Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates

Joana Damas[a,1], Graham M. Hughes[b,1], Kathleen C. Keough[c,d,1], Corrie A. Painter[e,f,1], Nicole S. Persky[e,g,1], Marco Corbo[a], Michael Hiller[h,i,j], Klaus-Peter Koepfli[k], Andreas R. Pfenning[l], Huabin Zhao[m], Diane P. Genereux[e], Ross Swofford[e], Katherine S. Pollard[c,d,n], Oliver A. Ryder[o,p], Martin T. Nweeia[q,r,s], Kerstin Lindblad-Toh[e,t], Emma C. Teeling[b], Elinor K. Karlsson[e,u,v], and Harris A. Lewin[a,w,x]

[a]The Genome Center, University of California Davis, Davis, CA 95616, USA; [b]School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland; [c]University of California San Francisco, San Francisco, CA 94117, USA; [d]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA 94158, USA; [e]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [f]Cancer Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [g]Genetic Perturbation Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [h]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany; [i]Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany; [j]Center for Systems Biology Dresden, 01307 Dresden, Germany; [k]Smithsonian Conservation Biology Institute, Center for Species Survival, National Zoological Park, Front Royal, VA 22630, Washington, DC 20008 USA; [l]Department of Computational Biology, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA; [m]Department of Ecology and Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan 430072, China; [n]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA; [o]San Diego Zoo Institute for Conservation Research, Escondido, CA 92027, USA; [p]Department of Evolution, Behavior, and Ecology, Division of Biology, University of California San Diego, La Jolla, CA 92093, USA; [q]Harvard School of Dental Medicine, Boston, MA 02115, USA; [r]Case Western Reserve University School of Dental Medicine, Cleveland, OH 44106, USA; [s]Marine Mammal Program, Department of Vertebrate Zoology, Smithsonian Institution, Washington, DC 20002, USA; [t]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, 751 23, Sweden; [u]Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01655, USA; [v]Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01655, USA; [w]Department of Evolution and Ecology, University of California Davis, Davis, CA 95616, USA; [x]John Muir Institute for the Environment, University of California Davis, Davis, CA 95616, USA.

Corresponding author: Harris A. Lewin, Department of Evolution and Ecology, College of Biological Sciences, University of California Davis, CA 95616, USA, lewin@ucdavis.edu, Tel: +1 530-754-5098

[1]These authors contributed equally to this work.

Abstract

The novel coronavirus SARS-CoV-2 is the cause of Coronavirus Disease-2019 (COVID-19). The main receptor of SARS-CoV-2, angiotensin I converting enzyme 2 (ACE2), is now undergoing extensive scrutiny to understand the routes of transmission and sensitivity in different species. Here, we utilized a unique dataset of 410 vertebrates, including 252 mammals, to study cross-species conservation of ACE2 and its likelihood to function as a SARS-CoV-2 receptor. We designed a five-category ranking score based on the conservation properties of 25 amino acids important for the binding between receptor and virus, classifying all species from *very high* to *very low*. Only mammals fell into the *medium* to *very high* categories, and only catarrhine primates in the *very high* category, suggesting that they are at high risk for SARS-CoV-2 infection. We employed a protein structural analysis to qualitatively assess whether amino acid changes at variable residues would be likely to disrupt ACE2/SARS-CoV-2 binding, and found the number of predicted unfavorable changes significantly correlated with the binding score. Extending this analysis to human population data, we found only rare (<0.1%) variants in 10/25 binding sites. In addition, we observed evidence of positive selection in ACE2 in multiple species, including bats. Utilized appropriately, our results may lead to the identification of intermediate host species for SARS-CoV-2, justify the selection of animal models of COVID-19, and assist the conservation of animals both in native habitats and in human care.

Keywords: SARS-CoV-2, COVID-19, ACE2, comparative genomics, host range, species conservation, evolution

Introduction

The 2019-novel coronavirus (2019-nCoV, also, SARS-CoV-2 and COVID-19 virus) is the cause of Coronavirus Disease-2019 (COVID-19), a major pandemic that threatens millions of lives and the global economy (1). Comparative analysis of SARS-CoV-2 and related coronavirus sequences has shown that SARS-CoV and SARS-CoV-2 likely originated in bats, followed by transmission to an intermediate host, and that both viruses may have an extended host range that includes primates and other mammals (1–3). However, the immediate source population/species for SARS-CoV and SARS-CoV-2 viruses has not yet been identified. Several mammalian species host coronaviruses, and these infections are frequently associated with severe clinical diseases, such as respiratory and enteric disease in pigs and cattle (4, 5). Molecular phylogenetics revealed that at least one human coronavirus (HCov-OC43), may have originated in cattle or swine (6), and that this virus was associated with a human pandemic that emerged in the late 19th century (7). Recent data indicate that coronaviruses can move from bats to other wildlife species and humans (8) and from humans to tigers (9) and pigs (10). Therefore, understanding the host range of SARS-CoV-2 and related coronaviruses is essential for improving our ability to predict and control future pandemics. It is also crucial for protecting populations of wildlife species in native habitats and under human care, particularly non-human primates, who may also be susceptible to COVID-19 (11).

The angiotensin I converting enzyme 2 (ACE2) serves as a functional receptor for the spike protein (S) of SARS-CoV and SARS-CoV-2 (12, 13). Under normal physiological conditions, ACE2 is a dipeptidyl carboxypeptidase that catalyzes the conversion of angiotensin I into angiotensin 1-9, a peptide of unknown function, and angiotensin II, a vasoconstrictor that is important in the regulation of blood pressure (14). ACE2 also converts angiotensin II into angiotensin 1-7, a vasodilator that affects the cardiovascular system (14) and may regulate other components of the renin-angiotensin system (15). The host range of SARS-CoV-2 may be extremely broad due to the conservation of ACE2 in mammals (2, 13) and its expression on ciliated bronchial epithelial cells and type II pneumocytes (10). While coronaviruses related to SARS-CoV-2 use ACE2 as a primary receptor, coronaviruses may use other proteases as receptors, such as CD26 (DPP4) for MERS-CoV (16), thus limiting or extending their host range.

In humans, ACE2 may be a cell membrane protein or it may be secreted (14). The secreted form is created primarily by enzymatic cleavage of surface-bound ACE2 by ADAM17 and other proteases (14). Sequence variation in *ACE2* affects the protein's functions. *ACE2* is polymorphic in humans, with many

3

synonymous and nonsynonymous mutations identified, although most are rare at the population level (17) and few are believed to affect cellular susceptibility to human coronavirus infections (18). Site-directed mutagenesis and co-precipitation of SARS-CoV constructs have revealed critical residues on the ACE2 tertiary structure that are essential for binding to the virus receptor binding domain (RBD) (19). These findings have been strongly supported by co-crystallization and the structural determination of the SARS-CoV and SARS-CoV-2 S proteins with human ACE2 (13, 20, 21), as well as binding-affinity with heterologous ACE2 (19). The RBD of human coronaviruses may mutate to change the binding affinity of S for ACE2, and thus lead to adaptation in humans or other hosts. The best studied example is the palm civet, believed to have been the intermediate host between bats and humans for SARS-CoV (2). To date, an intermediate host for SARS-CoV-2 has not been identified definitively, although Malayan pangolins (*Manis javanica*) have been proposed as a possible reservoir (22).

Comparative analysis of ACE2 nucleotide and protein sequences can predict their ability to bind SARS-CoV-2 S and therefore will yield important insights into the biology and potential zoonotic transmission of SARS-CoV-2 infection. Recent work has examined ACE2 from different vertebrate species and predicted its ability to bind SARS-CoV-2 S, but phylogenetic sampling was extremely limited (11, 23). Here, we made use of sequenced genomes of 410 vertebrates and protein structural analysis, to identify ACE2 homologs in all vertebrate classes (fishes, amphibians, birds, reptiles, and mammals) that have the potential to serve as a receptor for SARS-CoV-2, and to understand the evolution of ACE2 SARS-CoV-2 S binding sites. Our results reinforce earlier findings on the natural host range of SARS-CoV-2, and predict a broader group of species that may serve as a reservoir or intermediate host for this virus. Importantly, many threatened and endangered species were found to be at potential risk for SARS-CoV-2 infection, suggesting that as the pandemic spreads, humans could inadvertently introduce a potentially devastating new threat to these already vulnerable populations, especially for great apes and other primates.

Results

Comparison of vertebrate ACE2 sequences and their predicted ability to bind SARS-CoV-2. We identified 410 unique vertebrate species with *ACE2* orthologs (Dataset S1) that included representatives of all vertebrate taxonomic classes. Among these were 252 mammals, 72 birds, 65 fishes, 17 reptiles and 4 amphibians. Twenty-five amino acids corresponding to known SARS-CoV-2 S-binding residues (11, 13, 21) were examined for their similarity to the residues in human ACE2 (Fig. 1, Dataset S1). On the basis of

known interactions between specific residues on ACE2 and the RBD of SARS-CoV-2 S, a set of rules was developed for predicting the likelihood of S binding to ACE2 from each species (see Materials and Methods). Five score categories were predicted: *very high*, *high*, *medium*, *low* and *very low*. Results for all species and all SARS-CoV-2 S binding scores are shown in Dataset S1, and results for mammalian species are also shown in Fig. 1. The *very high* classification had at least 23/25 ACE2 residues identical to their human homolog and other constraints on substitutions at SARS-CoV-2 S binding hot spots (see Materials and Methods). The 18 species predicted as *very high* were all Old World primates and apes completely identical to human across the 25 ACE2 binding residues. The ACE2 proteins of 28 species were classified as having a *high* likelihood of binding the S RBD. Among them are twelve cetaceans, seven rodents, three cervids (deer), three lemuriform primates, two representatives of the order Pilosa (Giant anteater and Southern tamandua), and one Old World primate (Angola colobus, Fig. 1). Fifty-seven species scored as *medium* for the ability of their ACE2 to bind SARS-CoV-2 S. Like the *high* score, this category has at least 20/25 residues identical to human ACE2 but more relaxed constraints for critical binding residues. All species with *medium* score are mammals distributed across six orders.

Among Carnivora, 9/43 scored *medium*, 9/43 scored *low*, and 25/43 scored *very low* (Fig. 1). The carnivores scoring *medium* were only felids, including the domestic cat and Siberian tiger. Among the 13 Primates scoring *medium* there were 10 New World primates and three lemurs. Of 45 Rodentia species, 11 scored *medium*. Twenty-one Artiodactyls scored *medium*, including several important wild and domesticated ruminants, such as domesticated cattle, bison, sheep, goat, water buffalo, Masai giraffe, and Tibetan antelope. Species scoring *medium* also included 2/3 Lagomorphs and one Cetacean (sperm whale).

