (*Canis lupus familiaris*)<sup>17</sup>, and tiger (*Panthera tigris*)<sup>19</sup> where SARS-CoV-2 infections were detected, and ferret (*Mustela putorius furo*)<sup>20,21</sup>, macaque (*Macaca mulatta*)<sup>22</sup> and *Macaca fascicularis*<sup>23</sup>, and grivet (*Cercopithecus aethiops*)<sup>24</sup> that were experimentally infected by SARS-CoV-2. One additional mammal that is known to be infected by SARS-CoV-2 is the mink (*Neovison vison*)<sup>18</sup>, a relative of the ferret. However, the mink was not included in Fig. 2 because there were no *ACE2* gene records for this species in the NCBI gene database (last accessed April 25, 2020). Indeed, the *ACE2* genes in all these species share high similarity with human *ACE2* at key binding sites, except for ferret, which shares medium similarity. Wild-type mouse (*Mus musculus*) is not susceptible to SARS-CoV-2 infection<sup>26</sup>, and expectedly, the *ACE2* gene in mouse shares poor similarity with human *ACE2* at key binding sites (Fig. 2).

Among mammals that are known to be infected by SARS-CoV-2, nine key human *ACE2* sites (Ser19, Lys31, Glu35, Glu37, Tyr41, Gln42, Tyr83, Lys353 and Arg393) are conserved by their mammalian *ACE2* genes, but the other six key sites (Gln24, Asp30, His34, Asp38, Leu79, and Met82) are not well conserved (Fig. 2). One may reason that only these nine conserved *ACE2* sites could be important in S protein-binding. Nonetheless,

Order	Family	Species	Host coronavirus	ACE2 binding hotspots similarities with hACE2															Total similarity
				S19	Q24	D30	K31	H34	E35	E37	D38	Y41	Q42	L79	M82	Y83	K353	R393	
Primates	Hominidae	<u>Homo sapiens</u>	SARS-CoV-2, SARS-CoV, MERS	S19	Q24	D30	K31	H34	E35	E37	D38	Y41	Q42	L79	M82	Y83	K353	R393	15/15
		<u>Macaca mulatta</u>	n/a	S19	Q24	D30	K31	H34	E35	E37	D38	Y41	Q42	L79	M82	Y83	K353	R393	15/15
	Cercopithecoidea	<u>Macaca fascicularis</u>	n/a	S19	Q24	D30	K31	H34	E35	E37	D38	Y41	Q42	L79	M82	Y83	K353	R393	15/15
		<u>Cercopithecus aethiops</u>	n/a	S19	Q24	D30	K31	H34	E35	E37	D38	Y41	Q42	L79	M82	Y83	K353	R393	15/15
