



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



A novel structure-based approach for identification of vertebrate susceptibility to SARS-CoV-2: Implications for future surveillance programmes

Rahul Kaushik^{a,1}, Naveen Kumar^{b,1}, Kam Y.J. Zhang^a, Pratiksha Srivastava^b, Sandeep Bhatia^b, Yashpal Singh Malik^{c,*}

^a Laboratory for Structural Bioinformatics, Center for Biosystems Dynamics Research, RIKEN, 1-7-22 Suehiro, Yokohama, Kanagawa, 230-0045, Japan

^b Zoonotic Diseases Group, ICAR- National Institute of High Security Animal Diseases, Bhopal, 462022, India

^c College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Science University (GADVASU), Ludhiana, 141004, Punjab, India

ARTICLE INFO

Keywords:

SARS-CoV-2
Vertebrate species
ACE2
Spike protein
Binding affinity
Dissociation constant
Host range

ABSTRACT

Understanding the origin of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been a highly debatable and unresolved issue for scientific communities all over the world. Understanding the mechanism of virus entry to the host cells is crucial to deciphering the susceptibility profiles of animal species to SARS-CoV-2. The interaction of SARS-CoV-2 ligands (receptor-binding domain on spike protein) with its host cell receptor, angiotensin-converting enzyme 2 (ACE2), is a critical determinant of host range and cross-species transmission. In this study, we developed and implemented a rigorous computational approach for predicting binding affinity between 299 ACE2 orthologs from diverse vertebrate species and the SARS-CoV-2 spike protein. The findings show that the SARS-CoV-2 spike protein can bind to a wide range of vertebrate species carrying evolutionary divergent ACE2, implying a broad host range at the virus entry level, which may contribute to cross-species transmission and further viral evolution. Furthermore, the current study facilitated the identification of genetic determinants that may differentiate susceptible from resistant host species based on the conservation of ACE2-spike protein interacting residues in vertebrate host species known to facilitate SARS-CoV-2 infection; however, these genetic determinants warrant *in vivo* experimental confirmation. The molecular interactions associated with varied binding affinity of distinct ACE2 isoforms in a specific bat species were identified using protein structure analysis, implying the existence of diversified bat species' susceptibility to SARS-CoV-2. The current study's findings highlight the importance of intensive surveillance programmes aimed at identifying susceptible hosts, especially those with the potential to transmit zoonotic pathogens, in order to prevent future outbreaks.

1. Introduction

The most conclusive approach for identifying the zoonotic origins of Severe Acute Respiratory Syndrome- Coronavirus- 2 (SARS-CoV-2) is to detect closely related viruses from animal sources. Multiple evidences have emerged supporting a bat species as the immediate ancestor of SARS-CoV-2 (Kumar et al., 2021a; Lytras et al., 2021; Malik et al., 2020; Zhou et al., 2020, 2021), including the role of the intermediate host in transmission of SARS-CoV-2 to humans (Zhang et al., 2020). Although RaTG13, sampled from a *Rhinolophus affinis* bat in Yunnan, China (Zhou

et al., 2021), has the highest nucleotide similarity to SARS-CoV-2, but it cannot be the recent progenitor of SARS-CoV-2 because it lacks evolutionary signatures possessed by SARS-CoV-2 (Kumar et al., 2021a). Following the emergence of SARS-CoV-2, three other bat viruses - RpYN06, RmYN02, and PrC31 – were discovered to exhibit a high nucleotide similarity to that of SARS-CoV-2, throughout much of the virus genome, notably ORF1b, implying that these viruses share a more recent common ancestor with SARS-CoV-2 (Li et al., 2021; Lytras et al., 2021; Zhou et al., 2021). Collectively, these studies demonstrate the zoonotic origin of SARS-CoV-2. However, no bat reservoirs or

* Corresponding author.

E-mail address: malikyps@gmail.com (Y.S. Malik).

¹ These authors contributed equally.

intermediate animal hosts for SARS-CoV-2 have been identified to date. Therefore, the spillover events leading to SARS-CoV-2 transmission directly from bats to humans or via an intermediate host are yet unknown.