All chiropterans (bats) scored *low* (N=8) or *very low* (N=29) (Fig. 1), including the Chinese rufous horseshoe bat (*Rhinolophus sinicus*), from which a coronavirus very similar to SARS-CoV-2 was identified (1). Only 7.7% (3/39) primate species' ACE2 scored *low* or *very low,* and 61% of rodent species scored *low* (10/46) or *very low* (18/46). All monotremes (N=1) and marsupials (N=4) scored *very low*. All birds, fish, amphibians, and reptiles scored *very low*, with less than 18/25 ACE2 residues identical to the human and many non-conservative residues at the remaining non-identical sites (Dataset S1). Notable species scoring *very low* include the Chinese pangolin (*Manis pentadactyla*), Sunda pangolin (*Manis javanica*), and white-bellied pangolin (*Phataginus tricuspis*) (Fig. 1, Dataset S1).

Structural analysis of the ACE2/SARS-CoV-2 S binding interface. We complemented the sequence-identity based scoring scheme with a qualitative approach that combined structural homology

modeling and best fit rotamer positioning. We examined the 25 ACE2 binding residues in a subset of 28 representative species (Fig. S1) and 17 sites were variable and not glycosylation sites. First, we assessed the similarity of every contact at the binding interface between two recently solved crystal structures for the human ACE2/SARS-CoV-2 S RBD complex in humans, 6M0J and 6WV1 (13, 21). Both structures were in agreement except for the position of S19, which was excluded from subsequent analysis (24). We then generated homology models, and aligned them to the human ACE2/SARS-CoV-2 S RBD 6M0J structure. This showed a high degree of similarity along the $C\alpha$ backbone (25) for each of the 28 species. We selected the most favorable rotamer at each residue using CHIMERA (Fig. S2).

We examined a total of 55 substitutions and assigned each to one of three types: *neutral* (N; likely to maintain similar contacts; 18 substitutions); *weaken* (W; likely to weaken the interaction; 14 substitutions); or *unfavorable* (U; likely to introduce unfavorable interactions; 23 substitutions) (Fig. S1). Our assignments show good agreement with those made in a second study (26) based on experimental data, with 83.4% of the 55 substitutions evaluated concordant between the two approaches (Fig. S1). The structural homology binding assessments support the sequence identity analysis, with the fraction of residues ranked as U, correlating very strongly with the substitution scoring scheme (Spearman correlation rho=0.76; p< 2.2e-16; Fig. 2).

Structural analysis of variation in human *ACE2*. We applied the same approach used to compare species, sequence identity and protein structural analysis, to examine the variation in ACE binding residues within humans, some of which have been proposed to alter binding affinity (18, 27–30). We integrated data from six different sources: dbSNP (31), 1KGP (32), Topmed (33), UK10K (34) and CHINAMAP (28), and identified a total of 11 variants in ten of the 25 ACE2 binding residues (Dataset S2). All variants found are rare, with allele frequency less than 0.01 in any populations, and less than 0.0007 over all populations. Three of the 11 variants were synonymous changes, seven were conservative missense variants, and one, S19P, was a semi-conservative substitution. S19P has the highest allele frequency of the 11 variants, with a global frequency of 0.0003 (17). We evaluated, by structural homology, six missense variants. Four were *neutral* and two weakening (E35K, frequency=0.000016; E35D, frequency=0.000279799). S19P was not included in our structural homology assessment, but a recent study predicted it would increase binding affinity (26). Thus, with an estimated summed frequency of 0.001, genetic variation in the ACE2 S-binding interface is overall rare, and it is unclear whether the variation that does exist increases or decreases susceptibility to infection.

Evolution of ACE2 across mammals. We next investigated the evolution of ACE2 variation in vertebrates, including how patterns of positive selection compare between bats, a mammalian lineage known to harbor a diversity of coronaviruses (35), and other mammalian clades. We first inferred the phylogeny of *ACE2* using our 410-vertebrate alignment and IQTREE, using the best-fit model of sequence evolution (JTT+F+R7) and rooting the topology on fishes (Dataset S3; Fig. S3). We then assayed sequence conservation with PhyloP (36). The majority of ACE2 codons are significantly conserved across vertebrates and across mammals, likely reflecting its critical function in the renin-angiotensin system (37) (Dataset S4.1), with ten residues in the ACE2 binding domain exceptionally conserved in Chiroptera and/or Rodentia (Dataset S4.2).

We next used phyloP and CODEML to test for acceleration and positive selection (36). PhyloP compares the rate of evolution at each codon to the expected rate in a model estimated from third nucleotide positions of the codon, and is agnostic to synonymous versus nonsynonymous substitutions (dN/dS). CODEML uses ω=dN/dS>1 and Bayes Empirical Bayes (BEB) scores to identify codons under positive selection, and was run on a subset of 64 representative mammals (see Materials and Methods).

ACE2 shows significant evidence of positive selection across mammals (ω=1.83, LRT=194.13, p<0.001; Dataset S4.3, 4.4). Almost 10% of codons (N=73; 9 near the RBD) are accelerated within mammals (Dataset S4.1, 4.5), and 18 of these have BEB scores greater than 0.95, indicating positively selected residues (Dataset S4.5, 4.6, Fig. S4). Nineteen accelerated residues, including two positively-selected codons (Q24, H34), are critical for the binding of the ACE2 RBD and SARS-CoV-2 S (Dataset S4.5; Fig. 3; Fig. S5). Q24 has not been observed to be polymorphic within the human population, and H34 harbors a synonymous polymorphism (AF=0.00063) but no non-synonymous polymorphisms (Dataset S2).

This pattern of acceleration and positive selection in *ACE2* also holds for individual mammalian lineages. Using CODEML, positive selection was detected within the orders Chiroptera (LRT=346.40, ω=3.44 p<0.001), Cetartiodactyla (LRT=92.86, ω=3.83, p<0.001), Carnivora (LRT=65.66, ω=2.27. p<0.001), Primates (LRT=72.33, ω=3.16, p<0.001) and Rodentia (LRT=91.26, ω=1.77, p<0.001). Overall, bats had more positively selected sites with significant BEB scores (29 sites in Chiroptera compared to 10, 8, 7 and 15 sites in Cetartiodactyla, Carnivora, Primates and Rodentia, respectively). Positive selection at key sites for the binding of ACE2 and SARS-CoV-2 was only found in the bat-specific alignment. PhyloP was used to assess shifts in evolutionary rate within mammalian lineages, for each assessing signal relative to a neutral model trained on species from the specified lineage (Dataset S4.6-11, Fig. S6). We discovered six important binding residues, five of which showed evidence for positive selection, that are accelerated in

one or more of Chiroptera, Rodentia, or Carnivora, with G354 accelerated in all of these lineages (Dataset S4.12).

Given pervasive signatures of adaptive evolution in *ACE2* across mammals, we next sought to test if any mammalian lineages are evolving particularly rapidly compared to the others. CODEML branch-site tests identified positive selection in both the ancestral Chiroptera branch (1 amino acid, ω=26.7, LRT= 4.22, p=0.039) and ancestral Cetartiodactyla branch (2 amino acids, ω=10.38, LRT= 7.89, p=0.004, Dataset S4.3) using 64 mammals. These residues did not correspond to known viral binding sites. We found no evidence for lineage-specific positive selection in the ancestral primate, rodent or carnivore lineages. PhyloP identified lineage-specific acceleration in Chiroptera, Carnivora, Rodentia, Artiodactyla and Cetaceans relative to mammals (Dataset S4.13-17, Fig. S7). Bats have a particularly high level of accelerated evolution (18 codons; p<0.05). Of these accelerated residues, T27 and M82 are known to be important for binding SARS-CoV-2, with some bat subgroups having amino acids predicted to lead to less favorable binding of SARS-CoV-2 (Fig. S1, Fig. S8). Surprisingly, a residue that is conserved overall in our 410 species alignment and in the mammalian subset, Q728, is perfectly conserved in all 37 species of bats except for fruit bats (Pteropodidae), which have a substitution from Q to E. These results support the theory that *ACE2* is under lineage-specific selective pressures in bats relative to other mammals.

Positive selection in SARS-CoV-2 S protein. Positive selection was found using CODEML at sites L455, E484, F490 and S494 in the SARS-CoV-2 S sequence (ω=1.15, LRT=116.7, p<0.001); however this signal was not particularly high, possibly due to the small sample size (N=8). All of these sites lie within or near the ACE2 SARS-CoV-2 S RBD binding sites (Fig. 3) (38).

Discussion

Phylogenetic analysis of coronaviruses has demonstrated that SARS-CoV-2 most likely originated in a bat species (1). However, whether SARS-CoV-2 or the progenitor of this virus was transmitted directly to humans or through an intermediary host is not yet resolved. To determine if amino acid substitution analysis and structural information could be used to identify candidate intermediate host species, we undertook a deep comparative genomic, evolutionary and structural analysis of ACE2, the SARS-CoV-2 receptor in humans. To accomplish this we drew on the rapidly growing dataset of annotated vertebrate genomes as well as predicted protein sequences from recently acquired whole genome sequences produced by the Genomes 10K-affiliated Bat1K Consortium, Zoonomia, and Vertebrate

Genomes Project, and other sources (39, 40). We conducted a phylogenetic analysis of ACE2 orthologs from 410 vertebrate species and made predictions of their likelihood to bind the SARS-CoV-2 S using a score that was based on amino acid substitutions at 25 consensus human ACE2 binding residues (13, 21). We supported these predictions with comprehensive homology modeling of the ACE2 binding site. We also tested the hypothesis that the ACE2 receptor is under selective constraints in different mammalian lineages, and correlated these results with data on the known species distribution of coronaviruses.

Several recent studies examined the role of ACE2 in SARS-CoV-2 binding and cellular infection, and its relationship to experimental and natural infections in different species (30, 41–46). Our study design differs substantially from those studies in several aspects: 1) we analyzed a larger number of primates, carnivores, rodents, cetartiodactyls and other mammalian orders, and an extensive phylogenetic sampling of fishes, birds, amphibians and reptiles; 2) we analyzed the full complement of S-binding residues across the ACE2 binding site, which was based on a consensus set from two independent studies (13, 21); 3) we used different methodologies to assess ACE2 binding capacity for SARS-CoV-2 S; and, 4) our study tested for selection and accelerated evolution across the entire ACE2 protein. While our results are strongly consistent with the results and conclusions of Melin and colleagues (44) on the predicted susceptibility of primates to SARS-CoV-2, particularly Old World primates, our work made predictions for a larger number of primates (N=39 vs N=27), bats (N=37 vs N=7), other mammals (N=176 vs N=5) and other vertebrates (N=158 vs N=0). When ACE2 from species in our study were compared with results of other studies there were many consistencies, such as for rodents, but some predictions that differ, such as the relatively high risk described for SARS-CoV-2 binding in pangolin and horse (45), civet (46), *Rhinolophus sinicus* bats (46) and turtles (45). In one recent study, binding affinity of soluble ACE2 for the SARS-CoV-2 S RBD was analyzed by saturation mutagenesis (26). Results obtained at each ACE2 binding residue were generally consistent with ours, particularly in the binding hotspot region of ACE2 residues 353-357. Importantly, as compared with other studies, our results greatly expanded the potential number of intermediate hosts and identified many more threatened species that could be infected by SARS-CoV-2 via their ACE2 receptors.

