Comparative analyses of ACE2 and TMPRSS2 gene: Implications for the risk to which vertebrate animals are susceptible to SARS-CoV-2

Chen Huang1 | Yu Jiang2 | Jie Yan3

1Department of Veterinary Science, University of Kentucky, Lexington, Kentucky, USA
2Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, Tennessee, USA
3Department of Diagnosis, School of Medicine, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China

Correspondence
Jie Yan, Department of Diagnosis, School of Medicine, Cheeloo College of Medicine, Shandong University, Jinan, Shandong 250012, China. Email: jyan@sdu.edu.cn
Yu Jiang, Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health. University of Memphis, Memphis, TN 38111, USA. Email: yjiang4@memphis.edu

Funding information
Qilu Young Scholarship (Shandong University)

Abstract
Along with the control and prevention of coronavirus disease 2019 transmission, infected animals might have potential to carry the virus to spark new outbreaks. However, very few studies explore the susceptibility of animals to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Viral attachment as a crucial step for cross-species infection requires angiotensin-converting enzyme 2 (ACE2) as a receptor and depends on TMPRSS2 protease activity. Here, we searched the genomes of metazoans from different classes using an extensive BLASTP survey and found ACE2 and TMPRSS2 occur in vertebrates, but some vertebrates lack Tmprss2. We identified 6 amino acids among 25 known human ACE2 residues are highly associated with the binding of ACE2 to SARS-CoV-2 (p value < .01) by Fisher exact test, and following this, calculated the probability of viral attachment within each species by the randomForest function from R randomForest library. Furthermore, we observed that Ace2 selected from seven animals based on the above analysis lack hydrophobic contacts identified on human ACE2, indicating less affinity of SARS-CoV-2 to Ace2 in animals than humans. Finally, the alignment of 3D structure between human ACE2 and other animals by I-TASSER and TM-align displayed a reasonable structure for viral attachment within these species. Taken together, our data may shed light on the human-to-animal transmission of SARS-CoV-2.

Keywords
ACE2, infection, livestock, SARS-CoV-2, TMPRSS2

1 | INTRODUCTION

An unusual pneumonia caused by human severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel coronavirus with ~79% similarity of the genome sequence with SARS-CoV, is threatening global public health and economy.1-3 This disease was named coronavirus disease 2019 (COVID-19) by the World Health Organization. Until March 9, 2021, SARS-CoV-2 has spread to more than 188 countries and regions and caused 117,236,336 COVID-19 positive cases with 2,603,942 deaths (https://coronavirus.jhu.edu/map.html). The origin and transmission routes of SARS-CoV-2 is still unclear. Generally, other coronaviruses SARS-CoV and Middle East respiratory syndrome coronavirus transmitted from bats to an intermediate host and then spillover to humans.4-6 Bats are considered as the original host of SARS-CoV-2 according to isolation of bat coronavirus RaTG13, which has 96.1% genome identity with SARS-CoV-2.7-8 Malayan pangolin is a potential intermediate host of SARS-CoV-2 because of the isolation of pangolin-CoV with 100%, 98.6%, 97.8%, and 90.7% amino acid identity with human SARS-CoV-2 in the E, M, N, and S proteins. Meanwhile, infected pangolins showed...
clinical signs and circulating antibodies against pangolin-CoV, which can react with S protein of SARS-CoV-2.9–11

In addition to the unclear molecular events in which intermediate host during bat-to-human transmission of SARS-CoV-2, the transmission of SARS-CoV-2 between human and other mammals is becoming the new concern along with more and more reported emerging of cases in animals. In the experimental approach, rhesus macaques, golden hamster, domestic cats, dogs, ferret, Egyptian fruit bat, and rabbit are susceptible to infection by SARS-CoV-2.12–28 To date, pet dogs and cats in contact with COVID-19 patients,29–31 lion and tiger in New York city zoo,30 and mink on farms in the Netherlands have tested positive for COVID-19.31–33 At this moment, the risk of catching SARS-CoV-2 from an infected animal is much lower than from COVID-19 positive person. Nevertheless, along with the restrictions on the movement of COVID-19 person and numbers of infected people fall, the virus might jump back and forth between humans and animals to spark a new outbreak. Therefore, to effectively prevent new spillovers to control future pandemics, more studies should be carried out to understand the susceptibility of the virus to various livestock.

Effective SARS-CoV-2 cell entry is the first line for viral infection and determines the host range. Coronaviruses entry into target cells depends on: (1) the binding of the spike (S) protein of coronaviruses to cellular receptors, which facilitates virus attachment to the surface of cells; (2) the requirement of S protein priming by cellular proteases, which undertakes S protein cleavage to fuse cell membranes and virus.34,35 The present studies indicated that SARS-CoV-2 employs angiotensin-converting enzyme 2 (ACE2) as the cellular receptors and requires the cellular serine protease transmembrane serine protease 2 (TMPRSS2) for S protein priming.36–41

ACE2 is a type I transmembrane protein that was first discovered in humans as a homologous of ACE.42–44 ACE2 has two domains, namely peptidase_M2 domain in N-terminus and collectin domain in C-terminus.45 Evidence shows that human ACE2 is widely distributed in several tissues, including the heart, kidney, lung, liver, testis, and intestine.43 Biological functions of ACE2 could be divided into two categories: peptidase-dependent and -independent. The peptidase-independent function of ACE2 refers to a receptor of SARS-CoV-2 for cell entry.45 TMPRSS2 is a type II transmembrane serine protease that was first identified in prostate cancer.46 TMPRSS2 has three domains: Low-density lipoprotein receptor domain class A (LDLa), scavenger receptor cysteine-rich domain (SR), and trypsin-like serine protease (TrypSPc) domain. TMPRSS2 is widely expressed in the prostate, intestine, stomach, liver, lung, and kidney and so on.47

In this study, we performed an extensive BLASTP search to identify ACE2 and TMPRSS2 gene homologs in lower vertebrate classes (cyclostomata, fishes, amphibians, reptiles, birds) and each order and family in mammals. We found ACE2 and TMPRSS2 originated from vertebrates, but TMPRSS2 disappeared in some vertebrates. The phylogenetic tree of ACE2 demonstrates a structure consistent with a topology law of species evolution and two main clades within highly advanced mammals along with SARS-CoV-2 positive case animals and cells, which is similar to those observed in the TMPRSS2 gene family. Importantly, the distribution of 25 known human ACE2 residues affinity to SARS-CoV-2 among different species was analyzed using Fisher exact test by SAS 9.4, and 6 of those residues is highly associated with viral susceptibility with p < .01. Subsequently, the probability of the interplay of SARS-CoV-2 with ACE2 across different species was predicted by random forest analysis for those highly associated amino acids using randomForest function from R randomForest library. To further confirm our prediction, the hydrophobicity, the prediction, and alignment of three-dimensional (3D) structure of ACE2 from seven selected species were analyzed. Compared to human ACE2, less binding of SARS-CoV-2 to ACE2 was observed in seven animals, but with a reasonable structure for viral cell entry. Overall, our study is taken to identify viral entry in potential hosts through sequence comparison, prediction of probability for infectious and homology modeling, which can provide clues to better understand the potential human-to-animals transmission of SARS-CoV-2.

2 | MATERIALS AND METHODS

2.1 | Gene extraction

All ACE, ACE2, and TMPRSS2 genes from different species present in this study were retrieved as follows: the amino acid sequences of Human ACE (Genbank number: NP_000780), Human ACE2 (Genbank number: NP_001358344) and Human TMPRSS2 (Genbank number: NP_001128571) were downloaded from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/). According to the topology law of species evolution, representative animals from mammals, aves, peptile, amphibian, teleost fishes, cyclostomata, arthropoda, and nematomorpha were selected. The ACE, ACE2, and TMPRSS2 genes from those representative animals were extracted based on the best hits using extensive BLASTP against NCBI and Ensembl (https://useast.ensembl.org/index.html) database with human ACE, ACE2, and TMPRSS2 as the queries.

2.2 | Sequence analysis

Functional domains or motifs in ACE2 and TMPRSS2 genes were confirmed by SMART (http://smart.embl-heidelberg.de/). The presence of signal peptide and transmembrane regions of ACE2 and TMPRSS2 were predicted by SignalP 5.0 server (http://www.cbs.dtu.dk/services/SignalP/) and TMHMM Server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The locations of N-glycosylation site were predicted by NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc).