The limited information on potential reservoir hosts, as well as the risk to wildlife and livestock, necessitates an immediate need of thorough investigation. Multiple SARS-CoV-2 experimental investigations and natural infection observations have demonstrated the variable susceptibility of vertebrate host species (such as mink, tigers, cats, gorillas, dogs, raccoon dogs, white tailed deer, rabbits, and ferrets) to SARS-CoV-2 infection (WHO, 2021; Cool et al., 2021; Malik et al., 2021; 2021b; Mishra et al., 2021; Hossain et al., 2021). Recent outbreaks in minks have linked SARS-CoV-2 transmission back to humans (reverse zoonosis) and to other animals (van Aart et al., 2021). Except for minks, there is no concrete evidence that other vertebrate host species are either spreaders of SARS-CoV-2 to humans (reverse zoonosis) or reservoir hosts. In addition, susceptibility of the majority of animal species that come into close contact with humans is unknown.

A key to dissect the susceptibility profiles of animal species to SARS-CoV-2 is to understand how the virus enters into the host cells. The receptor-binding domain (RBD) of the SARS-CoV-2 spike protein initially binds to its receptor, angiotensin-converting enzyme 2 (ACE2), before being proteolytically activated by proteases and performing its activity (Shang et al., 2020). Theoretically, the presence of ACE2 receptor in any host species makes them susceptible to SARS-CoV-2 infection, but this is not the case, even in animals with significant ACE2 sequence similarity to human ACE2 (hACE2) (Li et al., 2021). Therefore, a comprehensive understanding of ACE2 diversity in vertebrates, combined with protein-protein interactions at the ACE2-RBD interface, could lead to some novel insights into SARS-CoV-2 susceptibility in various vertebrate species.

In this study, we not only developed and implemented a rigorous computational pipeline for predicting 299 vertebrate ACE2 binding affinities (via dissociation constant) with the spike protein of SARS-CoV-2, but we also demonstrated that dissociation constant is a better predictor of species susceptibility by benchmarking it against experimental data. By finding the best metric for assessing the ACE2-RBD interactions, the molecular interactions leading to a varied binding affinity of ACE2 isoforms in a particular bat species to the spike protein of SARS-CoV-2 are revealed. Furthermore, based on a comparison of key interacting ACE2 residues at the ACE2-RBD interface, the genetic determinants that could aid in differentiating the SARS-CoV-2 susceptible from the resistant species are identified. Overall, the current study identifies a broad host range of vertebrate species susceptible to SARS-CoV-2 for additional experimental investigations, as well as proposing a novel approach for assessing animal species' susceptibility profiles for viruses of interest.

2. Methodology

2.1. ACE2 protein sequences

The protein sequences of ACE2 orthologs ($n = 356$) originating from vertebrates were downloaded from the National Center for Biotechnology Information (NCBI) protein database (5 September 2021) (O'Leary et al., 2016). Partial or identical protein sequences were removed; however, ACE2 isoforms from a particular species were included in the final dataset. As a result, the final dataset included 299 unique ACE2 orthologs from 253 different species. The species-wise numbers of ACE2 orthologs retrieved from the NCBI protein database are provided in Supplementary Table S1.

2.2. Evolutionary divergence of ACE2 orthologs across the vertebrates

The full set of ACE2 protein sequences originating from vertebrates were aligned using MAFFT v.7.475 (Katoh et al., 2019) and the

GTR-Gamma substitution model was selected as the best fit for the dataset using jModelTest 2 (Darrriba et al., 2012). To explore the evolutionary relationship among the distinct vertebrate classes, a phylogenetic tree was constructed using the Maximum Likelihood statistical method with the GTR-Gamma model in RAXML v. 8.2.12 (Stamatakis, 2014). Furthermore, evolutionary diversity and divergence between and within the vertebrate classes were estimated using the Maximum Composite Likelihood model (Tamura et al., 2004). The number of base substitutions per site between sequences were used to calculate their evolutionary distance. Furthermore, the evolutionary divergence within and between the groups is represented by the number of base substitutions per site from averaging over all sequence pairs within each group and between the groups, respectively. The mean evolutionary diversity for the entire population ($\hat{\pi}_T$) was calculated using Eq. (1).

$$\hat{\pi}_T = \frac{q}{q-1} \sum_{i < j} \bar{X}_i \bar{X}_j \hat{d}_{ij} \quad (1)$$

where, 'q' denotes the total number of alleles examined, \bar{X}_i and \bar{X}_j are the estimates of average frequency of the *i*th and *j*th alleles, respectively in the entire population, and \hat{d}_{ij} is the frequency of nucleotide substitutions per site between the *i*th and *j*th alleles. While the mean evolutionary diversity within subpopulations ($\hat{\pi}_S$) was calculated as shown in Eq. (2).