		<u>Mustela putorius furo</u>	n/a	S19	L24	E30	K31	Y34	E35	E37	E38	Y41	Q42	H79	T82	Y83	K353	R393	9/15
Carnivora	Canidae	<u>Canis lupus familiaris</u>	Canine CRCoV	S19	L23	E29	K30	Y33	E34	E36	E37	Y40	Q41	L78	T81	Y82	K352	R392	10/15
		<u>Felis catus</u>	Feline-CoV	S19	L24	E30	K31	H34	E35	E37	E38	Y41	Q42	L79	T82	Y83	K353	R393	11/15
	Felidae	<u>Panthera tigris</u>	n/a	S11	L16	E22	K23	H26	E27	E29	E30	Y33	Q34	L71	T74	Y75	K345	R385	11/15
		<u>Paguma larvata</u>	Civet-SARS-CoV 007/2004	S19	L24	E30	T31	Y34	E35	Q37	E38	Y41	Q42	L79	T82	Y83	K353	R393	8/15
		<u>Mesocricetus auratus</u>	n/a	S19	Q24	D30	K31	Q34	E35	E37	D38	Y41	Q42	L79	N82	Y83	K353	R393	13/15
Rodentia	Muridae	<u>Mus musculus</u>	MHV	S19	N24	N30	N31	Q34	E35	E37	D38	Y41	Q42	T79	S82	F83	H353	R393	7/15
		<u>Bos taurus</u>	Bovine-CoV	S19	Q24	E30	K31	H34	E35	E37	D38	Y41	Q42	M79	T82	Y83	K352	R392	12/15
	Artiodactyla	<u>Camelus ferus</u>	Camel HKU23	S19	L24	E30	E31	H34	E35	E37	D38	Y41	Q42	T79	T82	Y83	K353	R393	10/15
		<u>Sus scrofa</u>	Swine PHEV	S19	L24	E30	K31	L34	E35	E37	D38	Y41	Q42	I79	T82	Y83	K353	R393	10/15
		<u>Rhinolophus ferrumequinum</u>	Bat SARS-like-CoV AT098145	S19	L24	D30	D31	S34	E35	E37	N38	H41	Q42	L79	N82	F83	K353	R393	8/15
Chiroptera	Rhinolophidae	<u>Rhinolophus sinicus</u>	BtRS-BetaCoV/YN2013	S19	R24	D30	E31	S34	E35	E37	N38	Y41	Q42	L79	N82	Y83	K353	R393	10/15
		<u>Rhinolophus affinis</u>	Bat RaTG13	S19	R24	D30	N31	H34	E35	E37	D38	Y41	Q42	L79	N82	Y83	K353	R393	12/15
	<u>Oryctolagus cuniculus</u>	Rabbit HKU14	S19	L24	E30	K31	Q34	E35	E37	D38	Y41	Q42	L79	T82	Y83	K353	R393	11/15	
Lagomorpha	Equidae	<u>Equus caballus</u>	Equine CoV	S19	L24	E30	K31	S34	E35	E37	E38	H41	Q42	L79	T82	Y83	K353	R393	9/15
		<u>Ceratotherium simum</u>	n/a	S19	L24	E30	K31	P34	E35	E37	D38	Y41	Q42	L79	T82	Y83	K353	R393	11/15
Perissodactyla	Rhinocerotidae	<u>Trichechus manatus latirostris</u>	n/a	S19	L24	D30	T31	Q34	E35	E37	D38	Y41	Q42	L79	N82	F83	K348	R388	10/15
		<u>Loxodonta africana</u>	n/a	S19	L24	D30	T31	Q34	E35	E37	D38	Y41	Q42	L79	D82	F83	K348	R388	10/15
Proboscidea	Elephantidae	<u>Manis javanica</u>	Pangolin-CoV	S19	E24	E30	K31	S34	E35	E37	E38	Y41	Q42	L79	N82	Y83	K353	R393	9/15