Evolution of *ACE2*. Variation of *ACE2* in the human population is rare (17). We examined a large set of *ACE2* variants for their potential differences in binding to SARS-CoV-2 S and their relationship to selected and accelerated sites. We found rare variants that would result in missense mutations in 7 out of the 25 binding residues (Dataset S2). Some of those (e.g. E35K with an AF of 0.00001636) could reduce the virus binding affinity, thus potentially lowering the susceptibility to the virus in a very small fraction of

the population. The analysis suggests that some variants (e.g. D38E) might not affect the binding while others (e.g. S19P) have uncertain effects. Further studies are needed to confirm and correctly address recent discoveries (18, 27, 28) and the data presented here, investigating the possible effect of these rare variants in specific populations.

When exploring patterns of codon evolution in ACE2, we found that a number of sites are evolving at different rates in the different lineages represented in our 410-species vertebrate alignment. Multiple ACE2 RBD residues important for the binding of SARS-CoV-2 are evolving rapidly across mammals, with two (Q24 and H34) under positive selection (Fig. 3, Fig. S5). Relative to other lineages analyzed, Chiroptera has a greater proportion of accelerated versus conserved residues, particularly at the SARS-CoV-2 S RBD, suggesting the possibility of selective forces on these codons in Chiroptera driven by their interactions with SARS-CoV-2-like viruses (Dataset S4.12, Fig. S8). Indeed, distinct signatures of positive selection found in bats and in the SARS-CoV S protein support this hypothesis that bats are evolving to tolerate SARS-CoV-2-like viruses.

Relationship of the ACE2 binding score to known infectivity of SARS-CoV-2. Data on susceptibility of wild animals to SARS-CoV-2 is still very limited. It has been reported that a captive Malayan tiger was infected by SARS-CoV-2 (9) and that domestic cats, ferrets (47), rhesus macaques (48) and Syrian golden hamsters (49) are susceptible to experimental infection by SARS-CoV-2. These results agree with our predictions of ACE2 binding ability to SARS-CoV-2 S (Fig. 1, Dataset S1); 4/5 five species with demonstrated susceptibility to SARS-CoV-2 score *very high* (Rhesus macaque) or *medium* (domestic cat, tiger and Golden hamster). The only inconsistency was observed for ferrets, which had a *low* ACE2 binding score. This inconsistency could be related to the high infectivity dose used for experimental infection that likely does not correspond to virus exposure in nature. Dogs have low susceptibility to SARS-CoV-2 under experimental conditions (47), and score *low* for binding of their ACE2 to SARS-CoV-2 S. However, kidney cell lines derived from dog showed ACE2-dependent SARS-CoV-2 S entry, suggesting that *in vitro* experiments may be overestimating true infectivity potential (39, 50). Pigs (*low*), ducks (*very low*) and chickens (*very low*) were similarly exposed to SARS-CoV-2 and showed no susceptibility (47), providing further support of our methodology. A recent publication reporting that SARS-CoV-2 could use pig, masked palm civet and Chinese rufous horseshoe bat ACE2 expressed in HeLa cells were inconsistent with our predictions, while data for mouse was in agreement (1). Indeed, while mouse ACE2 scored *very low* in our analysis, pig and Chinese rufous horseshoe bat score *low*, while the masked palm civet scored *very low*. As for the ferret, high-level exposure to the virus *in vitro* could potentially result in infection via low affinity interactions with ACE2. Another possibility is that other cellular machinery

present in the human HeLa cells is facilitating the infection, and that infectivity does not relate directly to ACE2 differences in these species. Confirmation of *in vitro* and *in vivo* susceptibility of these species under physiological conditions and with proper controls is clearly necessary. In addition, the expression of ACE2 varies across animal age, cell types, tissues and species (51, 52), which may lead to discrepancies between SARS-CoV-2 susceptibility gleaned from experimental infections or laboratory experiments and predictions made from the ACE2-based binding score.

Mammals with high predicted risk of SARS-CoV-2 infection. Of the 19 catarrhine primates analyzed, 18/19 scored *very high* for binding of their ACE2 to SARS-CoV-2 S and one scored *high* (the Angola colobus); the 18 species scoring *very high* had 25/25 identical binding residues to human ACE2, including rhesus macaques (*Macaca mulatta*), which are known to be infected by SARS-CoV-2 and develop COVID-19-like clinical symptoms (3, 48). Our analysis predicts that all Old World primates are susceptible to infection by SARS-CoV-2 via their ACE2 receptors. Thus, many of the 21 primate species native to China could be a potential reservoir for SARS-CoV-2. The remaining primate species were scored as *high* or *medium*, with only the Gray mouse lemur and the Philippine tarsier scoring as *low*.

We were surprised to find that all three species of Cervid deer and 12/14 cetacean species have *high* scores for binding of their ACE2s to SARS-CoV-2 S. There are 18 species of Cervid deer found in China. Therefore, Cervid deer cannot be ruled out as an intermediate host for SARS-CoV-2. While coronavirus sequences have been found in white tailed deer (53) and gammacoronaviruses have been found in beluga whales (54, 55) and bottlenose dolphins (56) and are associated with respiratory diseases, the cellular receptor used by these viruses is not known.

Other artiodactyls. A relatively large fraction (21/30) of artiodactyl mammals were classified with *medium* score for ACE2 binding to SARS-CoV-2 S. These include many species that are commonly found in Hubei Province and around the world, such as domesticated cattle, sheep and goats, as well as many species commonly found in zoos and wildlife parks (e.g., Masai giraffe, okapi, hippopotamus, water buffalo, scimitar horned oryx, and Dama gazelle). Although cattle MDBK cells were shown in one study to be resistant to SARS-CoV-2 *in vitro* (50), we propose immediate surveillance of common artiodactyl species for SARS-CoV-2 and studies of cellular infectivity, given our predictions. If ruminant artiodactyls can serve as a reservoir for SARS-CoV-2, it would have significant epidemiological implications as well as implications for food production and wildlife management (see below). It is noteworthy that camels and pigs, known for their ability to be infected by coronaviruses (35), both score *low* in our analysis. These

11

data are consistent with results (discussed above) indicating that pigs cannot be infected with SARS-CoV-2 both *in vivo* (47) and *in vitro* (50).

Rodents. Among the rodents, 7/46 species score *high* for ACE2 binding to SARS-CoV-2 S, with the remaining 11, 10 and 18 scoring *medium*, *low* or *very low*, respectively. Brown rats *(Rattus norvegicus)* and the house mouse (*Mus musculus*), scored *very low*, consistent with infectivity studies (1, 50). Given that wild rodent species likely come in contact with bats as well as with other predicted high risk species, we urge surveillance of *high* and *medium* binding likelihood rodents for the presence of SARS-CoV-2.

Bats and other species of interest. Chiroptera (bats) represent a clade of mammals that are of high interest in COVID-19 research because several bat species are known to harbor coronaviruses, including those most closely related to the betacoronavirus SARS-CoV-2 (1). We analyzed ACE2 from 37 bat species of which 8 and 29 scored *low* and *very low*, respectively. These results were unexpected because the three *Rhinolophus* spp. including the Chinese rufous horseshoe bat are major suspects in the transmission of SARS-CoV-2, or a closely related virus, to humans (1). Globally, bats have been shown to harbour the highest diversity of betacoronaviruses in mammals tested (35) and show little pathology carrying these viruses (57). We found evidence for accelerated evolution at six RBD binding domain residues within the bat lineage, which is more than in any other lineage tested. Bats also had far more sites showing evidence of positive selection, including four binding domain residues, compared to other mammalian orders. This suggests that the diversity observed in bat ACE2 sequences may be driven by selective pressure from coronaviruses. Our results suggest that SARS-CoV-2 is not likely to use the ACE2 receptor in bats, which challenges a recent study showing that SARS-CoV-2 can infect HeLa cells expressing *Rhinolophus sinicus* ACE2 (1). If bats can be infected with SARS-CoV-2, the virus likely uses a different receptor. For example, the MERS-CoV, a betacoronavirus, uses CD26/DPP4 (16) while the porcine transmissible enteritis virus, an alphacoronavirus uses aminopeptidase N (ANPEP) (58). As detailed above, further *in vitro* and *in vivo* infectivity studies are required to fully understand the mode of transmission of susceptibility of bats to SARS-CoV-2.

Carnivores. Recent reports of a Malayan tiger and a domestic cat infected by SARS-CoV-2 suggest that the virus can be transmitted to other felids (9, 47). Our results are consistent with these studies; 9/9 felids we analyzed scored *medium* for ACE2 binding of SARS-CoV-2 S. However, the masked palm civet (*Paguma larvata*), a member of the Viverridae family that is related to but distinct from Felidae, scored as *very low*. These results are inconsistent with transfection studies using civet ACE2 receptors expressed in HeLa cells (1), although these experiments have limitations as discussed above. While

12

carnivores closely related to dogs (dingos, wolves and foxes) all scored *low*, experimental data supporting infection in dogs were inconsistent (47, 50, 59) so no conclusions can be drawn.

Pangolins. Considerable controversy surrounds reports that pangolins can serve as an intermediate host for SARS-CoV-2. Pangolins were proposed as a possible intermediate host (22) and have been shown to harbor related coronaviruses. In our study, ACE2 of Chinese pangolin (*Manis pentadactyla*), Sunda pangolin (*Manis javanica*), and white bellied pangolin (*Phataginus tricuspis*) had *low* or *very low* binding score for SARS-CoV-2 S. Neither experimental infection nor *in vitro* infection with SARS-CoV-2 has been reported for pangolins. As for ferrets and bats, if SARS-CoV-2 infects pangolins it may be using a receptor other than ACE2, based on our analysis.