2.3 | Evolutionary analysis

Multiple alignments of ACE, ACE2 and TMPRSS2 from representative species were performed using Clustal X. Phylogenetic
tree was constructed using Neighbor-Joining (NJ) by MEGA X package, and in each case, branch confidence values were obtained by bootstrapping with 1000 replications. Substitutions model were analyzed with p-distance correction, and gaps were removed by Pairwise deletion.

2.4 | Statistical analysis for prediction

The frequencies of amino acids were summarized by the status of SARS-CoV-2 infection for each interested position for the 13 species including 9 known species has positive receptor binding and 4 species has negative receptor binding. The association of amino acid with the status virus infection were tested using Fisher exact test by SAS 9.4. The heatmap of amino acids on each interested position for the 13 species were generated using Heatmap function from R library ComplexHeatmap (https://jokergoo.github.io/ComplexHeatmap-reference/book/). To further predict the probability of the recognition of ACE2 to SARS-CoV-2 from other species, we first choose the six position that have a p value less than .01 from the Fisher exact test, which are Ser19, Lys26, Thr27, Asp30, Leu79, and Met82. For these positions, if the amino acids are in species which are recognized by the virus, then it will be coded as 0, otherwise, the amino acids will be coded as unknown. Using the transferred data, we conducted a random forest analysis for the five positions using randomForest function from R randomForest library. The 13 species are treated as training datasets and the virus infection status for species are predicted accordingly.

2.5 | Protein structure analysis

The amino acid sequences of ACE2 from seven representative species were selected from the multiway alignment and loaded into GOR IV (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html) for secondary protein structure, DNAMAN for the hydrophobicity of amino acids, and I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER), ranked as the Top one server in template-based protein structure modeling, for 3D structural prediction. The alignment of 3D structure was performed by TM-align (https://zhanglab.ccmb.med.umich.edu/TM-align/).48,49

3 | RESULTS

3.1 | Summary of animals and cells affected by SARS-CoV-2

To date, limited emergence of COVID-19 cases in animals and cells has been reported. For animals, rhesus macaques, golden hamsters, domestic cats, dogs, ferret, Egyptian fruit bat, and rabbits are susceptible to experimental infection by SARS-CoV-2. Lion and tiger were confirmed by the United States Department of Agriculture (Table S1). Viral RNA can be detected in the related tissues or cell lines in all species listed in Table S1. The clinical sign of respiratory tract was confirmed in rhesus macaques, golden hamster, domestic cats, tiger, lion, mink, and ferret; Viral transmission was observed in rhesus macaques, golden hamster, domestic cats, lion, tiger, mink, ferret, and Egyptian fruit bat.2,12-23,25-29,31-33,50 For cell lines, SARS-CoV-2 can infect and replicate in human Caco2, Calu3, NHBE, and lung epithelial line A549.24,51-56 Meanwhile, SARS-CoV-2 can replicate in rhesus monkey LLCMK2, rabbit RK-13, cat CRFK, and pig ST, and PK-15 cells lines.25,57 Overall, the above reports revealed that the SARS-CoV-2 cell entry, or at the least, the recognition of ACE2 to SARS-CoV-2 has widely existed in livestock and wide animals. Given the reported important role of ACE2 and TMPRSS2 in SARS-CoV-2 cell entry, we subsequently carried out the bioinformatics analysis for these two genes to explore the potential clues to the identification of intermediate host and human-to-animals transmission of SARS-CoV-2.

3.2 | Sequence identification of ACE2

To define ACE2 genes in different species, amino acids of human ACE2 were used as a means to identify homologues in NCBI and Ensembl database by BLASTP. As shown in Table 1, the number of ACE2 homologous in almost all vertebrates and invertebrates is 1, only in fresh-water polyp Hydra vulgaris is it 2. Vertebrate ACE2 contains peptidase M2 domain and signal peptide in N-terminus, and collectrin domain and transmembrane region in C-terminus; However, collectrin domain and trans-membrane region are not present in homolog/orthologues from invertebrate fruit fly D. melanogaster, roundworm C. elegans and fresh-water polyp H. vulgaris (Figure S1). The absence of collectrin domain in inveterate homologous/orthologues is similar to vertebrate ACE, the orthologues of the ACE2 gene family. Consistently with the domain composition, the homology of invertebrate homolog/orthologues to human ACE is higher than human ACE2. For instance, the identity of fruit fly ACE to human ACE is 44.35%, while the identity to human ACE2 is 36.47%. To further examine the evolutionary relationship of ACE and ACE2, an NJ tree was created by MEGA X based on their amino acid sequences. As shown in Figure S1, the tree can be split into two clades, one clade contains vertebrate ACE2, and the other clade contains invertebrate homologous/orthologues and vertebrate ACE, suggesting invertebrate homologous/orthologues is closed to vertebrate ACE. Therefore, the earliest of ACE2 occurs invertebrate. The cryo-electron microscopy structures of the full-length human ACE2 indicated that the extracellular region is highly glycosylated, and Asn90 was confirmed to prevent the binding of ACE2 and SARS-CoV-2.58 Based on some experimental evidence and bioinformatics predictions,59 the number of N-glycosylation sites appears to be varied across different species as following (Table 1). Firstly, the number of glycosylation sites is different in the range from 1 in very lowly advanced vertebrates (zebrafish and sea lamprey) to 8 in...
| Gene     | Species     | Common name       | Accession number | Identities to human ACE2 (%) | Length (aa) | Position of peptidase M2 domain 1 | Position of peptidase M2 domain 2 | Position of collectrin domain | Signal peptide | Trans-membrane region | N-Glycosylation sites |
|----------|-------------|-------------------|------------------|-------------------------------|-------------|-----------------------------------|-----------------------------------|-------------------------------|-----------------|----------------------|----------------------|
| ACE2     | *H. sapiens* | human             | NP_001358344.1   | 100                           | 805         | 19-606                            | N                                 | 617-770                      | 1-17            | 741-763              | 53, 90, 103, 322, 432, 546, 690 |
| M. mulatta | Rhesus macaque | NP_001129168.1   | 94.96            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 90, 103, 322, 432, 546, 690 |
| M. fascicularis | Crab-eating macaque | XP_005593094.1 | 95.21            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 90, 103, 322, 432, 546, 690 |
| M. auratus | Golden hamster | XP_005074266.1   | 84.26            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 82, 90, 432, 658, 690, 728, 772 |
| M. musculus | Mouse        | NP_001123985.1   | 81.86            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 536, 546, 660, 690, 722 |
| C. porcellus | Guinea pig   | XP_023417808.1   | 77.27            | 813                           | 27-614      | 740-762                          | 625-778                          | 1-25                         | 748-770         | 61, 98, 224, 440, 698, 809 |
| O. cuniculus | Rabbit       | QHX39726.1       | 84.76            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 741-763         | 53, 90, 134, 432 |
| F. catus    | Cat          | NP_001034545.1   | 85.59            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 90, 216, 299, 322, 690 |
| P. concolor | Cougar       | XP_025790417.1   | 85.70            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 90, 216, 299, 322, 690 |
| P. tigris altaica | Tiger      | XP_007090142.1   | 85.70            | 797 (partial)                 | 11-598      | N                                 | 609-762                          | 1-13                         | 732-754         | 45, 82, 208, 291, 314, 682 |
| C. crocuta  | Spotted hyena | KAF0878287.1     | 83.35            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 216, 322, 690 |
| S. suricatta | Meerket      | XP_029786256.1   | 82.86            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 90, 216, 690 |
| P. larvata  | Masked palm civet | Q56NL1.1         | 83.48            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 216, 322, 432, 690 |
| U. arctos horribilis | Grizzly bear | XP_026333865.1   | 83.88            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 90, 216, 322, 432, 690 |
| U. maritimus | Polar bear   | XP_008694637.1   | 83.92            | 790 (partial)                 | 4-591       | N                                 | 602-755                          | 1-31                         | 725-747         | 38, 75, 201, 307, 417, 675 |
| A. melanoleuca | Panda       | EFB23904.1       | 83.38            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 53, 90, 216, 322, 432, 690 |
| P. lotor    | Raccoon     | BAE72462.1       | 83.88            | 805                           | 19-606      | N                                 | 617-770                          | 1-17                         | 740-762         | 34, 53, 216, 322, 432, 690 |
| Gene Species | Common name | Accession number | Identities to human ACE2 (%) | Length (aa) | Position of peptidase M2 domain 1 | Position of peptidase M2 domain 2 | Position of collectrin domain | Signal peptide | Trans-membrane region | N-Glycosylation sites |
|--------------|-------------|-----------------|-------------------------------|------------|----------------------------------|----------------------------------|-----------------------------|---------------|----------------------|----------------------|
| *M. putorius furo* | Ferret | NP_001297119.1 | 82.74 | 805 | 19–606 | N | 617–770 | 1–17 | 740–762 | 53, 322, 690 |
| *C. lupus dingo* | Dingo (Dog) | XP_025292925.1 | 84.01 | 804 | 18–605 | N | 616–769 | 1–17 | 739–761 | 53, 133, 215, 298, 321, 689 |
| *C. lupus familiaris* | Dog | NP_001158732.1 | 83.50 | 804 | 18–605 | N | 616–769 | 1–17 | 739–761 | 53, 133, 215, 298, 321, 689 |
| *N. procyonoides* | Raccoon dog | ABW16956.1 | 84.01 | 804 | 18–605 | N | 616–769 | 1–17 | 739–761 | 53, 133, 215, 298, 321, 689 |
| *V. vulpes* | Red fox | XP_025842512.1 | 83.63 | 804 | 18–605 | N | 616–769 | 1–17 | 739–761 | 53, 133, 215, 298, 321, 689 |
| *M. javanica* | Malayan pangolin | XP_017505752.1 | 84.76 | 805 | 19–606 | N | 617–770 | 1–17 | 740–762 | 53, 90, 216, 299, 432, 690 |
| *O. aries* | Sheep | XP_011961657.1 | 81.74 | 804 | 19–605 | N | 616–769 | 1–17 | 739–761 | 53, 90, 431, 545, 659, 689 |
| *C. hircus* | Goat | NP_001277036.1 | 81.74 | 804 | 19–605 | N | 616–769 | 1–17 | 739–761 | 53, 90, 431, 545, 659, 689 |
| *B. taurus* | Taurine cattle | XP_005228486.1 | 81.12 | 804 | 19–605 | N | 616–769 | 1–17 | 739–761 | 53, 90, 298, 431, 545, 659, 689 |
| *B. mutus* | Wild Yak | XP_005903173.1 | 81.37 | 804 | 19–605 | N | 616–769 | 1–17 | 739–761 | 53, 90, 298, 431, 545, 659, 689 |
| *B. indicus* | Humped cattle | XP_019811720.1 | 81.12 | 804 | 19–605 | N | 616–769 | 1–17 | 739–761 | 53, 90, 298, 431, 545, 659, 689 |
| *M. muntjak* | Barking deer | KAB0345583.1 | 81.61 | 804 | 19–605 | N | 616–769 | 1–17 | 739–761 | 53, 90, 298, 431, 545, 689 |
| *S. scrofa domesticus* | Pig | ACT66265.1 | 81.94 | 787 (partial) | 19–606 | N | 617–770 | 1–17 | 740–762 | 53, 299, 322, 690 |
| *S. scrofa* | Wide pig | XP_020935033.1 | 81.74 | 805 | 19–606 | N | 617–770 | 1–17 | 740–762 | 53, 299, 322, 660, 690 |
| *C. ferox* | Camel | XP_006194263.1 | 82.23 | 805 | 19–606 | N | 617–770 | 1–17 | 739–761 | 53, 90, 322, 660, 690 |
| *E. caballus* | Horse | XP_001490241.1 | 86.78 | 805 | 19–606 | N | 617–770 | 1–17 | 739–761 | 53, 90, 134, 257, 322, 660, 690 |