$$\hat{\pi}_S = \sum_{k=1}^s w_k \hat{\pi}_k \quad (2)$$

where, 's' represents the subpopulations, the relative size of the *k*th subpopulation is w_k , and $\hat{\pi}_k$ is the estimate of average nucleotide diversity in the *k*th subpopulation. The inter-population evolutionary diversity ($\hat{\delta}_{ST}$) was calculated as per Eq. (3)

$$\hat{\delta}_{ST} = \hat{\pi}_T - \hat{\pi}_S \quad (3)$$

2.3. Model structures for ACE2 proteins

The recent accomplishments in bioinformatics and computational biology have paved the way for the development of some very reliable de-novo, knowledge-based, and hybrid methods for protein structure prediction that utilize the information inherent to the experimentally solved protein structures (Jumper et al., 2021; Yang et al., 2020; Yang et al., 2015; Bitencourt-Ferreira and de Azevedo, 2019; Kaushik and Jayaram, 2016, Kaushik et al., 2018). We devised and executed a rigorous computational approach to model the 3-D structures for all of the ACE2 orthologs. Fig. 1 depicts an overall workflow of the adopted computational pipeline in this study. Four state-of-the-art methods were used to predict model structures for each ACE2 protein: online version AlphaFold2 (available through Colab-Notebook) (Jumper et al., 2021), standalone versions of Rosetta (*de novo* modeling) (Yang et al., 2020), I-Tasser (version 5.1) (*ab initio* and homology-based hybrid methodology) (Yang et al., 2015), and Modeller (version 10.1) (template-based modeling) (Bitencourt-Ferreira and de Azevedo, 2019), yielding twenty model structures (4 sets of 5 models from each method). The top scoring model structure was chosen from each set of model structures by using a consensus approach for protein structure quality assessment by implementing ProTSAV (Singh et al., 2016), ProFitFun (Kaushik and Zhang, 2020, Kaushik and Zhang, 2022), and ModFold8 (McGuffin et al., 2021). The top scored model structure for each ACE2 protein was refined through a comprehensive protein structure refinement approach implemented in GalaxyRefine (Heo et al., 2013). The refined model structures were re-evaluated using the aforementioned quality assessment methods to identify the best scoring model structure for each ACE2 protein. The comprehensive approaches for protein structure refinement and quality assessment confirmed their reliability and applicability for studying various protein-protein interactions of ACE2 from vertebrates