Other vertebrates. Our analysis of 29 orders of fishes, 29 orders of birds, 3 orders of reptiles and 2 orders of amphibians predicts that the ACE2 proteins of species within these vertebrate classes are not likely to bind SARS-CoV-2 S. Thus, vertebrate classes other than mammals are not likely to be an intermediate host or reservoir for the virus, despite predictions reported in a recent study (45), unless SARS-CoV-2 can use another receptor for infection. With many different non-mammal vertebrates sold in the seafood and wildlife markets of Asia and elsewhere, it is still important to determine if SARS-CoV-2 can be found in non-mammalian vertebrates.

Relevance to Threatened Species. Among the 103 species that scored *very high*, *high* and *medium* for ACE2 SARS-CoV-2 S RBD binding, 41 (40%) are classified in one of three 'Threatened' categories (*Vulnerable*, *Endangered*, and *Critically Endangered*) on the IUCN Red List of Threatened Species, five are classified as *Near Threatened*, and two species are classified as *Extinct in the Wild* (Dataset S1)(60). This represents only a small fraction of the threatened species potentially susceptible to SARS-CoV-2. For example, all 20 catarrhine primate species in our analysis, representing three families (Cercopithecidae, Hylobatidae, and Hominidae) scored *very high*, suggesting that all 185 species of catarrhine primates, most of which are classified Threatened (62), are potentially susceptible to SARS-CoV-2. Similarly, all three species of deer, representatives of a family of ~92 species (Cervidae), scored as *high* risk, as did species representing Cetacea (baleen and toothed whales), and both groups contain a number of threatened species. Toothed whales have potential for viral outbreaks and have lost function of a gene key to the antiviral response in other mammalian lineages (61). If they are susceptible to SARS-CoV-2 , human-to-animal transmission could pose a risk through sewage outfall (62) and contaminated refuse from cities, commercial vessels and cruise liners (63). In contrast, some threatened species scored *low* or *very low*, such as the giant panda (*low*), potentially positive news for these at risk populations.

Our results have practical implications for populations of threatened species in the wild and those under human care (including those in zoos). Established guidelines for minimizing potential human to animal transmission should be implemented and strictly followed. Guidelines for field researchers working on great apes established by the IUCN have been in place since 2015 in response to previous human disease outbreaks (64) and have received renewed attention because of SARS-CoV-2 (64–66). For zoos, guidelines in response to SARS-CoV-2 have been distributed by several Taxon Advisory Groups of the North American Association of Zoos and Aquariums (AZA), the American Association of Zoo Veterinarians (AAZV), and the European Association of Zoo and Wildlife Veterinarians (EAZWV), and these organizations are actively monitoring and updating knowledge of species in human care considered to be potentially sensitive to infection (67, 68). Although *in silico* studies suggest potential susceptibility of diverse species, verification of infection potential is warranted, using cell cultures, stem cells, organoids, and other methods that do not require direct animal infection studies. Zoos and other facilities that maintain living animal collections are in a position to provide such samples for generating crucial research resources by banking tissues, and cryobanking viable cell cultures in support of these efforts.

Animal models for COVID-19. A variety of animal models have been developed for studying SARS and MERS coronavirus infections (69). Presently, there is a tremendous need for animal models for studying SARS-CoV-2 infection and pathogenesis, as the only species currently known to be infected and show similar symptoms of COVID-19 is rhesus macaque. Non-human primate models have proven to be highly valuable for other infectious diseases, but are expensive to maintain and numbers of experimental animals are limited. Our results provide an extended list of potential species that might be useful as animal models for SARS-CoV-2 infection and pathogenesis, including Chinese hamster and Syrian/Golden hamster (49), and large animals maintained for biomedical and agricultural research (e.g., domesticated sheep and cattle).

Conclusions. We predict that species scored as *very high* and *high* for SARS-CoV-2 S binding to ACE2 will have a high probability of becoming infected by the virus. We also predict that many species having a *medium* score have some risk of infection, and species scored as *very low* and *low* are unlikely to be infected by SARS-CoV-2 via the ACE2 receptor. Importantly, our predictions are based solely on *in silico* analyses and must be confirmed by direct experimental data. Until such time, other than for species in which SARS-CoV-2 infection has been demonstrated to occur using ACE2, we urge caution not to over-interpret the predictions made in the present study. This is especially important with regards to species, endangered or otherwise, in human care. While species ranked *high* or *medium* may be

14

susceptible to infection based on the features of their ACE2 residues, pathological outcomes may be very different among species depending on other mechanisms that could affect virus replication and spread to target cells, tissues, and organs within the host. Furthermore, we cannot exclude the possibility that infection in any species occurs via another cellular receptor, as has been shown for other betacoronaviruses. Nonetheless, our predictions provide a useful starting point for selection of appropriate animal models for COVID-19 research and for identification of species that may be at risk for human-to-animal or animal-to-animal transmissions by SARS-CoV-2. The approach we used for ACE2 can be extended to other cellular proteins known to be involved in coronavirus infection and immunity to better understand infection, transmission, inflammatory responses and disease progression.

Materials and Methods

Angiotensin I converting enzyme 2 (ACE2) coding and protein sequences. All human ACE2 orthologs for vertebrate species, and their respective coding sequences, were retrieved from NCBI Protein (March 20, 2020) (70). ACE2 coding DNA sequences were extracted from available or recently sequenced unpublished genome assemblies for 123 other mammalian species, with the help of genome alignments and the human or within-family ACE2 orthologs. The protein sequences were predicted using AUGUSTUS v3.3.2 (71) or CESAR v2.0 (72) and the translated protein sequences were checked against the human ACE2 orthologue. ACE2 gene predictions were inspected and manually curated if necessary. For four bat species (*Micronycteris hirsuta*, *Mormoops blainvillei*, *Tadarida brasiliensis* and *Pteronotus parnellii*) the ACE2 coding region was split into two scaffolds which were merged, and for *Eonycteris spelaea* a putative 1bp frameshift base error was corrected. Eighty ACE2 predictions were obtained from the Zoonomia project, 19 from the Hiller Lab, 12 from the Koepfli lab, 8 from the Lewin lab and 4 from the Zhou lab. The source, and accession numbers for the genomes or proteins retrieved from NCBI are listed in Dataset S1. The final set of ACE2 sequences comprises 410 vertebrate species. To assure alignment robustness, the full set of coding and protein sequences were aligned independently using Clustal Omega (73), MUSCLE (74) and COBALT (75) all with default parameters. All resulting protein alignments were identical. Clustal Omega alignments were used in the subsequent analysis. Each amino acid replacement present in our dataset was classified as neutral, semi-conservative and non-conservative as in Clustal Omega.

Identification of ACE2 residues involved in binding to SARS-CoV-2 S protein. We identified 22 ACE2 protein residues that were previously reported to be critical for the effective binding of ACE2 RBD and

SARS-CoV-2 S (13, 21). These residues include S19, Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L45, L79, M82, Y83, N330, K353, G354, D355, R357, and R393. All these residues were identified from the co-crystallization and structural determination of SARS-CoV-2 S and ACE2 RBD (13, 21). The known human ACE2 RBD glycosylation sites N53, N90 and N322 were also included in the analyzed residue set (11).

ACE2 and SARS-CoV-2 binding ability prediction. Based on the known interactions of ACE2 and SARS-CoV-2 residues, we developed a set of rules for predicting the likelihood of the SARS-CoV-2 S binding to ACE2. Each species was classified in one of five categories: *very high*, *high*, *medium*, *low* or *very low* likelihood of binding SARS-CoV-2 S. Species in the *very high* category have at least 23/25 critical residues identical to the human; have K353, K31, E35, M82, N53, N90 and N322; do not have N79; and have only conservative substitutions among the non-identical 2/25 residues. Species in the *high* group have at least 20/25 residues identical to the human; have K353; have only conservative substitutions at K31 and E35; do not have N79; and can only have one non-conservative substitution among the 5/25 non-identical residues. Species scoring *medium* have at least 20/25 residues identical to the human; can only have conservative substitutions at K353, K31, and E35; and can have up to two non-conservative substitutions in the 5/25 non-identical residues. Species in the *low* category have at least 18/25 residues identical to the human; can only have conservative substitutions at K353; can have up to three non-conservative substitutions on the remaining 7/25 non-identical residues. Lastly, species in the *very low* group have less than 18/25 residues identical to the human or have at least four non-conservative substitutions in the non-identical residues.

Protein structure analysis. We applied an orthogonal approach to assess the likelihood of binding of a sampling of species that were predicted to bind SARS-CoV-2 across the categories of *high*, *medium*, *low* or *very low* likelihood of binding. ACE2 amino acid sequences from 28 species were extracted from the multiway alignment and loaded into SWISS-MODEL (25) in order to generate homology derived models. The output files were aligned to the crystal structure 6MOJ (13) in order to assess the overall similarities to human ACE2. We used two recently solved crystal structures of the complex for ACE2 and SARS-CoV-2 S RBD, 6MOJ (13) and 6VW1 (21) as ground truth for the human ACE2/SARS-CoV-2 S interaction. In the program CHIMERA (76), we utilized the rotamer function to model each individual variant that species exhibit separately, and chose the rotamer with the least number of clashes, retaining the most initial hydrogen bonds and containing the highest probability of formation as calculated by CHIMERA from the Dunbrack 2010 backbone-dependent rotamer library (77). The rotamer was then evaluated in the context of its structural environment and assigned a score based on likelihood of interface disruption.

16

Neutral (N) was assigned if the residue maintained a similar environment as the original residue, and was predicted to maintain or in some cases increase affinity. Weakened (W) was assigned if hydrophobic contacts were lost and contacts that appear disruptive are introduced that are not technically clashes. Unfavorable (U) was assigned if clashes are introduced and/or a hydrogen bond is broken. Additional structural visualizations were generated in Pymol (78).

Human variants analysis. All variants at the 25 residues critical for effective SARS-CoV-2-ACE2 binding (11, 21, 79) were compiled from from dbSNP (31), 1KGP (32), Topmed (33), UK10K (34) and CHINAMAP (28). Specific population frequencies were obtained from gnomAD v.2.1.1 (17).

Phylogenetic reconstruction of the vertebrate *ACE2* species tree. The multiple sequence alignment of 410 ACE2 orthologous protein sequences from mammals, birds, fishes, reptiles and amphibians was used to generate a gene tree using the maximum likelihood method of reconstruction, as implemented in IQTREE (80). The best fit model of sequence evolution was determined using ModelFinder (81) and used to generate the species phylogeny. A total of 1000 bootstrap replicates were used to determine node support using UFBoot (82).