(Continues)
| Gene     | Species          | Common name                  | Accession number | Identities to human ACE2 (%) | Length (aa) | Position of peptidase M2 domain 1 | Position of peptidase M2 domain 2 | Position of collectrin domain | Signal peptide | Trans-membrane region | N-Glycosylation sites |
|----------|------------------|------------------------------|------------------|-----------------------------|-------------|-----------------------------------|-----------------------------------|-----------------------------|---------------|------------------------|---------------------|
| H. armiger | Great Roundleaf Bat | XP_019522954.1 | 80.38 | 806 | 19-606 | N | 618-771 | 1-17 | 741-763 | | 53, 90, 322, 432, 661 |
| R. ferrumequinum | greater horseshoe bat | XP_032963186.1 | 81.36 | 805 | 19-606 | N | 617-770 | 1-17 | 740-762 | | 38, 53, 82, 90, 322, 432, 546, 690 |
| T. chinensis | Chinese tree shrew | XP_006164754.1 | 80.75 | 805 | 19-606 | N | 617-770 | 1-17 | 740-762 | | 53, 78, 218, 257, 322, 690 |
| D. novemcinctus | Nine-banded armadillo | XP_004449124.1 | 79.13 | 804 | 18-605 | N | 616-769 | 1-19 | 739-761 | | 53, 81, 89, 217, 321, 545 |
| E. europaeus | Hedgehog | XP_007538670.1 | 79.01 | 804 | 19-606 | N | 617-769 | 1-17 | 739-761 | | 52, 81, 89, 217, 321, 545 |
| L. africana | Elephant | XP_023410960.1 | 80.50 | 800 | 19-601 | N | 612-765 | 1-17 | 735-757 | | 38, 53, 90, 103, 154, 216, 432 |
| T. manatus latirostris | Sea Cow | XP_004386381.1 | 81.49 | 800 | 19-601 | N | 612-765 | 1-17 | 735-757 | | 53, 154, 252, 298, 685 |
| O. afer afer | Aardvark | XP_007951028.1 | 79.38 | 799 | 19-600 | N | 611-764 | 1-17 | 734-756 | | 38, 53, 90, 153, 297, 488, 601 |
| E. edwardii | Cape elephant shrew | XP_006892457.1 | 77.52 | 798 | 19-601 | N | 612-763 | 1-17 | 734-756 | | 53, 154, 398, 317, 427, 683 |
| P. cinereus | Koala | XP_020863153.1 | 71.48 | 807 | 19-606 | N | 617-769 | 1-17 | 739-761 | | 53, 78, 216, 322, 660 |
| O. anatinus | Platypus | XP_001515597.2 | 68.02 | 806 | 18-605 | N | 621-773 | 1-17 | 743-765 | | 52, 78, 83, 321, 678, 790 |
| G. gallus | Chicken | XP_416822.2 | 65.97 | 808 | 18-606 | N | 617-770 | 1-17 | 741-763 | | 52, 78, 194, 257, 280, 322, 534, 690 |
| A. platyrhynchos | Duck | XP_012949915.2 | 64.81 | 805 | 18-606 | N | 617-767 | 1-17 | 738-760 | | 52, 78, 322, 534, 634, 687 |
| A. forsteri | Penguin | XP_009275140.1 | 66.75 | 809 | 18-606 | N | 617-770 | 1-17 | 741-763 | | 37, 52, 78, 257, 534, 572, 690 |
| P. bivittatus | Snake | XP_007431942.2 | 61.77 | 827 | 43-628 | N | 639-793 | 1-42 | 763-785 | | 77, 103 |
| Gene     | Species       | Common Name | Accession number | Identities to human ACE2 (%) | Identities to human ACE (%) | Length (aa) | Position of peptidase M2 domain 1 | Position of peptidase M2 domain 2 | Position of collectrin domain | Signal peptide | Trans-membrane region | N-Glycosylation sites |
|----------|---------------|-------------|------------------|-------------------------------|-------------------------------|-------------|-----------------------------------|-----------------------------------|-------------------------------|-----------------------------|------------------------|---------------------|
| X. tropicalis | Frog         | XP_002938293.2 | 60.03            | 862                           | 20–610                        | N           | 622–773                           | 1–17                             | 744–766                      | 54, 76                     |
| D. rerio  | Zebrafish    | NP_001007298.1 | 57.83            | 785                           | 18–607                        | N           | 630–783                           | 1–17                             | 752–774                      | 545                        |
| P. marinus | Sea lamprey  | XP_032835032.1 | 56.39            | 849                           | 19–579                        | N           | 590–742                           | 1–15                             | 711–733                      | 88                         |