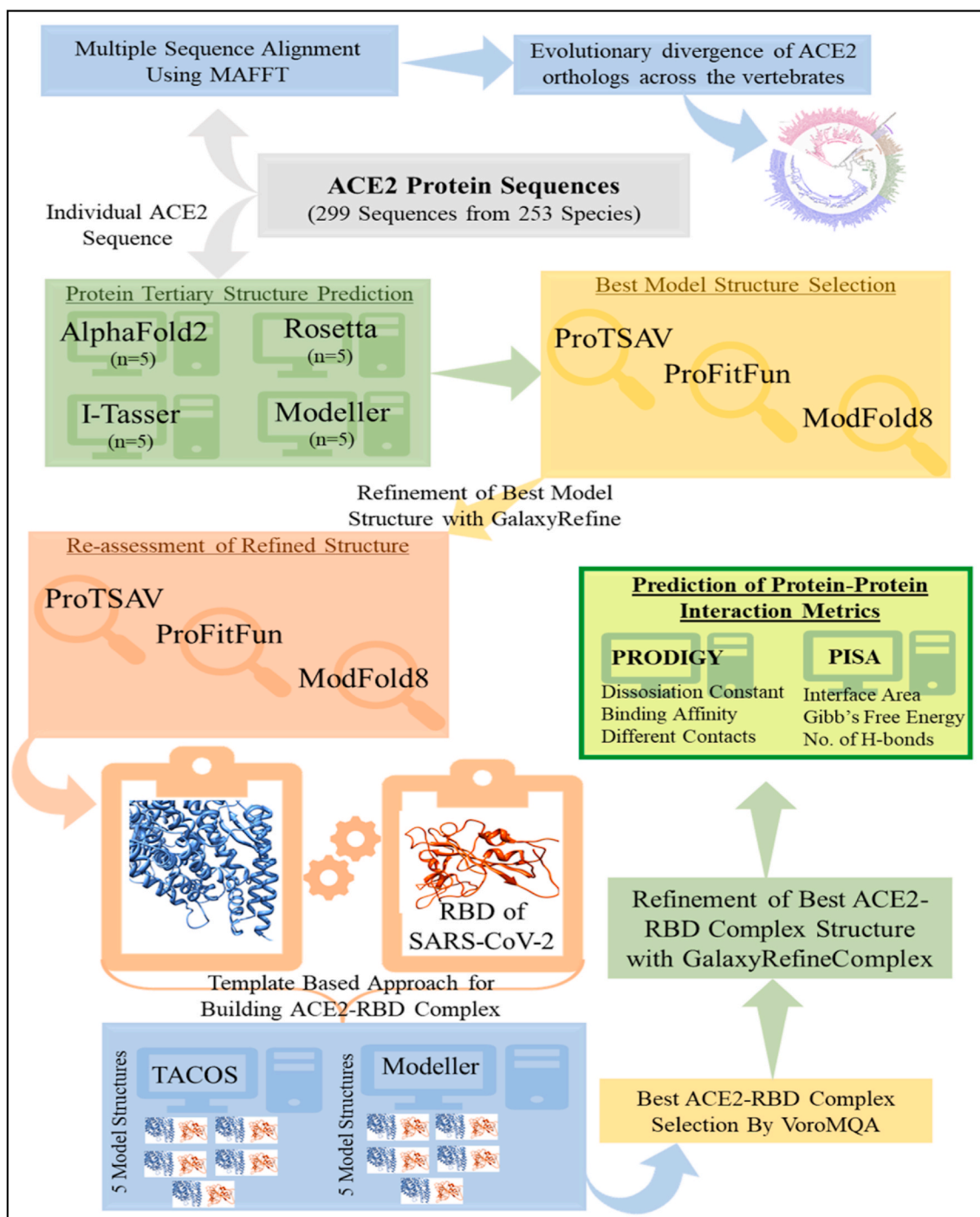


Fig. 1. An overview of the computational methodologies used for predicting binding affinities of 299 vertebrate ACE2 proteins with the RBD of the SARS-CoV-2 spike protein in order to investigate the virus's host range.

with the RBD of SARS-CoV-2 spike protein.

2.4. Template based approach for building ACE2-RBD complex

As the number of experimentally determined protein-protein complexes has increased, the reliability of template-based protein-protein complex predicting algorithms has improved. The major challenge of modeling a protein-protein complex is finding an appropriate template that is well aligned with the target protein sequence. In general, the capacity to identify the interacting protein partners extracted by comparative modeling determines the efficacy of protein-protein

docking in structural biology. Here, a template-based protein-protein docking was performed by implementing a standalone package of TACOS (Template-based Assembly of Complex Structures) (Szilagyí and Zhang, 2014), and the top ranked models of ACE2-RBD protein complex for each ACE2 protein were selected. In addition, using the ACE2-RBD complex of *Rhinolophus macrotis* (PDB ID: 7C8J), *Homo sapiens* (PDB ID: 6M0J), *Canis lupus familiaris* (PDB ID: 7E3J), and *Felis catus* (PDB ID: 7C8D) as reference structures, four more ACE2-RBD complexes for each ACE2 protein were predicted in the Modeller (version 10.1) (template-based modeling) (Bitencourt-Ferreira and de Azevedo, 2019). For each ACE2 protein, this resulted in 5 ACE2-RBD complexes.

2.5. Refinement and optimization of ACE2-RBD complexes

The standalone version of GalaxyRefineComplex (Heo et al., 2016) was used to refine all of the ACE2-RBD docked complexes ($n = 5 \times 299$). For each ACE2-RBD complex, five refined protein complexes were created, resulting in 25 ACE2-RBD complexes for each ACE2 protein. These protein complexes were ranked using VoroMQA's protein-protein complex quality assessment (Olechnovič and Venclovas, 2019) in order to choose the best refined and optimized ACE2-RBD complex for each ACE2 protein ($n = 299$). The top-ranked ACE2-RBD protein complex for each ACE2 protein was utilized for carrying out further structural interaction analysis.

2.6. Benchmarking of binding prediction metrics on dimeric complexes

Different approaches for assessing protein-protein interactions in a protein complex yield different metrics for quantifying the strength of the interactions. Some of these metrics include predicted binding affinity, predicted dissociation constant (Kd), change in Gibbs free energy (ΔG), binding energy, number of interactions (hydrophobic-hydrophobic, polar-polar, and polar-hydrophobic), number of H-bonds, and number of di-sulfide bonds. Here, the PDBbind database was used to create a dataset of heterodimeric protein complexes ($n = 400$) in order to discover the best metric for assessing protein-protein interactions (Liu et al., 2015).