Identifying sites undergoing positive selection. Signatures of site-specific positive selection in the *ACE2* receptor were explored using CODEML, part of the Phylogenetic Analysis using Maximum Likelihood (PAML, (83)) suite of software. Given CODEML's computational complexity, a smaller subset of mammalian taxa (N=64, Dataset S1), which included species from all prediction categories mentioned above, was used for selection analyses. To calculate likelihood-derived dN/dS rates (ω), CODEML utilises both a species tree and a codon alignment. The species tree for all 64 taxa was calculated using IQTREE (80) and the inferred best-fit model of sequence evolution (JTT+F+R4). This gene topology was generally in agreement with the 410 taxa tree, however bats were now sister taxa to Perissodactyla. Therefore all selection analyses were run using both the inferred gene tree, and a modified tree with the position of bats manually modified to reflect the 410 taxa topology. All species trees used were unrooted. A codon alignment of the 64 mammals was generated using pal2nal (84) with protein alignments generated with Clustal Omega (73) and their respective CDS sequences.

Site-models M7 (null model) and M8 (alternative model) were used to identify *ACE2* sites undergoing positive selection in mammals. Both M7 and M8 estimate ω using a beta distribution and 10 rate categories per site with $\omega<=1$ (neutral or purifying selection), but with an additional 11th category allowing $\omega >1$ (positive selection) in M8. A likelihood ratio test (LRT) calculated as $2*(lnL_{alt} - lnL_{null})$, comparing the fit of both null and alternative model likelihoods was carried out, with a p-value

17

calculated assuming a chi-squared distribution. Sites showing evidence of positive selection were identified by a significant (>0.95) Bayes Empirical Bayes (BEB) score, and validated by visual inspection of the protein alignment. To explore order-specific instances of positive selection, separate multiple sequence alignments and gene trees for Chiroptera (N=37), Cetartiodactyla (N=45), Carnivora (N=44), Rodentia (N=46) and Primates (N=39) were also generated and explored using M7 vs. M8 in CODEML.

In addition to site-models, branch-site model A1 (null model) and model A (alternative model) were also implemented targeting various mammalian orders, specifically Chiroptera, Cetartiodactyla, Rodentia and Primates, to identify lineage-specific positive selection in the *ACE2* receptor sequence. Branch-site Model A1 constrains both the target foreground branch (Carnivora, Chiroptera, Cetartiodactyla, Rodentia and Primates) and background branches to $\omega<=1$, while the alternative Model A allows positive selection to occur in the foreground branch. Null and alternative models were compared using LRTs as above, with significant BEB sites identified.

We also looked for positively selected sites in the viral spike protein, using SARS-CoV-2 (MN908947.3), Bat coronavirus RaTg13 (MN996532.1), Bat SARS-like coronavirus isolate Rs4231 (KY417146.1), SARS-related coronavirus strain BtKY72 (KY352407.1), SARS coronavirus Urbani (AY278741.1), SARS coronavirus PC4-227 (AY613950.1), Coronavirus BtRs-BetaCoV/YN2018B (MK211376.1) and the more divergent Bat Hp-betacoronavirus/Zhejiang2013 (NC_025217.1) viral strains. Protein and codon alignments were generated as above, with the viral species tree inferred using full genome alignments of all strains generated with Clustal Omega (73). Site-test models were applied using CODEML, and significant BEB sites identified.

Analysis for departure from neutral evolutionary rate in ACE2 with PHAST. Neutral models were trained on the specified species sets (Dataset S4) using the REV nucleotide substitution model implemented in phyloFit using an expectation maximization algorithm for parameter optimization. The neutral model fit was based on third codon positions to approximate the neutral evolution rate specific to the *ACE2* gene, using a 410-species phylogenetic tree generated by IQTREE as described above and rooted on fishes. The program phyloP was then used to identify codons undergoing accelerated or conserved evolution relative to the neutral model using --features to specify codons, --method LRT --mode CONACC, and --subtree for lineage-specific tests, with p-values thus assigned per codon based on a likelihood ratio test. P-values were corrected for multiple testing using the Benjamini-Hochberg method (36) and sites with a corrected p-value less than 0.05 were considered significant. PhyloFit and phyloP are both part of the PHAST package v1.4 (85, 86).

Acknowledgements

References

1. P. Zhou, *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).

2. G. Lu, Q. Wang, G. F. Gao, Bat-to-human: spike features determining "host jump" of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* **23**, 468–478 (2015).

3. C. Shan, *et al.*, Infection with Novel Coronavirus (SARS-CoV-2) Causes Pneumonia in the Rhesus Macaques https:/doi.org/10.21203/rs.2.25200/v1.

4. H. Laude, K. Van Reeth, M. Pensaert, Porcine respiratory coronavirus: molecular features and virus-host interactions. *Vet. Res.* **24**, 125–150 (1993).

5. L. J. Saif, Bovine respiratory coronavirus. *Vet. Clin. North Am. Food Anim. Pract.* **26**, 349–364 (2010).

6. W. Chen, *et al.*, SARS-associated Coronavirus Transmitted from Human to Pig. *Emerg. Infect. Dis.* **11**, 446 (2005).

7. L. Vijgen, *et al.*, Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J. Virol.* **79**, 1595–1604 (2005).

8. T. T.-Y. Lam, *et al.*, Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* (2020) https:/doi.org/10.1038/s41586-020-2169-0.

9. United States Department of Agriculture Animal and Plant Health Inspection Service, USDA APHIS | USDA Statement on the Confirmation of COVID-19 in a Tiger in New York (April 13, 2020).

10. Z. Qian, *et al.*, Innate Immune Response of Human Alveolar Type II Cells Infected with Severe Acute Respiratory Syndrome–Coronavirus. *Am. J. Respir. Cell Mol. Biol.* **48**, 742–748 (2013).

11. J. Sun, *et al.*, COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. *Trends Mol. Med.* (2020) https:/doi.org/10.1016/j.molmed.2020.02.008.

12. W. Li, *et al.*, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).

13. J. Lan, *et al.*, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* (2020) https:/doi.org/10.1038/s41586-020-2180-5.

14. V. B. Patel, J.-C. Zhong, M. B. Grant, G. Y. Oudit, Role of the ACE2/Angiotensin 1-7 Axis of the Renin-Angiotensin System in Heart Failure. *Circ. Res.* **118**, 1313–1326 (2016).

15. Y. Feng, *et al.*, Angiotensin-converting enzyme 2 overexpression in the subfornical organ prevents the angiotensin II--mediated pressor and drinking responses and is associated with angiotensin II type 1 receptor downregulation. *Circ. Res.* **102**, 729–736 (2008).

16. V. S. Raj, *et al.*, Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495**, 251–254 (2013).

17. K. J. Karczewski, *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv*, 531210 (2020).

18. E. W. Stawiski, *et al.*, Human ACE2 receptor polymorphisms predict SARS-CoV-2 susceptibility. *bioRxiv*, 2020.04.07.024752 (2020).

19. F. Li, Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Res.* **100**, 246–254 (2013).

20. F. Li, W. Li, M. Farzan, S. C. Harrison, Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).

21. J. Shang, *et al.*, Structural basis of receptor recognition by SARS-CoV-2. *Nature* (2020) https:/doi.org/10.1038/s41586-020-2179-y.

22. T. Zhang, Q. Wu, Z. Zhang, Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr. Biol.* **30**, 1346–1351.e2 (2020).

23. Z. Liu, *et al.*, Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J. Med. Virol.* (2020) https:/doi.org/10.1002/jmv.25726.

24. E. F. Pettersen, *et al.*, UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

25. A. Waterhouse, *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

26. E. Procko, The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS

coronavirus 2. *bioRxiv*, 2020.03.16.994236 (2020).

27. M. Hussain, *et al.*, Structural Variations in Human ACE2 may Influence its Binding with SARS-CoV-2 Spike Protein. *J. Med. Virol.* (2020) https:/doi.org/10.1002/jmv.25832.

28. Y. Cao, *et al.*, Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discov* **6**, 11 (2020).

29. A. Renieri, *et al.*, ACE2 variants underlie interindividual variability and susceptibility to COVID-19 in Italian population. *medRxiv* (2020) https:/doi.org/10.1101/2020.04.03.20047977.

30. H. Othman, *et al.*, Interaction of the spike protein RBD from SARS-CoV-2 with ACE2: similarity with SARS-CoV, hot-spot analysis and effect of the receptor polymorphism. *bioRxiv*, 2020.03.04.976027 (2020).

31. S. T. Sherry, dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* **29**, 308–311 (2001).

32. B. F. Voight, *et al.*, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

33. NHLBI, Trans-Omics for Precision Medicine WGS-About TOPMed (April 14, 2020).

34. UK10K Consortium, *et al.*, The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).

35. S. J. Anthony, *et al.*, Global patterns in coronavirus diversity. *Virus Evol* **3**, vex012 (2017).

36. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

37. G. Y. Oudit, M. A. Crackower, P. H. Backx, J. M. Penninger, The role of ACE2 in cardiovascular physiology. *Trends Cardiovasc. Med.* **13**, 93–101 (2003).

38. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).

39. D. Jebb, *et al.*, Six new reference-quality bat genomes illuminate the molecular basis and evolution of bat adaptations. *bioRxiv*, 836874 (2019).

40. K.-P. Koepfli, B. Paten, G. 10k C. of Scientists, S. J. O'Brien, The Genome 10K Project: a way forward. *Annu. Rev. Anim. Biosci.* **3**, 57–111 (2015).

41. E. S. Brielle, D. Schneidman-Duhovny, M. Linial, The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor. *bioRxiv*, 2020.03.10.986398 (2020).

42. M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **5**, 562–569 (2020).

43. J. Luan, Y. Lu, X. Jin, L. Zhang, Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochem. Biophys. Res. Commun.* (2020) https:/doi.org/10.1016/j.bbrc.2020.03.047.

44. A. D. Melin, M. C. Janiak, F. Marrone, P. S. Arora, J. P. Higham, Comparative ACE2 variation and primate COVID-19 risk. *bioRxiv*, 2020.04.09.034967 (2020).

45. Y. Qiu, *et al.*, Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2. *Microbes Infect.* (2020) https:/doi.org/10.1016/j.micinf.2020.03.003.

46. Y. Wan, J. Shang, R. Graham, R. S. Baric, F. Li, Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **94** (2020).

47. J. Shi, *et al.*, Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* (2020) https:/doi.org/10.1126/science.abb7015.

48. V. J. Munster, *et al.*, Respiratory disease and virus shedding in rhesus macaques inoculated with SARS-CoV-2. *bioRxiv*, 2020.03.21.001628 (2020).

49. J. F.-W. Chan, *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. *Clin. Infect. Dis.* (2020).

50. M. Hoffmann, *et al.*, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* (2020) https:/doi.org/10.1016/j.cell.2020.02.052.

51. X. Xie, J. Chen, X. Wang, F. Zhang, Y. Liu, Age- and gender-related difference of ACE2 expression in rat lung. *Life Sci.* **78**, 2166–2171 (2006).