| Gene     | Species       | Common Name | Accession number | Identities to human ACE2 (%) | Identities to human ACE (%) | Length (aa) | Position of peptidase M2 domain 1 | Position of peptidase M2 domain 2 | Position of collectrin domain | Signal peptide | Trans-membrane region | N-Glycosylation sites |
|----------|---------------|-------------|------------------|-------------------------------|-------------------------------|-------------|-----------------------------------|-----------------------------------|-------------------------------|-----------------------------|------------------------|---------------------|
| ACE      | H. sapiens    | human       | NP_000780.1      | 100                           | 1306                          | 40–623      | 643–1221                          | N                                | 1–29                         | 1257–1276                   |
| G. gallus | Chicken      | NP_001161204.1 | 68.03            | 1281                          | 28–609                        | 629–1207    | N                                | 1–17                             | 1240–1262                   | 1240–1262                   |
| A. carolinensis | Green Alone | XP_00811340.1 | 66.88            | 1305                          | 52–635                        | 655–1233    | N                                | 1–40                             | 1264–1286                   | 1264–1286                   |
| X. tropicalis | Frog       | NP_001166882.1 | 65.6             | 1284                          | 28–611                        | 631–1209    | N                                | 1–16                             | 1244–1266                   | 1244–1266                   |
| D. rerio  | Zebrafish    | XP_694336.5  | 65.18            | 1324                          | 69–651                        | 671–1249    | N                                | 1–58                             | 1283–1305                   | 1283–1305                   |
| M. musculus | Mouse       | NP_997507.1  | 83.08            | 1312                          | 45–628                        | 648–1226    | N                                | 1–29                             | 1264–1286                   | 1264–1286                   |
| P. marinus | Sea lamprey  | XP_032807619.1 | 62.47            | 1281                          | 32–615                        | 635–1212    | N                                | 1–20                             | 1248–1274                   | 1248–1274                   |
| P. marinus-like | Sea lamprey | XP_032821231.1 | 60.59            | 1290                          | 38–623                        | 643–1220    | N                                | 1–27                             | 1252–1274                   | 1252–1274                   |
| D. melanogaster | Fruit fly  | AAB02171.1   | 36.47            | 44.35                         | 615                            | 19–606      | N                                | N                                | 1–17                         | N                          |
| C. elegans | Roundworm    | NP_001024453.1 | 26.59            | 906                            | 175–764                       | N            | N                                | 1–19                             | N                            | N                          |
| H. vulgaris | Hydra vulgaris | XP_004208490.1 | 38.63            | 639                            | 20–607                        | N            | N                                | 1–19                             | N                            | N                          |
| H. vulgaris | Hydra vulgaris | XP_012555870.1 | 38.15            | 616                            | 21–605                        | N            | N                                | 1–23                             | N                            | N                          |

Note: "N" means not found.
mammals (chicken, bat, and golden hamster). Human, monkey, dog, cattle, horse, aardvark, elephant, and penguin have 7 sites, while other animals have 2–6. Secondly, the position of N-glycosylation sites is varied. Asn53 (50/51) is the highest conserved sites among different species and then subsequently is Asn690 (39/51), Asn322 (29/51), Asn90 (28/51), Asn432/18/51), Asn546 (10/51), and Asn103 (3/51), whereas Asn103 only appears in primates.

### 3.3 Sequence identification of TMPRSS2

A blast search identified TMPRSS2 homologous in representative organisms (Table 2), including 1 homolog in mammals, aves, reptiles, and teleost fishes, 10 homologous in frog *Xenopus tropicalis*, but not in lower vertebrate sea lamprey *Petromyzon marinus* (cyclostomata) and invertebrates, such as fruit fly *D. melanogaster* (arthropoda) and roundworm *C. elegans* (nematomorpha). Interestingly, possibly due to the incompleteness of the genome sequencing or lineage-specific gene loss/evolution in vertebrates, masked palm civet *Paguma larvata*, raccoon dog *Nyctereutes procyonoides*, pig *Sus scrofa domesticus*, and cape elephant shrew *Elephantulus edwardii* lacked Tmprss2 genes.

Compared to the homology of human ACE2 with other species, the homology between human TMPRSS2 and other species is much lower. For example, in rhesus macaque, *Macaca mulatta*, an experimental model widely used in SARS-CoV-2 research, the homology of ACE2 with humans is 94.96%, while the homology of TMPRSS2 with humans is 88.06%. Vertebrate TMPRSS2 is composed of three domains: low-density LDLa for calcium binding, SR for the binding to other cell surface or extracellular molecules, and TrypSPc for cleaving peptide bonds after lysine or arginine residues. Interestingly, LDLa domain is not represented in ferret *Mustela putorius furo*, hedgehog *Erinaceus europaeus* and sea cow *Trichechus manatus*: X. tropicalis tmprss 2.2, 2.4, 2.5, 2.6, 2.7, 2.14, and 2.15 contains many \( \geq 2 \) LDLa. SR domain is not identified in cougar *Puma concolor* and frog *tmprss* 2.4. Conversely, TrypSPc domain was conserved in all representative species except frog *tmprss* 2.4. To date, all characterized vertebrate TMPRSS2 has a signal peptide and transmembrane region. However, hedgehog lacked signal peptide, while ferret and great roundleaf bat *Hipposideros armiger* and frog *tmprss* 2.5 lacked trans-membrane region.

### 3.4 Phylogenetic analysis of ACE2 and TMPRSS2 gene family

To examine the evolutionary relationship of ACE2 gene family, a NJ tree with bootstrapping branch confidence was constructed with MEGA X using ACE2 homologs from vertebrates, including mammals, birds, reptiles, amphibians, fishes, and cyclostomata (Figure 1). Entirely, the tree is consistent with the topology law of species evolution. The fishes (D. rerio) and cyclostomata (P. marinus) locate at the base, suggesting they are likely the ancestors of the ACE2 gene family. In addition to ACE2 from platypus *Ornithorhynchus anatinus* (protopheria) and koala *Plethodon cinereus* (Metatheria) located outside the branches of vertebrate ACE2, ACE2 homologous from eutheria clusters together. In most of Eutheria, except for cape elephant shrew *E. edwardii* (macroscelidea), aardvark *Orycteropus afer afer* (tubulidentata), sea Cow *T. manatus latirostris* (sirenia), elephant *Loxodonta africana* (proboscidia), hedgehog *E. europaeus* (insectivora), nine-banded armadillo *Dasypus novemcinctus* (edentata), and tupaia *Tupaia chinensis* (scandentia), ACE2 splits into two main branches: one branch contains species from primates, rodentia and lagomorpha; the other branch contains species from carnivora, pholidota, artiodactyla, perissodactyla, and chiroptera.

Moreover, an NJ tree of TMPRSS2 from different species was created to detect its evolutionary relationship. Entirely, the TMPRSS2 tree is similar to ACE2 tree, consistently with the species phylogeny (Figure 1). In addition to frog *tmprss* 2.1, other 9 frog *tmprss* 2 genes situ in the root of the tree, indicating a lineage-specific duplication or gene conversion to produce massive homologs occurs in frogs. Compared to the ACE2 tree, hedgehog *E. europaeus* (insectivora) in TMPRSS2 tree locates at the base of vertebrates, while tupaia *T. chinensis* (scandentia) clusters closely to primates. Given the reported cases of animals or cell lines (Table S1) that appear in both branches 1 and 2, it is therefore possible that animals from those two branches is susceptible to SARS-CoV-2 infections.