There are 154 complexes in all in this heterodimeric protein complexes dataset, each having at least one enzymatic protein (Supplementary Table S2). The experimental values of dissociation constants for these complexes, adopted from the PDBbind Database (Liu et al., 2015), were used to compare the predicted protein-protein binding metrics. The standalone versions of PRODIGY (Xue et al., 2016) and PISA (Battelle, 2016) were used to calculate various types of contacts (intermolecular, charged-charged, charged-polar, polar-polar, apolar-polar, and apolar-apolar), predicted binding affinity (Kcal/mol), predicted dissociation constant (Kd), interface area (\AA), number of hydrogen bonds, salt bridges, disulfide bonds, and change in Gibbs free energy (ΔG). A correlation analysis of these parameters with the experimental dissociation constant values was performed to identify the most accurately predicted metric. The result showed that the predicted dissociation constant had the best correlation with the experimental values. As a result, the predicted dissociation constant was chosen as the main metric for quantifying the protein-protein interactions in the ACE2-RBD complexes.

2.7. Estimation of binding prediction metrics at the interface of ACE2-RBD complexes

The interacting interfaces of RBD and ACE2 were extracted from the best ACE2-RBD protein complex for each ACE2 protein and analyzed to calculate the dissociation constant using the standalone version of PRODIGY (Xue et al., 2016), a state-of-the-art method for predicting the binding in protein-protein complexes. PRODIGY was also used to predict the binding affinity of various types of contacts (intermolecular, charged-charged, charged-polar, polar-polar, apolar-polar, and apolar-apolar). Furthermore, the interface area, number of hydrogen bonds, salt bridges, and disulfide bonds, and change in Gibbs free energy (ΔG) were also estimated using the standalone version of PISA (Battelle, 2016), implemented in CPP4 software suite (Supplementary Table S3). The predicted dissociation constants were utilized to quantify and build a better understanding of SARS-CoV-2 host preferences.

2.8. Statistical analysis

We used the GraphPad Prism 7.01 (GraphPad Software, San Diego, CA) for all the statistical analysis. The multiple t-tests with Holm-Sidak adjustments method was employed to assess the variations in evolutionary distances across the vertebrate classes. A p-value of less than

0.01 was considered statistically significant. GraphPad Prism 7.01 software was used to create all of the graphs.

3. Results

3.1. ACE2 protein's evolutionary diversity across a wide range of vertebrate species

The protein sequences of 299 ACE2 orthologs downloaded from the NCBI protein database were analyzed and classified into six vertebrate taxonomic classes: 86 *Actinopterygii* (bony fishes), 6 *Amphibia* (amphibians), 43 *Aves* (birds), 2 *Chondrichthyes* (cartilaginous fishes), 143 *Mammalia* (mammals), and 19 *Reptilia* (reptiles), and subsequently their evolutionary diversity and divergence were estimated (Fig. 2). All the ACE2 protein sequences have a mean evolutionary diversity ($d \pm SE$) of 0.43 ± 0.02 , indicating a huge diversity among these protein sequences. Furthermore, the mean evolutionary diversity of ACE2 protein sequences between and within classes was 0.20 ± 0.01 , and 0.22 ± 0.01 , respectively. In addition, the average evolutionary divergence ($d \pm SE$) of ACE2 protein sequences among the six vertebrate taxonomic classes ranged from 0.27 ± 0.01 to 0.13 ± 0.009 , with *Aves* having the least average evolutionary divergence among all the vertebrate classes ($P < 0.0001$) (Fig. 3A). Furthermore, ACE2 protein sequences from *Mammalia* have a statistically higher mean evolutionary divergence than ACE2 from *Aves* ($P < 0.0001$), but a lower mean evolutionary divergence than ACE2 from *Actinopterygii* ($P = 0.0006$) (Fig. 3B). Nevertheless, the mean evolutionary divergence of ACE2 protein sequences from *Mammalia* is non-significantly related to *Amphibia*, *Reptilia*, and *Chondrichthyes* ($P = 0.018-0.569$).

In addition, the ACE2 protein sequences of *Homo sapiens* showed the highest sequence identity of 99.38% with that of *Hominidae* family members (*Gorilla gorilla gorilla*, *Pan paniscus*, and *Pan troglodytes*) followed by the *Hylobatidae* family members, with sequence identities of 98.76% and 98.29% (*Hylobates moloch*, and *Nomascus leucogenys*, respectively), and *Cercopithecidae* family members (*Chlorocebus sp.*, *Macaca sp.*, *Papio anubis*, *Ptilocolobus tephrosceles*, *Rhinopithecus roxellana*, and *Theropithecus gelada*) with sequence identities of 96.05–96.86% (Supplementary Table S4). The ACE2 protein sequences of *Homo sapiens*, on the other hand, showed the poor sequence identities with the distinct ACE2 isoforms of *Rhinolophus sp.* (77.0–78.14%) and also with different species of bats, such as *Pteropus sp.* (76.31–78.59%), *Pipistrellus sp.* (70.47–71.48%), *Myotis sp.* (77.46%), *Desmodus rotundus* (76.54%), and *Eptesicus fuscus* (77.92%). These findings show that the ACE2 protein sequences in *Mammalia* have a great deal of evolutionary diversity.