52. K. Sun, L. Gu, L. Ma, Y. Duan, Atlas of ACE2 gene expression in mammals reveals novel insights in transmisson of SARS-Cov-2. *Genomics*, 536 (2020).

53. K. P. Alekseev, *et al.*, Bovine-like coronaviruses isolated from four species of captive wild ruminants are homologous to bovine coronaviruses, based on complete genomic sequences. *J. Virol.* **82**, 12422–12431 (2008).

54. K. A. Mihindukulasuriya, G. Wu, J. St. Leger, R. W. Nordhausen, D. Wang, Identification of a Novel Coronavirus from a Beluga Whale by Using a Panviral Microarray. *J. Virol.* **82**, 5084–5088 (2008).

55. H. Schütze, "Coronaviruses in Aquatic Organisms" in *Aquaculture Virology*, (Academic Press, 2016), pp. 327–335.

56. P. C. Y. Woo, *et al.*, Discovery of a Novel Bottlenose Dolphin Coronavirus Reveals a Distinct Species of Marine Mammal Coronavirus in Gammacoronavirus. *J. Virol.* **88**, 1318–1331 (2014).

57. A. Banerjee, *et al.*, Novel Insights Into Immune Systems of Bats. *Front. Immunol.* **11**, 26 (2020).

58. B. Delmas, *et al.*, Aminopeptidase N is a major receptor for the enteropathogenic coronavirus TGEV. *Nature* **357**, 417–420 (1992).

59. S. Temmam, A. Barbarino, D. Maso, S. Behillil, V. Enouf, Absence of SARS-CoV-2 infection in cats and dogs in close contact with a cluster of COVID-19 patients in a veterinary campus. *bioRxiv* (2020).

60. IUCN, The IUCN Red List of Threatened Species. Version 2019-2. *http://www.iucnredlist.org* (2019).

61. B. A. Braun, A. Marcovitz, J. G. Camp, R. Jia, G. Bejerano, Mx1 and Mx2 key antiviral proteins are surprisingly lost in toothed whales. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8036–8040 (2015).

62. A. Bosch, F. Xavier Abad, R. M. Pintó, "Human Pathogenic Viruses in the Marine Environment" in *Oceans and Health: Pathogens in the Marine Environment*, (Springer, Boston, MA, 2005), pp. 109–131.

63. C. Copeland, Cruise Ship Pollution: Background, Laws and Regulations, and Key Issues. *Congressional Research Service, The Library of Congress* **RL32450** (2005).

64. K. V. K. Gilardi, *et al.*, Best practice guidelines for health monitoring and disease control in great ape populations (2015).

65. A. Estrada, *et al.*, Impending extinction crisis of the world's primates: Why primates matter. *Sci Adv* **3**, e1600946 (2017).

66. T. R. Gillespie, F. H. Leendertz, COVID-19: protect great apes during human pandemics. *Nature* **579**, 497 (2020).

67. J. Johnson, A. Moresco, S. Han, SARS-COV-2 Considerations and Precautions. *AZA Small Carnivore Taxon Advisory Group* (2020).

68. A. Lecu, M. Bertelsen, C. Walzer, EAZWV Infectious Diseases Working Group, Science-based facts & knowledge about wild animals, zoos, and SARS-CoV-2 Virus. *European Association of Zoo and Wildlife Veterinarians - Transmissible Diseases Handbook* (2020).

69. T. C. Sutton, K. Subbarao, Development of animal models against emerging coronaviruses: From SARS to MERS coronavirus. *Virology* **479-480**, 247–258 (2015).

70. NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7 (2016).

71. B. M. Mario Stanke, AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465 (2005).

72. V. Sharma, P. Schwede, M. Hiller, CESAR 2.0 substantially improves speed and accuracy of comparative gene annotation. *Bioinformatics* **33**, 3985–3987 (2017).

73. F. Sievers, D. G. Higgins, Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. *Methods Mol. Biol.* **1079** (2014).

74. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

75. J. S. Papadopoulos, R. Agarwala, COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073–1079 (2007).

76. E. F. Pettersen, *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J.*

Comput. Chem. **25**, 1605–1612 (2004).

77. M. V. Shapovalov, R. L. Dunbrack Jr, A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **19**, 844–858 (2011).

78. PyMOL, The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. *https://pymol.org/2/* (2020).

79. J. Lan, *et al.*, Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor https:/doi.org/10.1101/2020.02.19.956235.

80. B. Q. Minh, *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* (2020) https:/doi.org/10.1093/molbev/msaa015.

81. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

82. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

83. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

84. M. Suyama, D. Torrents, P. Bork, PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–12 (2006).

85. M. J. Hubisz, K. S. Pollard, A. Siepel, PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).

86. R. Ramani, K. Krumholz, Y.-F. Huang, A. Siepel, PhastWeb: a web interface for evolutionary conservation scoring of multiple sequence alignments using phastCons and phyloP. *Bioinformatics* **35**, 2320–2322 (2019).

Figures and Tables

Column positions (headers): ID, S19, Q24, T27, F28, D30, K31*, H34, E35*, E37, D38, Y41, Q42, L45, N53*, L79, M82*, Y83, N90*, N322*, N330, K353*, G354, D355, R357, R393

VERY HIGH

Species	ID
Homo sapiens (Human)	25
Gorilla gorilla gorilla (Western lowland gorilla)	25
Nomascus leucogenys (Northern white-cheeked gibbon)	25
Pongo abelii (Sumatran orangutan)	25
Macaca fascicularis (Crab-eating macaque)	25
Mandrillus leucophaeus (Drill)	25
Nasalis larvatus (Proboscis monkey)	25
Pan paniscus (Bonobo)	25
Pan troglodytes (Chimpanzee)	25
Piliocolobus tephrosceles (Ugandan red colobus)	25
Pygathrix nemaeus (Red-shanked douc)	25
Rhinopithecus roxellana (Golden snub-nosed monkey)	25
Chlorocebus sabaeus (Green monkey)	25
Erythrocebus patas (Patas monkey)	25
Macaca mulatta (Rhesus macaque)	25
Papio anubis (Olive baboon)	25
Theropithecus gelada (Gelada)	25
Cercocebus atys (Sooty mangabey)	25
Macaca nemestrina (Southern pig-tailed macaque)	25

HIGH

Species	ID	notable substitutions
Colobus angolensis (Angola colobus)	24	R (N322*)
Propithecus coquereli (Coquerel's sifaka)	24	T (M82*)
Cricetomys gambianus (Gambian pouched rat)	22	A (T27), Q (D30), N (L79)
Cricetulus griseus (Chinese hamster)	22	Q (D30), N (L79), H (N90*)
Ctenodactylus gundi (Common gundi)	22	Q (D30), N (L79), T (M82*)
Delphinapterus leucas (Beluga whale)	22	Q (D30), I (L79), T (M82*)
Eulemur flavifrons (Blue-eyed black lemur)	22	E (S19), A (Q24), T (M82*)
Indri indri (Indri)	22	N (H34), Q (E35*), N (L79)
Monodon monoceros (Narwhal)	22	Q (D30), I (L79), T (M82*)
Neophocaena asiaeorientalis (Narrow-ridged finless porpoise)	22	Q (D30), I (L79), T (M82*)
Phocoena phocoena (Harbour porpoise)	22	Q (D30), I (L79), T (M82*)
Balaenoptera acutorostrata scammoni (Minke whale)	21	Q (D30), R (Y83), I T (L79, M82*)
Balaenoptera bonaerensis (Antarctic minke whale)	21	Q (D30), R (Y83), I T (L79, M82*)
Eschrichtius robustus (Gray whale)	21	Q (D30), E (N53*), I T (L79, M82*)
Nannospalax galili (Spalax)	21	Q (F28), K (N322*), D Q (K353*, G354)
Odocoileus virginianus texanus (White-tailed deer)	21	E (D30), M T (L79, M82*), H (N90*)
Rangifer tarandus (Reindeer)	21	E (D30), M T (L79, M82*), H (N90*)
Tamandua tetradactyla (Southern tamandua)	21	E (D30), Q (Q42), I T (L79, M82*)
Dipodomys stephensi (Stephens's kangaroo rat)	20	L (Q24), N Q (H34, E35*), I (L79), K (N330)
Elaphurus davidianus (Pere David's deer)	20	E N (D30, K31*), M T (L79, M82*), H (N90*)
Ellobius lutescens (Transcaucasian mole vole)	20	D (Q24), Q (E35*), S (N322*), H (N330), D (D355)
Globicephala melas (Long-finned pilot whale)	20	R (S19), Q (F28), R (H34), I T (L79, M82*)
Lagenorhynchus obliquidens (Pacific white-sided dolphin)	20	R (S19), Q (F28), R (H34), I T (L79, M82*)
Lipotes vexillifer (Baiji)	20	R (S19), Q (F28), I T F (L79, M82*, Y83)
Myrmecophaga tridactyla (Giant anteater)	20	E (D30), Q (E35*), N (L45), I T (L79, M82*)
Ondatra zibethicus (Muskrat)	20	N (K31*), Q (E35*), S (N322*), H (N330), D (D355)
Orcinus orca (Killer whale)	20	R (S19), Q (F28), R (H34), I T (L79, M82*)
Tursiops truncatus (Common bottlenose dolphin)	20	R (S19), Q (F28), R (H34), I T (L79, M82*)