### 3.5 Comparison of interacting amino acids of ACE2 and SARS-CoV-2

Cocrystallization and the structural determination of SARS-CoV-2 S protein with full-length human ACE2 identified a total of 25 amino acids in ACE2 was critical for its binding to SARS-CoV-2. To confirm the conservation of these 25 residues across different species, the amino acid sequence alignments by Clustal X were performed. As shown in Table 3, Gly352, Gly355, and Gly357 were conserved among all vertebrate ACE2; Phe28 (50/51), Arg393 (50/51), Leu45 (47/51), and Asn330 (48/51) is highly conserved across all vertebrates; Met82 only presents in primates. To identify which residues are highly associated with the recognition of ACE2 to SARS-CoV-2, we performed Fisher exact test by SAS 9.4. Based on the receptor recognition summarized in Table S1, we selected nine species as the positive samples and four lower advanced species including snake, frog, fish, and sea lamprey, as the negative samples. We observed the Phe28, Asp355, Arg357, and Arg93 are present in both negative and positive animals, indicating that they might be not relevant to the recognition of ACE2 to SARS-CoV-2 among different species (Table S2). Nevertheless, the \( \rho \)-value of His34, Glu35, Asp38, Tyr41, Gin42, Leu45, Asn330, and Gly354 is above .05, whereas other residues are under .05 (Figure S2). In particular, the \( \rho \)-value of Ser19, Lys26, Thr27, Asp30, Leu79, and Met82 is lower than .01 (Figure 2A), implying those amino acids are highly associated with the binding of ACE2 to SARS-CoV-2.
| Species          | Common Name                  | Accession number       | Identities to human (%) | Length (aa) | Position of LDLa domain (low-density lipoprotein receptor domain class A) | Position of SR domain (Scavenger receptor Cys-rich) | Position of Tryp_SPc domain (Trypsin-like serine protease) | Signal peptide | Trans-membrane region |
|------------------|------------------------------|------------------------|-------------------------|-------------|-------------------------------------------------------------------------|-----------------------------------------------|-------------------------------------------------------------|----------------|----------------------|
| H. sapiens       | Human                        | NP_001128571.1         | 100                     | 529         | 149–187| 186–279| 292–521| 1–49| 121–143 | |
| M. mulatta       | Rhesus macaque              | XP_028701148.1         | 88.06                   | 534         | 154–192| 191–284| 297–526| 1–29| 126–148 | |
| M. fascicularis  | Crab-eating macaque         | XP_005548700.1         | 86.6                    | 525         | 145–183| 182–275| 288–517| 1–46| 117–139 | |
| M. auratus       | Golden hamster              | XP_012971684.1         | 76.67                   | 491         | 112–150| 119–242| 254–483| 1–13| 85–107 | |
| M. musculus      | Mouse                        | NP_056590.2            | 78.41                   | 490         | 111–149| 148–241| 253–482| 1–13| 86–108 | |
| C. porcellus     | Guinea pig                  | XP_013001524.1         | 86.6                    | 523         | 109–147| 146–239| 251–480| 1–38| 83–105 | |
| O. cuniculus     | Rabbit                      | XP_008250697.1         | 76.67                   | 275         | 135–171| 170–262| 274–503| 1–36| 107–129 | |
| F. cat          | Cat                         | XP_023094477.1         | 79.67                   | 492         | 112–150| 149–242| 255–484| 1–13| 84–106 | |
| P. concolor     | Cougar                      | XP_025768873.1         | 58.42                   | 446         | 112–150| N      | 212–402| 1–13| 84–106 | |
| P. tigris altaica | Tiger                      | XP_015396688.1         | 69.11                   | 518         | 112–150| 149–242| 224–510| 1–13| 84–106 | |
| C. crocuta       | Spotted hyena               | KAF0886132.1           | 79.38                   | 492         | 112–150| 149–242| 255–484| 1–50| 84–106 | |
| S. suricatta    | Meerket                     | XP_029794796.1         | 77.8                    | 492         | 112–150| 149–242| 255–484| 1–50| 84–106 | |
| P. larvata       | Masked palm civet           | N                      |                         |             |             |             |             |             |             |             | |
| U. arctos horribilis | Grizzly bear               | XP_026355755.1         | 79.43                   | 492         | 112–150| 149–242| 255–484| 1–41| 84–106 | |
| U. maritimus     | Polar bear                  | XP_008708425.1         | 78.86                   | 524         | 112–150| 149–242| 255–487| 1–41| 84–106 | |
| A. melanoleuca  | Panda                       | XP_034511296.1         | 78.41                   | 492         | 112–150| 149–242| 255–484| 1–41| 84–106 | |
| P. lotor        | Raccoon                     | N                      |                         |             |             |             |             |             |             |             | |
| M. putorius furo | Ferret                     | XP_012916721.1         | 72.89                   | 551         | N      | 236–301| 314–543| 1–46| N    | |
| C. lupus dingo  | Dingo (Dog)                 | XP_025317827.1         | 80.24                   | 492         | 112–150| 149–242| 255–484| 1–13| 84–106 | |
| C. lupus familiaris | Dog                       | BBD33861.1             | 80.04                   | 492         | 112–150| 149–242| 255–484| 1–13| 84–106 | |
| N. procyonoides | Raccoon dog                 | N                      |                         |             |             |             |             |             |             |             | |
| V. vulpes       | Red fox                     | XP_025839165.1         | 79.84                   | 492         | 112–150| 149–242| 255–484| 1–13| 84–106 | |
| M. javanica     | Malayan pangolin            | XP_017508124.1         | 78.16                   | 527         | 112–150| 149–242| 255–484| 1–50| 84–106 | |
| O. aries        | Sheep                       | XP_027816505.1         | 79.02                   | 490         | 112–148| 147–240| 253–482| 1–13| 84–106 | |