3.2. Prediction of vertebrate ACE2 orthologs binding affinities to spike protein of SARS-CoV-2

In order to discover the best metric for assessing ACE2-RBD interactions, the predicted dissociation constant (Kd), and predicted change in Gibbs free energy (ΔG) were benchmarked against the experimental binding affinity values of dimeric protein complexes ($n = 400$) obtained from the PDBbind database (Liu et al., 2015). The benchmarking of these two matrices on predicted and experimentally values derived from 400 dimeric protein complexes revealed that the predicted dissociation constant is highly correlated with experimental values ($r = 0.858$, $P < 0.0001$), whereas the predicted change in Gibbs free energy (ΔG) is poorly correlated with the experimental values ($r = 0.125$, $P < 0.012$). Therefore, we first devised and conducted a rigorous computational approach to construct ACE2-RBD models for all 299 unique ACE2 orthologs, and then predicted dissociation constant (Kd) for assessing their interactions. Intriguingly, based on the predicted dissociation constant, the present study found that the binding affinity of dACE2 with RBD is 4.15 times lower than that of hACE2, which is comparable (6.65 times lower) to an experimental study that calculated

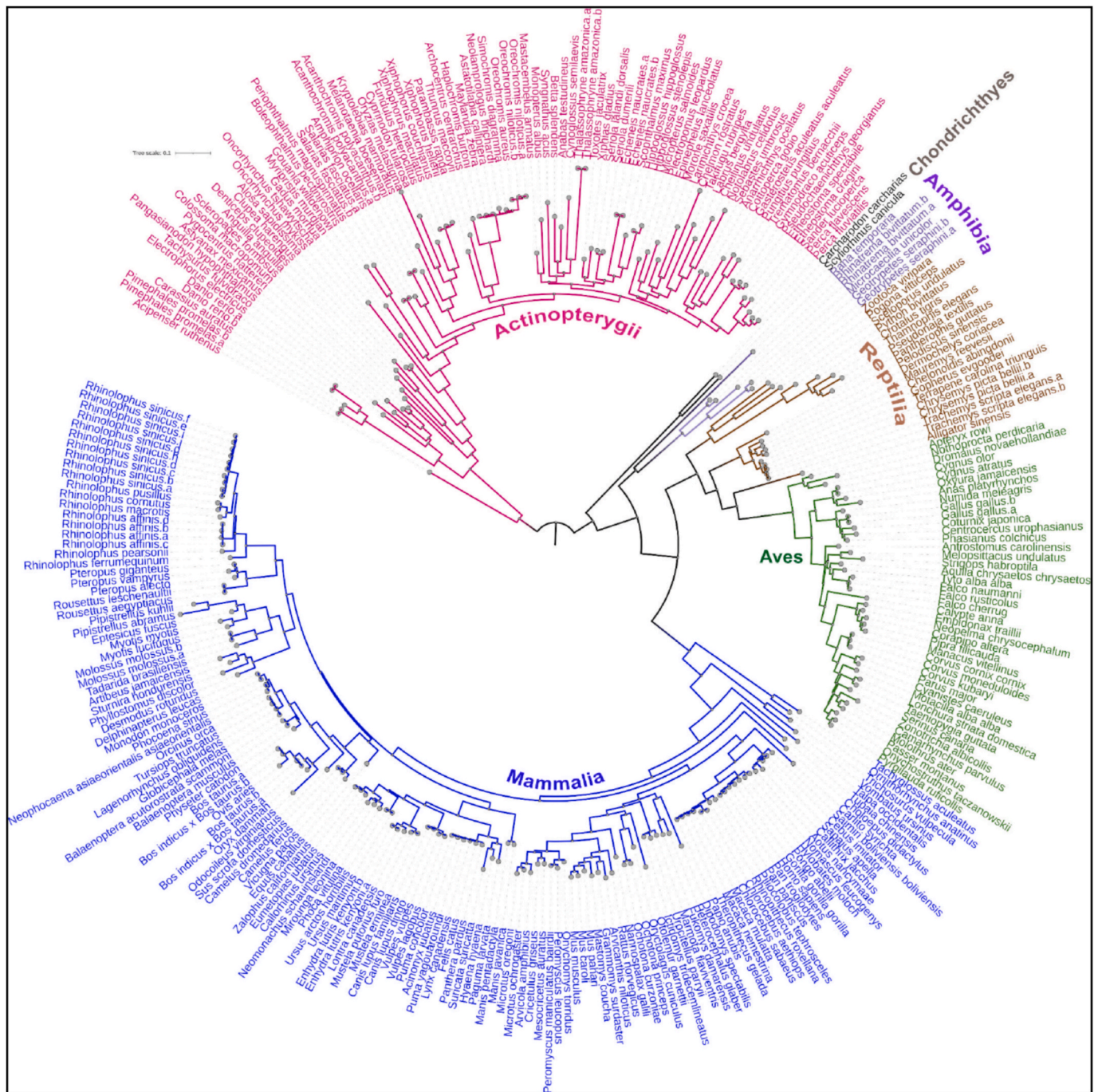


Fig. 2. Maximum-likelihood (ML) tree displaying the phylogenetic relationships of all 299 vertebrate ACE2 protein sequences. The ACE2 protein sequences are classified into six vertebrate taxonomic classes; *Actinopterygii* (bony fishes), *Amphibia* (amphibians), *Aves* (birds), *Chondrichthyes* (cartilaginous fishes), *Mammalia* (mammals), and *Reptilia* (reptiles), and are colored differently.