MEDIUM

Species	ID	notable substitutions
Daubentonia madagascariensis (Aye-aye)	23	F (S19), T (M82*)
Cheirogaleus medius (Fat-tailed dwarf lemur)	22	T (M82*), K Q (K353*, G354)
Ictidomys tridecemlineatus (Thirteen-lined ground squirrel)	22	L (Q24), Q (E35*), A (N53*)
Marmota flaviventris (Yellow-bellied marmot)	22	L (Q24), Q (E35*), A (N53*)
Marmota marmota marmota (Alpine marmot)	22	L (Q24), Q (E35*), A (N53*)
Mesocricetus auratus (Golden hamster)	22	Q (D30), N (L79), Y (Y83)
Physeter catodon (Sperm whale)	22	Q (D30), T T (M82*, Y83)
Spermophilus dauricus (Daurian ground squirrel)	22	L (Q24), Q (E35*), A (N53*)
Allactaga bullata (Gobi jerboa)	21	T (Q24), Q (E35*), T (M82*), H (N90*)
Ammotragus lervia (Barbary sheep)	21	E (D30), M T (L79, M82*), Y (Y83)
Antilocapra americana (Pronghorn)	21	E (D30), M T (L79, M82*), Y (Y83)
Aotus nancymaae (Nancy Ma's night monkey)	21	H E (N53*, L79), T (M82*), Q (G354)
Beatragus hunteri (Hirola)	21	E (D30), M T (L79, M82*), Y (Y83)
Bison bison bison (American bison)	21	E (D30), M T (L79, M82*), Y (Y83)
Bos indicus (Zebu)	21	E (D30), M T (L79, M82*), Y (Y83)
Bos mutus (Wild yak)	21	K (K31*), M T (L79, M82*), Y (Y83)
Bos taurus (Cattle)	21	E (D30), M T (L79, M82*), Y (Y83)
Bubalus bubalis (Water buffalo)	21	E (D30), M T (L79, M82*), Y (Y83)
Callicebus donacophilus (White-eared titi)	21	H E (N53*, L79), T (M82*), Q (G354)
Callithrix jacchus (Common marmoset)	21	H E (N53*, L79), T (M82*), Q (G354)
Capra aegagrus (Wild goat)	21	E (D30), M T (L79, M82*), Y (Y83)
Capra hircus (Goat)	21	E (D30), M T (L79, M82*), Y (Y83)
Cebus capucinus imitator (Panamanian white-faced capuchin)	21	H E (N53*, L79), T (M82*), Q (G354)
Felis catus (Cat)	21	L (Q24), E (D30), E (N53*), T (M82*)
Giraffa tippelskirchi (Masai giraffe)	21	E (D30), M T (L79, M82*), Y (Y83)
Hemitragus hylocrius (Nilgiri tahr)	21	E (D30), M T (L79, M82*), Y (Y83)
Lynx canadensis (Canadian lynx)	21	L (Q24), E (D30), E (N53*), T (M82*)
Mirza coquereli (Coquerel's giant mouse lemur)	21	I (T27), N (H34), T (M82*), K (N330)
Moschus moschiferus (Siberian musk deer)	21	E (D30), M T (L79, M82*), Y (Y83)
Neofelis diardi (Sunda clouded leopard)	21	L (Q24), E (D30), E (N53*), T (M82*)
Neofelis nebulosa (Clouded leopard)	21	L (Q24), E (D30), E (N53*), T (M82*)
Okapia johnstoni (Okapi)	21	E (D30), M T (L79, M82*), Y (Y83)
Ovis aries (Sheep)	21	E (D30), M T (L79, M82*), Y (Y83)
Panthera onca (Jaguar)	21	L (Q24), E (D30), E (N53*), T (M82*)
Panthera pardus (Leopard)	21	L (Q24), E (D30), E (N53*), T (M82*)
Panthera tigris altaica (Siberian tiger)	21	L (Q24), E (D30), E (N53*), T (M82*)
Pantholops hodgsonii (Tibetan antelope)	21	E (D30), M T (L79, M82*), Y (Y83)
Perognathus longimembris (Little pocket mouse)	21	V (Q24), Q (E35*), T (M82*), K (N330)
Peromyscus maniculatus bairdii (Deer mouse)	21	I (T27), Q (E35*), N (L79), H (N90*)
Pithecia pithecia (White-faced saki)	21	H E (N53*, L79), T (M82*), Q (G354)
Puma concolor (Cougar)	21	L (Q24), E (D30), E (N53*), T (M82*)
Saimiri boliviensis boliviensis (Black-capped squirrel monkey)	21	H E (N53*, L79), T (M82*), Q (G354)
Sapajus apella (Tufted capuchin)	21	H E (N53*, L79), T (M82*), Q (G354)
Urocitellus parryii (Arctic ground squirrel)	21	L (Q24), Q (E35*), H (N90*), D (N322*)
Bos indicus x Bos taurus	21	E (D30), M T (L79, M82*), Y (Y83)
Acinonyx jubatus (Cheetah)	20	L (Q24), E (D30), E (N53*), T (M82*), K (N330)
Alouatta palliata (Mantled howler)	20	E (D30), H E (N53*, L79), T (M82*), Q (G354)
Ateles geoffroyi (Geoffroy's spider monkey)	20	A (S19), H E (N53*, L79), T (M82*), Q (G354)
Fukomys damarensis (Damaraland mole-rat)	20	L (S19), Q (E35*), A (M82*), H (N90*), N (N330)
Heterocephalus glaber (Naked mole-rat)	20	L (S19), Q (E35*), A (M82*), H (N90*), D (N330)
Hippopotamus amphibius (Hippopotamus)	20	L (N53*), D (L79), A (M82*), F D (N90*, N322*)
Lepus americanus (Snowshoe hare)	20	L (Q24), E (D30), Q (Q42), T (M82*), S (N322*)
Nanger dama (Dama gazelle)	20	E (D30), M T F (L79, M82*, Y83), Y (N90*)
Oryctolagus cuniculus (European rabbit)	20	L (Q24), E (D30), Y (Q42), T (M82*), S (N322*)
Oryx dammah (Scimitar oryx)	20	* (S19), E (D30), M T (L79, M82*), Y (Y83)
Saguinus imperator (Emperor tamarin)	20	A (Q24), H E (N53*, L79), T (M82*), Q (G354)
Vicugna pacos (Alpaca)	20	L (Q24), K E (K31*, D30), A I (M82*, Y83)

LOW

Species	ID	notable substitutions
Ceratotherium simum cottoni (Northern white rhinoceros)	21	L (Q24), E (D30), P (H34), T (M82*)
Ceratotherium simum simum (Southern white rhinoceros)	21	L (Q24), E (D30), P (H34), T (M82*)
Diceros bicornis (Black rhinoceros)	21	L (Q24), E (D30), P (H34), T (M82*)
Galeopterus variegatus (Sunda flying lemur)	21	F (S19), L (Q24), E (L79), N (N90*)
Peromyscus leucopus (White-footed mouse)	21	I (T27), Q (E35*), N (L79), P (K353*)
Ailuropoda melanoleuca (Giant panda)	20	L (Q24), E (D30), Y (Q42), H T (L79, M82*)
Camelus bactrianus (Bactrian camel)	20	L (Q24), E E (D30, K31*), T T (M82*, Y83)
Camelus dromedarius (Dromedary)	20	L (Q24), E E (D30, K31*), T T (M82*, Y83)
Camelus ferus (Wild bactrian camel)	20	L (Q24), E E (D30, K31*), T T (M82*, Y83)
Dicerorhinus sumatrensis (Sumatran rhinoceros)	20	L (Q24), E (D30), P (H34), H (N53*), T (M82*)
Graphiurus murinus (Woodland dormouse)	20	T (Q24), N R (H34, E35*), H N (N90*, N322*)
Tapirus indicus (Malayan tapir)	20	L (Q24), E (D30), P (H34), H (N53*), T (M82*)
Tapirus terrestris (South American tapir)	20	L (Q24), E (D30), P (H34), H (N53*), T (M82*)
Ursus arctos horribilis (Grizzly bear)	20	L (Q24), E (D30), Y (Q42), H T (L79, M82*)
Ursus maritimus (Polar bear)	20	L (Q24), E (D30), Y (Q42), H T (L79, M82*)
Canis lupus dingo (Dingo)	19	L (Q24), E (D30), Y (Q42), E (L79), T (M82*), D (N322*)
Canis lupus familiaris (Dog)	19	L (Q24), E (D30), Y (Q42), E (L79), T (M82*), D (N322*)
Chinchilla lanigera (Long-tailed chinchilla)	19	N E K (H34, E35*, E37), A (M82*), H (N90*), D (N322*)
Chrysocyon brachyurus (Maned wolf)	19	L (Q24), E (D30), Y (Q42), E (L79), T (M82*), D (N322*)
Dipodomys ordii (Ord's kangaroo rat)	19	Q L (S19, Q24), N Q (H34, E35*), I (L79), K (N330)
Eonycteris spelaea (Lesser dawn bat)	19	L (Q24), E (D30), T (H34), T (M82*), D (N322*), K (N330)
Equus asinus (Donkey)	19	L (Q24), E (D30), S (E35*), E H (L45, N53*), H (N90*)
Equus caballus (Horse)	19	L (Q24), E (D30), S (E35*), E H (L45, N53*), H (N90*)

Column header positions (both tables): ID, S19, Q24, T27, F28, D30, K31*, H34, E35*, E37, D38, Y41, Q42, L45, N53*, L79, Y83, M82*, N90*, N322*, N330, K353*, G354, D355, R357, R393

LOW (continued)

ID	Species
19	*Equus przewalskii* (Przewalski's horse)
19	*Hydrochoerus hydrochaeris* (Capybara)
19	*Hystrix cristata* (Crested porcupine)
19	*Megaderma lyra* (Indian false vampire)
19	*Microtus ochrogaster* (Prairie vole)
19	*Rhinolophus pearsonii* (Pearson's horseshoe bat)
19	*Rhinolophus sinicus* (Chinese rufous horseshoe bat)
19	*Rousettus aegyptiacus* (Egyptian rousette)
19	*Speothos venaticus* (Bush dog)
19	*Sus scrofa* (Pig)
19	*Tragulus javanicus* (Java mouse-deer)
19	*Vulpes lagopus* (Arctic fox)
19	*Vulpes vulpes* (Red fox)
18	*Balaena mysticetus* (Bowhead whale)
18	***Carlito syrichta* (Philippine tarsier)**
18	*Dasyprocta punctata* (Central American agouti)
18	***Dolichotis patagonum* (Patagonian mara)**
18	***Eidolon helvum* (Straw-colored fruit bat)**
18	*Loxodonta africana* (African elephant)
18	*Microcebus murinus* (Gray mouse lemur)
18	*Ochotona princeps* (American pika)
18	*Octodon degus* (Common degu)
18	*Procavia capensis* (Rock hyrax)
18	*Pteropus alecto* (Black flying fox)
18	***Pteropus vampyrus* (Large flying fox)**
18	***Trichechus manatus latirostris* (West Indian manatee)**