(Continues)
| Species          | Common Name          | Accession number | Identities to human (%) | Length (aa) | Position of LDLa domain (low-density lipoprotein receptor domain class A) | Position of SR domain (Scavenger receptor Cys-rich) | Position of Tryp.SPc domain (Trypsin-like serine protease) | Signal peptide | Trans-membrane region |
|------------------|----------------------|------------------|--------------------------|------------|--------------------------------------------------------------------------------|-------------------------------------------------------------------|----------------------------------------------------------------|----------------|------------------------|
| C. hircus        | Goat                 | XP_005675686.1   | 77.39                    | 490        | 112–148                                                                         | 112–148                                                           | 112–148                                                       | 1–13          | 84–106                 |
| B. taurus        | Taurine cattle       | NP_001075054.1   | 78.41                    | 490        | 112–148                                                                         | 112–148                                                           | 112–148                                                       | 113           | 84–106                 |
| B. mutus         | Wild Yak             | XP_005893081.1   | 78.21                    | 490        | 112–148                                                                         | 112–148                                                           | 112–148                                                       | 1–13          | 84–106                 |
| B. indicus       | Humped cattle        | XP_019818444.1   | 78.41                    | 490        | 112–148                                                                         | 112–148                                                           | 112–148                                                       | 1–13          | 84–106                 |
| M. muntjak       | Barking deer         | KAB0356950.1     | 77.50                    | 503        | 110–146                                                                         | 110–146                                                           | 110–146                                                       | 1–45          | 82–104                 |
| S. scrofa        | Pig                  |                  |                         |            |                                                                                   |                                                                   |                                                              |               |                       |
| S. scrofa        | Wide pig             | BAF76737.1       | 76.92                    | 495        | 114–152                                                                         | 114–152                                                           | 114–152                                                       | 1–13          | 86–108                 |
| C. fenus         | Camel                | XP_032317729.1   | 77.6                     | 492        | 112–150                                                                         | 112–150                                                           | 112–150                                                       | 1–13          | 84–106                 |
| E. caballus      | Horse                | XP_005606217.1   | 74.75                    | 490        | 112–148                                                                         | 112–148                                                           | 112–148                                                       | 1–13          | 84–106                 |
| H. armiger       | Great Roundleaf Bat  | XP_019481585.1   | 84.33                    | 384        | 4–42                                                                             | 4–42                                                              | 4–42                                                         | 1–27          | N                      |
| R. ferrumequinum | greater horseshoe bat| XP_032944694.1   | 80.86                    | 492        | 112–150                                                                         | 112–150                                                           | 112–150                                                       | 1–13          | 84–106                 |
| T. chinensis     | Chinese tree shrew    | XP_027630699.1   | 70.92                    | 500        | 112–148                                                                         | 112–148                                                           | 112–148                                                       | 1–13          | 84–106                 |
| D. novemcinctus  | Nine-banded armadillo| XP_004466361.1   | 79.32                    | 460        | 80–118                                                                           | 80–118                                                            | 80–118                                                       | 1–44          | 52–74                  |
| E. europaeus     | Hedgehog             | XP_016049831.1   | 68.91                    | 479        | N                                                                                 |                                                                   | N                                                            | N             | 17–39                  |
| L. africana      | Elephant             | XP_023414879.1   | 74.49                    | 497        | 111–149                                                                         | 111–149                                                           | 111–149                                                       | 1–13          | 84–106                 |
| T. manatus       | Sea Cow              | XP_023594890.1   | 69.82                    | 491        | N                                                                                 |                                                                   | N                                                            | 1–38          | 81–103                 |
| O. afer afer     | Aardvark             | XP_007945507.1   | 78.94                    | 524        | 112–150                                                                         | 112–150                                                           | 112–150                                                       | 1–13          | 84–106                 |
| E. edwardii      | Cape elephant shrew  |                  |                         |            |                                                                                   |                                                                   |                                                              |               |                       |
| P. cinereus      | Koala                | XP_020831920.1   | 69.98                    | 492        | 111–149                                                                         | 111–149                                                           | 111–149                                                       | 1–48          | 83–105                 |
| Species       | Common Name | Accession number | Identities to human (%) | Length (aa) | Position of LDLa domain (low-density lipoprotein receptor domain class A) | Position of SR domain (Scavenger receptor Cys-rich) | Position of Tryp_SPc domain (Trypsin-like serine protease) | Signal peptide | Trans-membrane region |
|---------------|-------------|------------------|-------------------------|-------------|--------------------------------------------------------------------------|--------------------------------------------------|----------------------------------------------------------|----------------|----------------------|
| O. anatinus   | Platypus    | XP_028902467.1   | 71.49                   | 492         | 111–149                                                                  | 148–242                                          | 255–484                                                 | 1–57           | 83–105               |
| G. gallus     | Chicken     | XP_015156666.1   | 59.34                   | 486         | 104–142                                                                  | 141–235                                          | 249–478                                                 | 1–53           | 77–99                |
| A. platyrhynchos | Duck        | XP_021132301.2   | 56.22                   | 482         | 101–138                                                                  | 137–231                                          | 245–474                                                 | 1–61           | 75–97                |
| A. forsteri   | Penguin     | XP_009281370.1   | 58.51                   | 484         | 102–140                                                                  | 139–233                                          | 247–476                                                 | 1–49           | 76–98                |
| P. bivittatus | Snake       | XP_007441832.1   | 57.26                   | 487         | 105–143                                                                  | 142–236                                          | 250–479                                                 | 1–54           | 78–100               |
| X. tropicalis | Frog        | XP_004912265.1   | 48.47                   | 482         | 110–145                                                                  | 148–241                                          | 247–474                                                 | 1–59           | 82–104               |
|               |             | XP_004912262.1   | 51.73                   | 521         | 101–136, 141–176                                                          | 175–269                                          | 281–513                                                 | 1–60           | 67–89                |
|               |             | XP_031752391.1   | 28.57                   | 571         | 142–177, 203–241, 267–305, 338–376, 398–436, 441–476, 483–518, 525–560 | N                                               | N                                                        | 1–34           | 106–128              |
|               |             | XP_031752394.1   | 50.67                   | 497         | 2–33, 38–73, 78–113, 120–155                                             | 154–248                                          | 260–489                                                 | 1–18           | N                    |
|               |             | XP_002936535.2   | 51.21                   | 695         | 107–142, 149–184, 193–228, 235–270, 277–312, 319–354                     | 353–447                                          | 459–687                                                 | 1–60           | 72–94                |
|               |             | XP_031752397.1   | 50                      | 511         | 94–129, 130–166                                                          | 165–259                                          | 271–503                                                 | 1–41           | 67–89                |
|               |             | XP_031752400.1   | 45.72                   | 497         | 122–157                                                                  | 156–249                                          | 260–489                                                 | 1–37           | 96–108               |
|               |             | XP_002943001.1   | 45.52                   | 460         | 64–99                                                                    | 121–216                                          | 221–452                                                 | 1–47           | 35–57                |
|               |             | XP_031752402.1   | 50.13                   | 504         | 88–123, 126–161                                                          | 160–253                                          | 264–495                                                 | 1–23           | 55–77                |
|               |             | XP_031752206.1   | 43.65                   | 504         | 88–123, 126–161                                                          | 160–253                                          | 264–495                                                 | 1–23           | 55–77                |
| D. rerio      | Zebrafish   | NP_001008623.1   | 44.4                    | 486         | 133–170                                                                  | 173–224                                          | 251–479                                                 | 1–39           | 107–129              |
| P. marinus    | Sea lamprey | N                 |                         |             |                                                                         |                                                  |                                                         |                |                      |
| D. melanogaster | Fruit fly  | N                 |                         |             |                                                                         |                                                  |                                                         |                |                      |
| C. elegans    | Roundworm   | N                 |                         |             |                                                                         |                                                  |                                                         |                |                      |

Note: “N” means not found.

Abbreviations: LDLa, lipoprotein receptor domain class A; SR, scavenger receptor cysteine-rich domain; Tryp_SPc, trypsin-like serine protease.
To further dissect the probability of the recognition of ACE2 to SARS-CoV-2 in other unknown species, a random forest analysis using the randomForest function from the R randomForest library was performed. As Leu79 has too many species with unknown status of amino acids, only Ser19, Lys26, Thr27, Asp30, and Met82 were used for this analysis. The identified receptor recognition shown in "Yes" in Figure 2B is mainly distributed in two classes, carnivora and artiodactyla, and contains horse, malayan pangolin, and sea cow. These identified species comprise wide animals and many important economy livestock, such as sheep and cows.

3.6 Structure analysis of ACE2 gene family

To further explore the recognition of ACE2 to SARS-CoV-2 across different species, we selected ACE2 genes from 7 animals according to the reported cases in Table S1 (rabbit and pig), analysis on Figure 2 (the "Yes" for horse, cattle, and malayan pangolin; "NA" for platypus), and snake for negative binding and carried out their structure analysis. It is known that the networks of hydrophobic interactions and hydrogen-bonding within the interface between ACE2 and the receptor-binding domain (RBD) of SARS-CoV-2 could enhance this receptor binding.61 The hydrophobic contact of Phe486 of RBD with ACE2 is situated in a pocket fenced by Leu79, Met82, and Tyr83 at the N-terminal end of ACE2. Indeed, Tyr83 contributes a hydrogen bond to Asn487 of the RBD.61 Therefore, we first analyzed the hydrophobicity of the amino acids of 76 to 85 among various species by DNAMAN. As per Figure 3A, compared with the hydrophobicity value of human ACE2 above zero, the hydrophobicity value within the other seven animals is below zero, indicating the key part of ACE2 from the other seven species is not hydrophobic. Given the enhancement of hydrophobic contact on the binding of ACE2 and TMPRSS2 genes from the highly advanced animals could be split into two branches (Branches 1 and 2). *Labeled the animals or cells which has been confirmed to be recognized by SARS-CoV-2. *Labeled the species whose position is quite different between ACE2 and TMPRSS2 trees. ACE2, angiotensin-converting enzyme 2; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

FIGURE 1 Neighbor-joining phylogenetic tree of ACE2 (left) and TMPRSS2 (right) gene family. The confidence values of bootstrap derived from 1000 replications, and the percentage of bootstrap is shown on interior branches. The scale bar shows the number of substitutions per site. (A) ACE2 tree. (B) TMPRSS2 tree. X. tropicalis tmprss2 gene is shown as gene symbols from Genbank to species names distinguish different tmprss2 homologous. The gene accession numbers used in this figure refer to Tables 1 and 2. Entirely, both ACE2 and TMPRSS2 gene from different species clusters as species taxonomy, which is labeled next to the tree. ACE2 and TMPRSS2 genes from the highly advanced animals could be split into two branches (Branches 1 and 2). *Labeled the animals or cells which has been confirmed to be recognized by SARS-CoV-2. *Labeled the species whose position is quite different between ACE2 and TMPRSS2 trees. ACE2, angiotensin-converting enzyme 2; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.
RBD, ACE2 from horse, cow, and Malayan pangolin to RBD probably has less affinity of RBD to ACE2. Secondly, GOR IV prediction displays a similar composition of the secondary structure of ACE2 from different species, including alpha helix (average value is 41.7%), extended strand (average value is 17.53%), and random coil (average value is 40.78%), whereas the lower advanced species snake has a more extended strand than other higher species (Figure 3B). Furthermore, the 3D structure of the full-length ACE2 from the other 7 animals was predicted by I-TASSER program. C-score and TM-score of the first-ranked model from 7 animals are from \((-2, 0)\) and 0.49 ±0.15, respectively (data not shown). As scores range from \((-5, 2)\) and a large value means more confidence in the model, our model, therefore, is a good model (Figure S3B–H). For TM-score, a value > 0.5 means the model has correct topology. To gain an understanding of the structural basis for the binding of RBD with ACE2, we focused on the key binding contact region with those 25 amino acids. The structural alignments by TM-align serve showed that the TM-score between human ACE2 (1R42A) and horse, cow, and Malayan pangolin is 0.99889, 0.99699, and 0.99867 respectively, while the human and the snake has 0.87201. Consistently, the 3D structure of the binding site region of human ACE-SARS-CoV-2 is almost completely overlapped with horse, cattle or Malayan pangolin ACE2 (Figure S3I–K), while partially overlapped with snake ACE2 (Figure S3L).