**Figure 2.** Comparisons at key ACE2 sites between human and a sample of mammals belong to ten orders. Species that have been experimentally identified as susceptible to SARS-CoV-2 infection are underlined. Species that can be infected by their own host-specific coronaviruses are in black bold, and other species are in grey. Highlighted blue and red are matching and mis-matching amino acid residues, respectively, between mammalian ACE2 and human ACE2 at key SARS-CoV-2 S protein binding sites. Underlined amino acids are conserved among species known to be infected by SARS-CoV-2. Total similarity designates the total number of matching amino acid residues with respect to human ACE2. The total similarity score for mammalian ACE2 in blue highlights perfect site similarities (14–15 matching sites), in green highlights high similarity (12–13 matching sites), in light green highlights medium–high similarity (10–11 matching sites), in yellow highlights medium similarity (8–9 matching sites), in orange highlights medium–low similarity (6–7 matching sites), and in red highlights low similarity (5 or less matching sites). A comparison of key ACE2 sites involved in SARS-CoV-2 binding between human and 131 mammals is shown in Supplementary Figure S1.

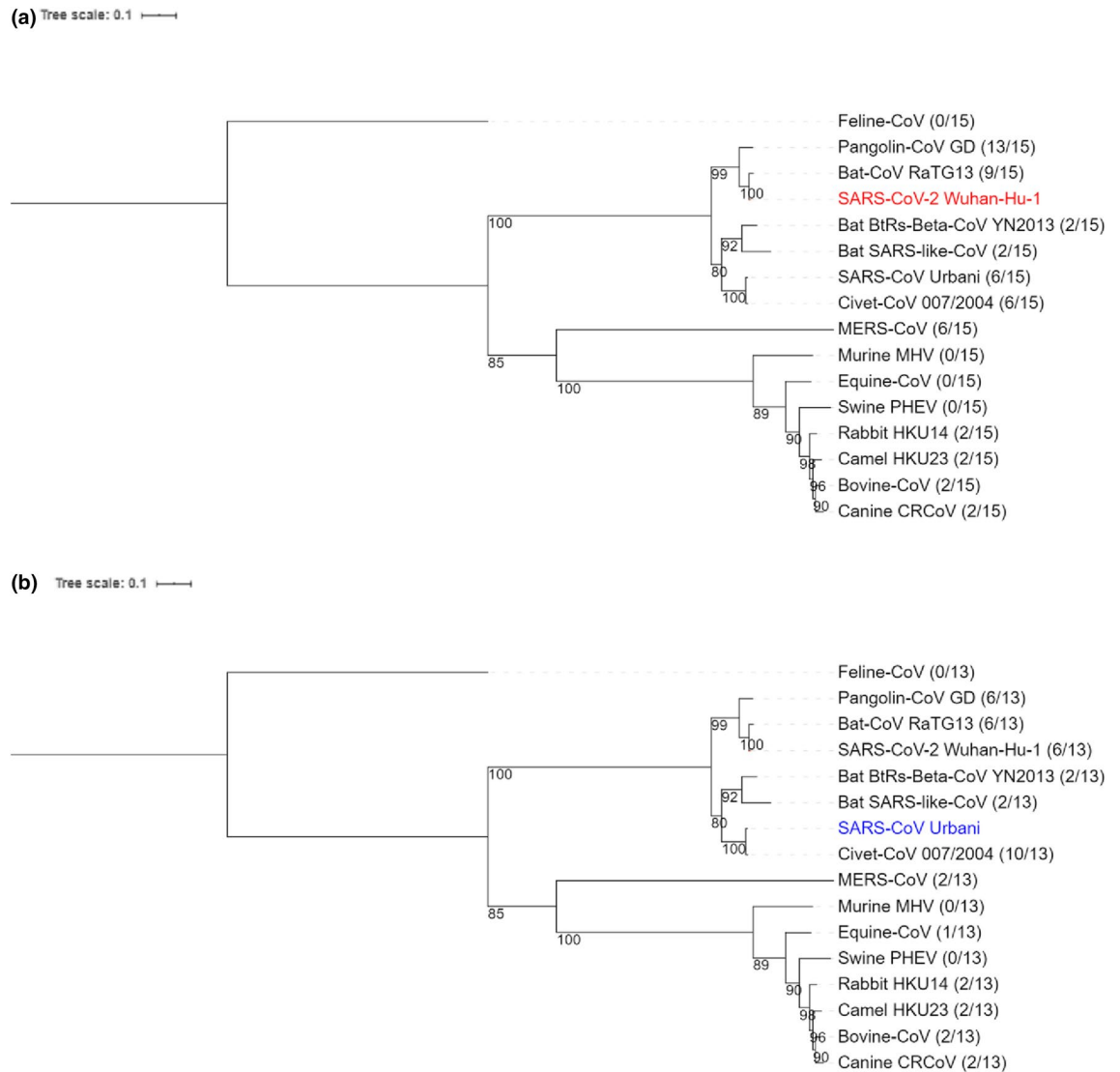
mis-match amino acids at the six non-conserved sites all share similar physiochemical properties with the human ACE2 residue (the replacements resulted in small Graham’s distance  $D^{31}$ , which implies similar composition, polarity, and molecular volume between the two amino acids). Aspartic acid at sites 30 and 38 are replaced by Glutamic acid (Grantham’s distance  $D = 45$ ), Glutamine at site 24 is replaced by Lysine ( $D = 53$ ), Histidine at site 34 is replaced by Tyrosine ( $D = 83$ ), and Methionine at site 82 is replaced by Threonine ( $D = 81$ ). Future ACE2 gene mutation studies may be required to better elucidate which key ACE2 sites among the 15 contacting residues are essential for SARS-CoV-2 infection.

**Key ACE2-binding sites on SARS-CoV-2 and SARS-CoV S proteins are distinctly conserved by mammalian coronaviruses.** Above results highlighted several mammals at high risk of SARS-CoV-2 infection. Figure 2 also highlights 12 species-specific coronaviruses that infect high-risk mammals (belonging to the Artiodactyla, Carnivora, Chiroptera, Perissodactyla, Pholidota, and Lagomorpha orders) and one MHV coronavirus that infects the low-risk mice, but mammals of the Proboscidea and Sirenia orders do not have records of species-specific coronaviruses. Some of these coronaviruses may have SARS-like infectivity because they infect mammals having human-like ACE2 receptors. We thus investigated whether key S protein sites in SARS-CoV-2 and SARS-CoV are conserved by mammalian-specific coronaviruses.