VERY LOW

ID	Species
20	***Catagonus wagneri* (Chacoan peccary)**
19	*Jaculus jaculus* (Lesser Egyptian jerboa)
18	*Cavia porcellus* (Guinea pig)
18	*Cavia tschudii* (Montane guinea pig)
18	*Hipposideros armiger* (Great roundleaf bat)
18	*Hipposideros pratti* (Pratt's roundleaf bat)
18	*Mesoplodon bidens* (Sowerby's beaked whale)
18	*Spilogale gracilis* (Western spotted skunk)
18	*Zapus hudsonius* (Meadow jumping mouse)
17	***Ctenomys sociabilis* (Social tuco-tuco)**
17	*Cynopterus brachyotis* (Lesser short-nosed fruit bat)
17	*Cynopterus sphinx* (Greater short-nosed fruit bat)
17	***Enhydra lutris kenyoni* (Sea otter)**
17	***Eumetopias jubatus* (Steller sea lion)**
17	*Grammomys surdaster* (African woodland thicket rat)
17	***Gulo gulo* (Wolverine)**
17	*Heterohyrax brucei* (Yellow-spotted rock hyrax)
17	*Macroglossus sobrinus* (Long-tongued fruit bat)
17	***Manis javanica* (Sunda pangolin)**
17	***Manis pentadactyla* (Chinese pangolin)**
17	***Mellivora capensis* (Honey badger)**
17	*Mus pahari* (Gairdner's shrewmouse)
17	***Mustela erminea* (Stoat)**
17	***Mustela lutreola* (European mink)**
17	***Mustela nigripes* (Black-footed ferret)**
17	*Mustela putorius furo* (Ferret)
17	***Neomonachus schauinslandi* (Hawaiian monk seal)**
17	*Petromus typicus* (Dassie rat)
17	*Phoca vitulina* (Harbor seal)
17	***Pteronura brasiliensis* (Giant otter)**
17	*Rhinolophus ferrumequinum* (Greater horseshoe bat)
17	***Taxidea taxus* (American badger)**
17	*Thryonomys swinderianus* (Greater cane rat)
17	*Zalophus californianus* (California sea lion)
16	*Acomys cahirinus* (Cairo spiny mouse)
16	*Anoura caudifer* (Tailed tailless bat)

VERY LOW (continued)

| ID	Species
16	*Artibeus jamaicensis* (Jamaican fruit-eating bat)
16	***Callorhinus ursinus* (Northern fur seal)**
16	*Choloepus hoffmanni* (Hoffmann's two-toed sloth)
16	*Condylura cristata* (Star-nosed mole)
16	***Cryptoprocta ferox* (Fossa)**
16	*Dasypus novemcinctus* (Nine-banded armadillo)
16	*Hipposideros galeritus* (Cantor's roundleaf bat)
16	***Hyaena hyaena* (Striped hyena)**
16	*Miniopterus natalensis* (Natal long-fingered bat)
16	***Miniopterus schreibersii* (Schreibers' long-fingered bat)**
16	*Mirounga angustirostris* (Northern elephant seal)
16	*Mus caroli* (Ryukyu mouse)
16	*Mus musculus* (House mouse)
16	*Mus spretus* (Algerian mouse)
16	*Myocastor coypus* (Coypu)
16	*Myotis davidii* (David's myotis)
16	*Myotis myotis* (Greater mouse-eared bat)
16	*Noctilio leporinus* (Greater bulldog bat)
16	*Odobenus rosmarus divergens* (Walrus)
16	*Otolemur garnettii* (Northern greater galago)
16	*Paguma larvata* (Masked palm civet)
16	***Phataginus tricuspis* (White-bellied pangolin)**
16	*Psammomys obesus* (Fat sand rat)
16	*Rattus norvegicus* (Brown rat)
16	***Sarcophilus harrisii* (Tasmanian devil)**
15	***Ailurus fulgens styani* (Red panda)**
15	*Carollia perspicillata* (Seba's short-tailed bat)
15	*Chrysochloris asiatica* (Cape golden mole)
15	*Elephantulus edwardii* (Cape elephant shrew)
15	*Eptesicus fuscus* (Big brown bat)
15	*Helogale parvula* (Common dwarf mongoose)
15	*Mastomys coucha* (Southern multimammate mouse)
15	*Meriones unguiculatus* (Mongolian gerbil)
15	*Monodelphis domestica* (Gray short-tailed opossum)
15	*Mungos mungo* (Banded mongoose)
15	*Murina feae* (Little tube-nosed bat)
15	*Myotis brandtii* (Brandt's bat)
15	*Myotis lucifugus* (Little brown bat)
15	*Orycteropus afer afer* (Aardvark)
15	*Paradoxurus hermaphroditus* (Asian palm civet)
15	*Phyllostomus discolor* (Pale spear-nosed bat)
15	*Scalopus aquaticus* (Eastern mole)
15	*Sorex araneus* (Common shrew)
15	*Suricata suricatta* (Meerkat)
15	*Tadarida brasiliensis* (Brazilian free-tailed bat)
15	*Tonatia saurophila* (Stripe-headed round-eared bat)
14	*Microgale talazaci* (Talazac's shrew tenrec)
14	*Molossus molossus* (Velvety free-tailed bat)
14	*Mormoops blainvillei* (Antillean ghost-faced bat)
14	*Neovison vison* (American mink)
14	***Phascolarctos cinereus* (Koala)**
14	*Pteronotus parnellii* (Parnell's mustached bat)
14	***Solenodon paradoxus* (Hispaniolan solenodon)**
14	***Vombatus ursinus* (Common wombat)**
13	*Desmodus rotundus* (Common vampire bat)
13	*Echinops telfairi* (Lesser hedgehog tenrec)
13	*Erinaceus europaeus* (European hedgehog)
13	*Micronycteris hirsuta* (Hairy big-eared bat)
13	***Ornithorhynchus anatinus* (Platypus)**
13	*Pipistrellus kuhlii* (Kuhl's pipistrelle)
13	*Pipistrellus pipistrellus* (Common pipistrelle)
13	*Tupaia chinensis* (Northern treeshrew)
11	*Uropsilus gracilis* (Gracile shrew mole)

Figure 1. Cross-species conservation of ACE2 and predictions of SARS-CoV-2 susceptibility. Species are sorted by binding score of ACE2 for SARS-CoV-2 S. The 'ID' column depicts the number of amino acids identical to human binding residues. Bold amino acid positions (also labeled with *) represent residues at binding hotspots and constrained in the scoring scheme. Each amino acid substitution is colored according to its classification as non-conservative (orange), semi-conservative (yellow) or neutral (blue), as compared to the human residue. Bold species names depict species with threatened IUCN risk status. The 410 vertebrate species dataset is available in Dataset S1.
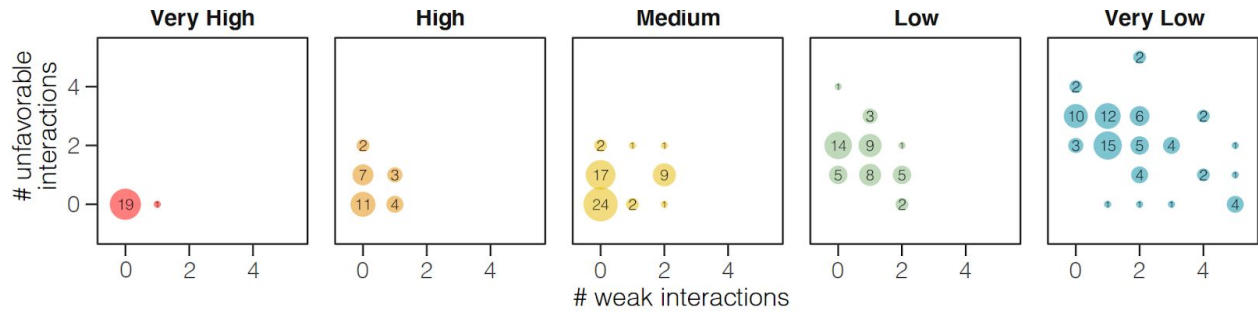
Figure 2. Congruence between binding score and structural homology analysis. Species classified by sequence identity to human ACE2 as *very high* (red) or *high* binding score (orange) have significantly fewer amino acid substitutions rated as potentially altering the binding interface between ACE2 and SARS-CoV-2 through protein structural analysis, as compared to *low* (green) or *very low* (blue) species. The more severe *unfavorable* variants are counted on y-axis and less severe *weaken* variants on the x-axis. Black numerical labels indicate species count.
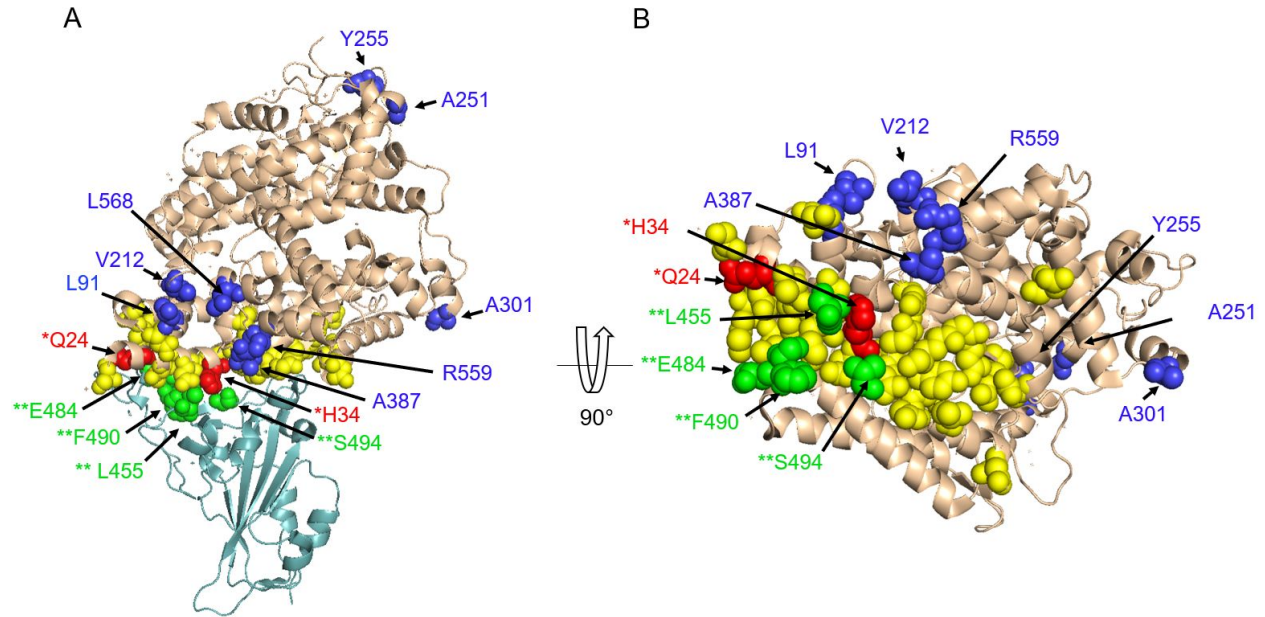
Figure 3. Residues under positive selection detected with CODEML and acceleration with phyloP in mammals. (**A**) ACE2 is represented in wheat cartoon with residues involved in the binding interface shown in yellow spheres. Dark blue and red spheres indicate residues in ACE2 that are accelerated and under positive selection. Red spheres represent residues that overlap with positions in the binding interface and are labeled with (*). The spike RBD is shown in light teal cartoon. Green spheres indicate residues on the SARS-CoV-2 spike protein under positive selection and are labeled with (**). (**B**) 90 degree rotation of the ACE2 protein.