### DISCUSSION

To our acknowledge, according to the limited information available to date, several species including pet dogs, cats, lions, tigers, and farmed mink can be naturally infected with SARS-CoV-2 from humans and cause COVID-19. There is no evidence that the infected pet dogs and cats could spread SARS-CoV-2 to humans, thereby the risk of animals transmitting SARS-CoV-2 to humans is considered to be very low. However, the genetic code of SARS-CoV-2 detected in three infected people demonstrate great resemblance to the viral genetic code previously found in the mink at the infected farm in the Netherlands, suggesting SARS-CoV-2 is probably able to transmit
FIGURE 2

(See caption on next page)
from mink to human (the COVID-19 update 209 on International Society for infectious Disease). Particularly, some virus could end up passing back and forth between animals and people. For example, the 2009 pandemic H1N1 influenza virus originated in pigs, transmitted to humans, spread worldwide and then jumped back to pigs.66,67 Notably, given the livestock and pets are closely connected to human beings and very few related investigations have been explored, more studies are needed to be conducted to understand if different animals could be affected by SARS-CoV-2, so as to avoid a potential outbreak by animal transmission in the further, which would become a critical aspect for controlling pneumonia.

4.1 | ACE2 and TMPRSS2 in SARS-CoV-2 cell entry

Coronavirus entry into host cells is an important determinant of viral pathogenesis. ACE2 and TMPRSS2 are known to function as a cell surface receptor for viral attachment and cell surface protease for viral-cell fusion, respectively.36,37,68 To gain insight to viral cell entry, we first determined ACE2 and TMPRSS2 homologs from invertebrates to vertebrates and found both of two genes occur in vertebrates. Interestingly, compared to the stable existence and domain composition of ACE2 in each order of vertebrate, TMPRSS2
vary considerably across different species as following. First, TMPRSS2 is absent in many animals, even in some mammals; Secondly, the identities of TMPRSS2 between human and animals is much lower than the identities of ACE2 between human and animals; Thirdly, TMPRSS2 from ferret and great roundleaf bat lack the trans-membrane region, indicating that TMPRSS2 on those two animals is not able to function as protease. Given the ferret and fruit bats is susceptible to SARS-CoV-2,27 TMPRSS2 might be not necessary for viral cell entry in these animals, or some other redundant genes could replace the TMPRSS2 role. Coincidentally, a single-cell RNA-seq data analysis of ACE2 and TMPRSS2 from multiple tissues from healthy donors demonstrated that although no co-expression of ACE2 and TMPRSS2 in placenta/decidua and fetal liver and thymus, co-expression is observed in multiple tissue including airways, cornea, esophagus, ileum, colon, liver, gallbladder, heart, kidney and testis, in the eight subsets of ACE2+ airway epithelial cells, including nasal, goblet, basal, and Type II alveolar, TMPRSS2 was only expressed in one subset of ACE2+ cells, while cathepsin B, another protease is expressed in more than 70%–90% of ACE2+ cells, indicating that cathepsin or other proteases probably functionally replaces TMPRSS2 during the cell entry of SARS-CoV-2, and along with that ACE2, rather than TMPRSS2, may be a limiting factor for viral entry.69 Therefore, in this study, we focused on ACE2 on the following analysis, including sequence comparison, prediction of probability for infectious, and homology modeling and prediction.

4.2 | The N-glycosylation and hydrophobicity on the recognition of ACE2 to SARS-CoV-2

Receptor recognition is mainly determined by two factors, the binding specificity and affinity. A recent study identified the glyco-sylation at Asp90 could partly disrupt the interaction of the SARS-CoV-2 with ACE2 receptors,58 whereas Wang et al.61 discovered the hydrophobic contacts at Leu79, Met82, and Tyr83 of ACE2 can enhance receptor recognition. Our predictions of the number and position of N-glycosylation sites appeared to vary across different species (Table 1). However, few studies about which N-glycosylation can play a role in receptor binding have been reported. It is therefore hard to calculate the correlation of glycosylation sites across species and their susceptibility to viral cell entry. To some extent, the various N-glycosylations might result in varied receptor affinity. Moreover, our prediction about hydrophobicity implied less receptor affinity in all animals than humans, consistent with the current phenomenon that the virus is now spreading from person to person.

4.3 | The prediction of SARS-CoV-2 transmissibility

The phylogenetic tree of ACE2 among different organisms imply that the ACE2 is highly conserved in mammals. In many different classes of mammals, there are organisms with confirmed receptor binding, indicating that other species from the same class are possibly susceptible to SARS-CoV-2 infections. Nevertheless, our analysis about the probability of viral attachment across different species in Figure 2B demonstrated the probability varies in the same class. Moreover, a recent study in the human community described the viral entry/receptor binding is not the only factor that determines viral infection in COVID-19 cases, while host age, underlying conditions, behavior, and population density can also determine infection.70 It will be the same in the animals’ world.

4.4 | SARS-CoV-2 transmissibility in pig

Chu et al.,25 examined the greater increase of SARS-CoV-2 replicates in pig and rabbit cell lines, revealing that ACE2 can recognize SARS-CoV-2. Therefore we selected pig and rabbit ACE2 as the positive case for its binding to SARS-CoV-2 in the following analysis, including Figure 1–3 and Table 3. However, pig orally/intranasally/in-trastracheally inoculation inoculated with SARS-CoV-2 was negative for viral RNA in all organ samples and contact animals during all checkpoints within 21 days infections, declaring that the pig is not susceptible to SARS-CoV-2 in vivo.27,57 The sharp contrast of studies between cell lines and in vivo raises a concern, the viral receptor binding predictions is not sufficient to determine the susceptibility to SARS-CoV-2 infection. In future, more experiments on viral transmission should be performed to study the susceptibility of animals to SARS-CoV-2.

ACKNOWLEDGMENTS
This study received support from Qilu Young Scholarship (Shandong University).

CONFLICT OF INTERESTS
The authors declare that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Chen Huang http://orcid.org/0000-0001-8631-3654
Jie Yan http://orcid.org/0000-0001-8362-4086