Figure 3 shows distinct differences in the conservation of key SARS-CoV-2 and SARS-CoV S protein sites by mammalian-specific coronaviruses. Between the two SARS viruses, key S protein sites of SARS-CoV-2 are weakly conserved by SARS-CoV (Fig. 3a) and vice versa (Fig. 3b). In addition, key S protein sites of both SARS viruses share a medium degree of similarity with S protein sites of bat coronavirus RaTG13 isolated from *Rhinolophus affinis*, but they share poor similarities with S protein sites of two other bat coronaviruses isolated from *Rhinolophus sinicus* and *Rhinolophus ferrumequinum*. Importantly, key S protein sites of SARS-CoV-2 share highest similarity with S protein sites of Pangolin-CoV strain Guangdong (GD) that was sequenced with high coverage, but share lower similarity with S protein sites of Pangolin-CoV strain Guangxi (GX) that was flagged as poorly sequenced by GISAID (Fig. 3a). Indeed, the unique 482–485 domain in SARS-CoV-2 (highlighted yellow, Fig. 3a) is perfectly conserved by Pangolin-CoV GD, partially conserved by bat RaTG13 and Pangolin-CoV GX, and not conserved by any other coronaviruses surveyed. In contrast, key S protein sites of SARS-CoV share highest similarity with S protein sites of Civet-CoV but share medium similarity with S protein sites of Pangolin-CoVs (Fig. 3b). As for other mammalian-specific coronaviruses, including the human MERS-CoV with a presumed camel origin, their S proteins share little similarities with SARS-CoV-2 and SARS-CoV at key sites. Together, these findings imply that the bat coronavirus RaTG13 is SARS-like in infectivity, but SARS-CoV-2 more closely resembles Pangolin-CoV and SARS-CoV more closely resembles Civet-CoV.

Global similarities among whole S protein-coding amino acid sequences (Fig. 4) also showed that SARS viruses are distinctly closely related to mammalian coronaviruses: SARS-CoV-2 closely relates to Pangolin-CoV and bat RaTG13 (Fig. 4a), and SARS-CoV closely relates to Civet-CoV (Fig. 4b). However, while SARS-CoV-2 is more similar to bat RaTG13 than to Pangolin-CoV in terms of global S protein similarity (Fig. 4a), key S protein sites are more similar between SARS-CoV-2 and Pangolin-CoV (Fig. 3a). The notion that comparing local but not





**Figure 4.** Phylogenetic reconstruction of 16 MAFFT G-INS-i aligned coronavirus S protein-coding genes. Infected mammalian species belong to the Artiodactyla, Canivora, Chiroptera, Lagomorpha, Perissodactyla, Primates, Pholidota and Rodentia orders. The phylogenetic tree is constructed 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 sequence coverage. The ratio appended to each coronavirus indicates the total number of amino acid matches when coronavirus S protein gene was compared to (a) SARS-CoV-2 S protein gene (in red) at its 15 key ACE2-binding sites and to (b) SARS-CoV S protein gene (in blue) at its 13 key ACE-2 binding sites.

key S protein sites among 17 coronaviruses infecting human, 12 high-risk mammals, and the low-risk mouse. We showed that key S protein sites of the two SARS viruses are modestly conserved by one bat coronavirus RaTG13. More importantly, key S protein sites in SARS-CoV and SARS-CoV-2 are distinctly most conserved by Civet-CoV and Pangolin-CoV GD, respectively (Fig. 3). Indeed, a recent study<sup>28</sup> had also found that SARS-CoV-2 is more similar to Pangolin-CoV GD than to bat RaTG13 or to SARS-CoV at the S1 binding domain of the S protein. However, 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 others have shown<sup>2,34</sup>. Hence, similarities in coronavirus infectivity can be better determined by local than global S protein sequence alignments. Nevertheless, both Figs. 3 and 4 suggest that the two SARS viruses are evolutionarily distinct in terms of infectivity.