the binding affinity of dACE2/hACE2 to RBD using surface plasmon resonance (Zhang et al., 2021). Subsequently, in order to assess the ACE2-RBD interactions based on the predicted dissociation constants, vertebrate species were categorized into four groups based on their propensity to bind to SARS-CoV-2: very high, high, medium, and low. The predicted dissociation constants values for all the vertebrate species are provided in Supplementary Table S5. The ACE2 proteins from three mammals (Southern elephant seal, Cat, and North American river otter), three reptiles (Chinese Alligator, Leatherback sea turtle, and Plateau fence lizard), and one bird species (White-ruffed manakin) were predicted to possess a very high propensity to bind to SARS-CoV-2. In addition, ACE2 proteins of 120 (16 bony fishes, 18 birds, 75 mammals,

and 11 reptiles), 155 (61 bony fishes, 23 birds, 62 mammals, 5 reptiles, 3 amphibians, and 1 cartilaginous fish), and 17 species (10 bony fishes, 3 mammals, 3 amphibians, and 1 cartilaginous fish), respectively were classified as having a high, medium, and low propensity to bind to SARS-CoV-2 (Supplementary Table S5).

Next, a comparison of the species' predicted propensity (based on predicted Kd) to bind to SARS-CoV-2 with the experimentally proven susceptibility of infection (both natural and experimental infection) was performed in 46 species (high, medium, low, and extremely low), as depicted in Fig. 4. Intriguingly, all 46 species were distributed across the very high, high, and medium propensity categories predicted in our study. For example, it was found that 13 species designated as highly