REFERENCES
1. Tay MZ, Poh CM, Renia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. Nat Rev Immunol. 2020;20(6):363-374.
2. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020;395(10224):565-574.
3. Li J, Huang DQ, Zou B, et al. Epidemiology of COVID-19: a systematic review and meta-analysis of clinical characteristics, risk factors, and outcomes. J Med Virol. 2021;93(3):1449-1458.
4. Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining ‘host jump’ of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiol. 2015;23(8):468-478.
5. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17(3):181-192.
6. Hu B, Ge X, Wang LF, Shi Z. Bat origin of human coronaviruses. Virol J. 2015;12:221.
7. Wang Q, Zhang Y, Wu L, et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell. 2020;181(4):894-904e899.
8. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-273.
9. Lam TT, Jia N, Zhang YW, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature. 2020;583(7815):282-285.
10. Xiao K, Zhai J, Feng Y, et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature. 2020;583(7815):286-289.
11. Liu P, Jiang JZ, Wan XF, et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLOS Pathog. 2020;16(5):e1008421.
12. Chandrashekar A, Liu J, Martinot AJ, et al. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. Science. 2020;369(6505):812-817.
13. Munster VJ, Feldmann F, Williamson BN, et al. Respiratory disease in rhesus macaques inoculated with SARS-CoV-2. Nature. 2020;585(7824):268-272.
14. Li Y, Bi Y, Xiao H, et al. A novel DNA and protein combination COVID-19 vaccine formulation provides full protection against SARS-CoV-2 in rhesus macaques. Emerg Microbes Infect. 2021;10(1):342-355.
15. Yu J, Tostanoski LH, Peter L, et al. DNA vaccine protection against SARS-CoV-2 in rhesus macaques. Science. 2020;369(6505):806-811.
16. Shan C, Yao YF, Yang XL, et al. Infection with novel coronavirus (SARS-CoV-2) causes pneumonia in rhesus macaques. Cell Res. 2020;30(8):670-677.
17. Sia SF, Yan LM, Chin AWH, et al. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. Nature. 2020;583(7818):834-838.
18. Imai M, Iwatsu-Horimoto K, Hatta M, et al. Syrian hamsters as a small animal model for SARS-CoV-2 infection and countermeasure development. Proc Natl Acad Sci USA. 2020;117(28):16587-16595.
19. Halfmann PJ, Hatta M, Chiba S, et al. Transmission of SARS-CoV-2 in domestic cats. N Engl J Med. 2020;383(6):592-594.
20. Braun KM, Moreno GK, Halfmann PJ, et al. Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck. PLOS Pathog. 2021;17(2):e1009373.
21. Gaudreault NN, Trujillo JD, Carrossino M, et al. SARS-CoV-2 infection, disease and transmission in domestic cats. Emerg Microbes Infect. 2020;9(1):2322-2322.
22. Kim YL, Kim SG, Kim SM, et al. Infection and Rapid Transmission of SARS-CoV-2 in Ferrets. Cell Host Microbe. 2020;27(5):704-709 e702.
23. Richard M, Kok A, de Meulder D, et al. SARS-CoV-2 is transmitted via contact and via the air between ferrets. Nat Commun. 2020;11(1):3496.
24. Myktyzen AZ, Lamers MM, Okba NMA, et al. Susceptibility of rabbits to SARS-CoV-2. Emerg Microbes Infect. 2021;10(1):1-7.
25. Chu H, Chan JF, Yuen TT, et al. Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. Lancet Microbe. 2020;1(1):e14-e23.
26. Shi J, Wen Z, Zhong G, et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. Science. 2020;368(6494):1016-1020.
27. Schlootka K, Rissmann M, Graaf A, et al. SARS-CoV-2 in fruit bats, ferrets, pigs and chickens: an experimental transmission study. Lancet Microbe. 2020;1(5):e218-e225.
28. Bosco-Lauth AM, Hartweg AE, Porter SM, et al. Experimental infection of domestic dogs and cats with SARS-CoV-2: pathogenesis, transmission, and response to reexposure in cats. Proc Natl Acad Sci USA. 2020;117(42):26382-26388.
29. Segalès J, Puig M, Rodon J, et al. Detection of SARS-CoV-2 in a cat owned by a COVID-19-affected patient in Spain. Proc Natl Acad Sci USA. 2020;117(40):24790-24793.
30. McAuleese D, Laverack M, Wang L, et al. From people to panthera: natural SARS-CoV-2 infection in tigers and lions at the Bronx zoo. mBio. 2020;11(5).
31. Koopmans M. SARS-CoV-2 and the human-animal interface: outbreaks on mink farms. Lancet Infect Dis. 2021;21(1):18-19.
32. Oreshkova N, Molenaar RJ, Vreman S, et al. SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. Euro Surveill. 2020;25(23).
33. Fenollar F, Mediannikov O, Maurin M, et al. Mink, SARS-CoV-2, and the human-animal interface. Front Microbiol. 2021;12:663815.
34. Li F. Structure, function, and evolution of coronavirus spike proteins. Annu Rev Virology. 2016;3(1):237-261.
35. Hofmann H, Pohllmann S. Cellular entry of the SARS coronavirus. Trends Microbiol. 2004;12(10):466-472.
36. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367(6483):1260-1263.
37. Bao L, Deng W, Huang B, et al. The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. Nature. 2020;583(7818):830-833.
38. Dong M, Zhang J, Ma X, et al. ACE2, TMPRSS2 distribution and extrapulmonary organ injury in patients with COVID-19. Biomed Pharmacother. 2020;131:110678.
39. Claussen TM, Sandoval DR, Spiliid CB, et al. SARS-CoV-2 infection depends on cellular Heparan sulfate and ACE2. Cell. 2020;183(4):1043-1057 e1015.
40. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020;181(2):271-280 e278.
41. Lukassen S, Chua RL, Trefzer T, et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. EMBO J. 2020;39(10):e105114.
42. Donoghue M, Hisieh F, Barones E, et al. A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. Circ Res. 2000;87(5):E1-E9.
43. Tipnis SR, Hooper NM, Hyde R, Karran E, Christie G, Turner AJ. A human homolog of angiotensin-converting enzyme. Cloning and functional expression as a captopril-insensitive carboxypeptidase. J Biol Chem. 2000;275(43):33238-33243.
44. Hussain M, Jabeen N, Raza F, et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. J Med Virol. 2020;92(9):1580-1586.
45. Liu M, Wang T, Zhou Y, Zhao Y, Zhang Y, Li J. Potential role of ACE2 in coronavirus disease 2019 (COVID-19) prevention and management. J Transl Med. 2020;8(1):9-19.
46. Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res. 1999;59(17):4180-4184.
47. Stopsack KH, Mucci LA, Antonarakis ES, Nelson PS, Kantoff PW. TMPRSS2 and COVID-19: serendipity or opportunity for intervention? Cancer Discov. 2020;10(6):779-782.
48. Roy A, Kucukural A, Zhang Y, I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010;5(4):725-738.
49. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005;33(7):2302-2309.
51. Ou X, Liu Y, Lei X, et al. Author correction: characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun. 2021;12(1):2144.

52. Kim JM, Chung YS, Jo HJ, et al. Identification of coronavirus isolated from a patient in Korea with COVID-19. Osong Public Health Res Perspect. 2020;11(1):3-7.

53. Blanco-Melo D, Nilsson-Payant BE, Liu WC, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell. 2020;181(5):1036-1045 e1039.

54. Chen F, Zhang Y, Sucgang R, et al. Meta-analysis of host transcriptional responses to SARS-CoV-2 infection reveals their manifestation in human tumors. Sci Rep. 2021;11(1):2459.

55. Maggi F, Rosellini A, Spezia PG, et al. Nicotine upregulates ACE2 expression and increases competence for SARS-CoV-2 in human pneumocytes. ERJ Open Res. 2021;7(2).

56. Xu J, Xu X, Jiang L, Dua K, Hansbro PM, Liu G. SARS-CoV-2 induces transcriptional signatures in human lung epithelial cells that promote lung fibrosis. Respir Res. 2020;21(1):182.

57. Meekins DA, Morozov I, Trujillo JD, et al. Susceptibility of swine cells and domestic pigs to SARS-CoV-2. Emerg Microbes Infect. 2020;9(1):2278-2288.

58. Suryamohan K, Diwanji D, Stawiski EW, et al. Human ACE2 receptor polymorphisms and altered susceptibility to SARS-CoV-2. Commun Biol. 2021;4(1):475.

59. Shajahan A, Archer-Hartmann S, Supekar NT, Gleinich AS, Heiss C, Azadi P. Comprehensive characterization of N- and O-glycosylation of SARS-CoV-2 human receptor angiotensin converting enzyme 2. Glycobiology. 2020;31:410-424.

60. Paolini-Giacobino A, Chen H, Peitsch MC, Rossier C, Antonarakis SE. Cloning of the TMPRSS2 gene, which encodes a novel serine protease with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3. Genomics. 1997;44(3):309-320.

61. Wang Y, Liu M, Gao J. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. Proc Natl Acad Sci USA. 2020;117(25):13967-13974.

62. Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020;581(7807):221-224.

63. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581(7807):215-220.

64. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020;367(6485):1444-1448.

65. Procko E. The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2. bioRxiv. 2020

66. Nelson MI, Gramer MR, Vincent AL, Holmes EC. Global transmission of influenza viruses from humans to swine. J Gen Virol. 2012;93(Pt 10):2195-2203.

67. Novel Swine-Origin Influenza A/VIT, Dawood FS, Jain S, et al. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med. 2009;360(25):2605-2615.

68. Shang J, Wan Y, Luo C, et al. Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci USA. 2020;117(21):11727-11734.

69. Sungnak W, Huang N, Bécavin C, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat Med. 2020;26(5):681-687.

70. Suresh SLPM. Factors influencing the epidemiological characteristics of pandemic COVID 19: A TISM approach. Intl J Healthcare Manage. 2020;13(2):89-98.

SUPPORTING INFORMATION
Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Huang C, Jiang Y, Yan J. Comparative analyses of ACE2 and TMPRSS2 gene: Implications for the risk to which vertebrate animals are susceptible to SARS-CoV-2. J Med Virol. 2021;93:5487–5504. https://doi.org/10.1002/jmv.27073