Our results corroborate the current hypothesis on the origin and evolution of SARS viruses. The progenitors of SARS-CoV-2 and SARS-CoV are likely to have a bat coronavirus origin<sup>1,35</sup>, and the bat serves as a potential reservoir for these viruses. However, transmissions of SARS viruses from bats to humans are unlikely, and the viruses may have required adaptation in distinct intermediate hosts, namely the palm civet for SARS-CoV<sup>4,11</sup> and the pangolin for SARS-CoV-2<sup>30,34,36</sup>. Indeed, the ACE2 genes in both palm civet and pangolin share similarities with human ACE2 gene at key S protein-binding sites. Additionally, key S protein sites are most conserved between Civet-CoV and SARS-CoV, and between Pangolin-CoV and SARS-CoV-2. It is plausible that, prior to

zoonotic transmission, rapid evolution of progenitor SARS virus in distinct intermediate hosts allowed it to adapt high binding potential between viral S protein and human-like ACE2 receptor and led to differences between SARS-CoV-2 and SARS-CoV at the S protein.

## Methods

**Retrieving and processing 132 mammalian ACE2 genes and 17 coronavirus S protein-coding genes.** The nucleotide sequences of 266 ACE2 gene variants from 132 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 gene name and “Mammalia” as 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 ACE2 gene entry, only the coding DNA sequence region was extracted in FASTA format. The coding DNA sequences were translated from nucleotides into amino acids using DAMBE<sup>37</sup> and verified with annotated amino acid sequences from NCBI GenBank files. Then, for sequence files, gene IDs were renamed as follow: ACE2\_NCBI gene accession ID\_Species name (Supplementary File S1). Similarly, the amino acid sequences of S proteins encoded by 17 host-specific coronaviruses infecting 12 mammalian species were retrieved from NCBI and extracted in FASTA format.

Multiple ACE2 isoforms were retrieved for some species, but only one ACE2 gene per species was selected for analysis. Many isoforms differ only in the 5' and 3' UTRs. For human, 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 protein. Furthermore, some isoforms were experimentally validated, while others were predictions by automated computational approaches. We selected experimentally verified variants when available, but when only predicted variants are available, we picked one among those whose (1) amino acid identities are conserved by most other variants, and (2) sequence is not truncated (selected ACE2 variants are listed in Supplementary File S1). Similarly, one representative coronavirus was picked out of several available strains (e.g., strain Urbani was picked for SARS-CoV).

Next, amino acid sequences of ACE2 genes and of S protein-coding genes were aligned with MAFFT<sup>38</sup> with the slow but accurate G-INS-i option. We then extracted the location and identity of amino acids that aligned to key human ACE2 sites and to key S protein-coding gene sites in SARS-CoV-2 and SARS-CoV listed in Table 1. Mammalian ACE2 match-mismatch heat-maps were then generated, and a total similarity score (the total number of matching amino acid identities between human and mammals at key ACE2 sites) was calculated for each mammal. Similarly, coronavirus S protein match-mismatch heat-maps were generated (one against SARS-CoV-2 S protein and another against SARS-CoV S protein), and the total similarity score was calculated for each mammalian-specific coronavirus.

**Phylogenetic reconstruction based on 132 ACE2 genes and 16 S protein-coding genes.** Three phylogenetic trees were constructed using MAFFT G-INS-i aligned amino acid sequences with the maximum-likelihood-based PHYML approach<sup>39</sup>: one tree for aligned ACE2 genes from 132 mammalian species (bootstrap = 500, model = JTT + G + I + F), another tree for ACE2 genes from 13 sample mammalian species (bootstrap = 100, model = JTT + G + I + F), and a third tree for 16 mammalian-specific coronavirus S proteins (bootstrap = 100, model = WAG + G + I + F). All were constructed using the PHYML model 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, branch lengths and rate parameters. Tree figure illustrations were made using the Interactive Tree Of Life (iTOL) v4<sup>40</sup>.

## Data availability

Supplementary file S1 contains data for mammalian ACE2 and coronavirus S protein gene accessions, the selected species used for phylogenetic reconstruction, and key site comparisons between 132 mammalian species at aligned ACE2 genes and between 17 coronaviruses at aligned S protein-coding genes. Supplementary file S2 contains Supplementary figures S1, S2, and S3.

Received: 30 April 2020; Accepted: 10 December 2020

Published online: 18 January 2021

## References

- Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
- Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
- 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).
- Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *Embo. J.* **24**, 1634–1643 (2005).
- Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **30**, 020–2179 (2020).
- Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
- Tian, X. *et al.* Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg. Microbes. Infect.* **9**, 382–385. <https://doi.org/10.1080/22221751.2020.1729069> (2020).
- Walls, A. C. *et al.* Structure, function, and antigenicity of the SARS-CoV-2 Spike glycoprotein. *Cell* **181**, 281–292 (2020).
- 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).
- 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).



11. Guan, Y. *et al.* Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276–278 (2003).
12. Song, H. D. *et al.* Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci. USA* **102**, 2430–2435 (2005).
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. Liu, P. *et al.* Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog.* **16** (2020).
16. Halfmann, P. J. *et al.* Transmission of SARS-CoV-2 in domestic cats. *N. Engl. J. Med.* **383**, 592–594 (2020).
17. Sit, T. H. C. *et al.* Infection of dogs with SARS-CoV-2. *Nature* <https://doi.org/10.1038/s41586-020-2334-5> (2020).
18. Oreshkova, N. *et al.* SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. *Euro. Surveill.* **25**, 1560–7917 (2020).
19. Wang, L. *et al.* Complete genome sequence of SARS-CoV-2 in a tiger from a US Zoological Collection. *Microbiol. Resour. Announc.* **9**, e00468. <https://doi.org/10.1128/mra.00468-20> (2020).
20. Kim, Y.-I. *et al.* Infection and rapid transmission of SARS-CoV-2 in ferrets. *Cell Host Microbe* **27**, 704–709. <https://doi.org/10.1016/j.chom.2020.03.023> (2020).
21. Shi, J. *et al.* Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* **368**, 1016–1020 (2020).
22. Chandrashekar, A. *et al.* SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science* **369**, 812–817 (2020).
23. Rockx, B. *et al.* Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. *Science* **368**, 1012–1015 (2020).
24. Woolsey, C. *et al.* Establishment of an African green monkey model for COVID-19. <https://doi.org/10.1101/2020.05.17.100289v1> (2020).
25. Chan, J.F.-W. *et al.* Simulation of the clinical and pathological manifestations of Coronavirus Disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis and transmissibility (Infect. Dis. Clin. 2020). <https://doi.org/10.1093/cid/ciaa325>.
26. Bao, L. *et al.* The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature* **583**, 830–833 (2020).
27. Gralinski, L. E. & Menachery, V. D. Return of the coronavirus: 2019-nCoV. *Viruses* **12**, 135 (2020).
28. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
29. Tang, X. *et al.* On the origin and continuing evolution of SARS-CoV-2. *Nat. Sci. Rev.* **7**, 1012–1023. <https://doi.org/10.1093/nsr/nwaa036> (2020).
30. Xiao, K. *et al.* Isolation and characterization of 2019-nCoV-like coronavirus from Malayan pangolins. <https://doi.org/10.1101/2020.02.17.951335v1> (2020).
31. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
32. 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. *J. Virol.* **94**, e00127–00120. <https://doi.org/10.1128/jvi.00127-20> (2020).
33. Schlottau, K. *et al.* SARS-CoV-2 in fruit bats, ferrets, pigs, and chickens: An experimental transmission study. *Lancet Microbe*. [https://doi.org/10.1016/S2666-5247\(20\)30089-6](https://doi.org/10.1016/S2666-5247(20)30089-6) (2020).
34. Zhang, T., Wu, Q. & Zhang, Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr. Biol.* **30**, 1578 (2020).
35. Perlman, S. Another decade, another coronavirus. *N. Engl. J. Med.* **382**, 760–762 (2020).
36. Lam, T.T.-Y. *et al.* Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. *Nature* **583**, 282–285 (2020).
37. Xia, X. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* **35**, 1550–1552 (2018).
38. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).
39. 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).
40. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259. <https://doi.org/10.1093/nar/gkz239> (2019).

## 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].

## Author contributions

Y.W. and X.X. designed the study and wrote the manuscript. Y.W., P.A., and H.F. collected and analyzed the data. Y.W. and P.A. prepared all figures. X.X. supervised the study. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80573-x>.

**Correspondence** and requests for materials should be addressed to X.X.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2021