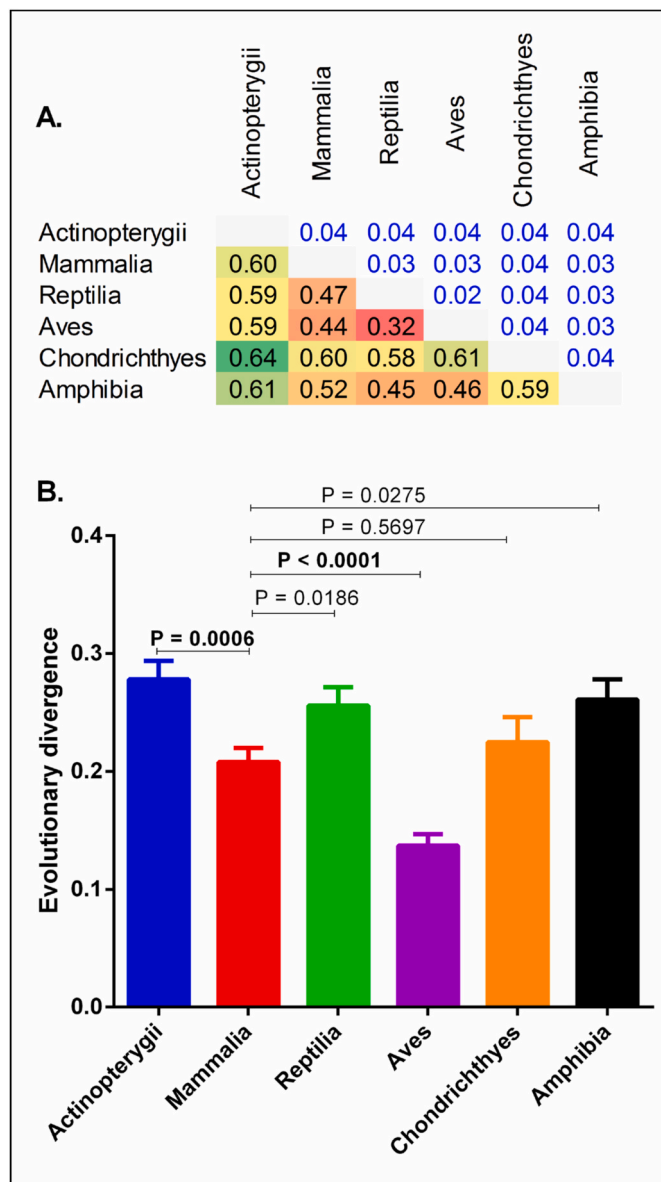


Fig. 3. (A) Depicts an evolutionary divergence matrix among the six vertebrate taxonomic classes. (B) Compares the mean evolutionary divergence of the six vertebrate taxonomic classes. A p-value less than 0.01 was considered statistically significant.

susceptible to SARS-CoV-2 infection based on experimental studies fall into very high ($n = 1$, cat), high ($n = 7$, White-tailed deer, Human, Western lowland gorilla, Ferret, Prairie deer mouse, Golden hamster, and Mountain lion), and medium ($n = 5$, Leopard, Rhesus macaque, Common marmoset, American mink, and Egyptian rousette) propensity categories of our study, while one species (Rabbit) designated as medium susceptible to SARS-CoV-2 infection occupied the medium propensity category of our study.

3.3. Comparison of predicted ACE2 binding affinity and their interacting residues among experimentally proven susceptible species

We identified a total of 54 species, 46 of which were proven to support SARS-CoV-2 infection either naturally or experimentally, while the remaining 8 species predicted to be resistant to SARS-CoV-2 infection were used as controls in the current study. Following that, we combined and compared their predicted ACE2 binding affinity (Kd) with

the fraction of interacting ACE2 residues similar to hACE2, in order to gain insight into their differential propensity to bind to SARS-CoV-2 (Fig. 4). The predicted dissociation constant is negatively correlated with the fraction of interacting ACE2 residues similar to hACE2 ($r = -0.5997$, $P < 0.0001$), implying that binding affinity of ACE2 to the spike protein of SARS-CoV-2 is associated with the number of interacting ACE2 residues similar to hACE2. Subsequently, we identified 30 amino acid residues in ACE2 that interact with the spike protein's RBD of SARS-CoV-2, and these residues were examined for their similarity to the residues in hACE2. The species susceptibility was classified into three categories (high, medium, and low) based on the number of residues identical to hACE2; at least 24 residues identical to hACE2 for high (at least 24/30), 15–23 residues for medium (15–23/30), and less than 14 residues for low (less than 14/30) susceptibility category. Nine species (Cat, White-tailed deer, Human, Golden snub-nosed monkey, Olive baboon, Western lowland gorilla, Chinese hamster, Gelada, and Sumatran orangutan) that were predicted to have a high propensity (based on the similarity of at least 24/30 residues with that of hACE2) also had a very high to high propensity based on the predicted dissociation constant. However, comparing ACE2 residues for their similarity to the residues in hACE2 alone does not truly reflect their propensity for SARS-CoV-2 because experimentally proven highly susceptible species to SARS-CoV-2 do not necessarily possess a comparatively higher number of ACE2 interacting residues than hACE2, for example, Ferret (17/30), American mink (16/30), and Leopard (20/30). Of note, interpreting species susceptibility for SARS-CoV-2 should be done with caution; thus, combining predicted ACE2 binding affinity (Kd) with the fraction of interacting ACE2 residues similar to hACE2 could yield more confidence in predicting propensity of species susceptibility to SARS-CoV-2 infection. The species predicted to be resistant to SARS-CoV-2 infection carried $\leq 14/30$ residues similar to that of hACE2, and also had high Kd values (low binding affinity).

Previous studies have identified five critical ACE2 interacting residues based on their conservation in known susceptible species (31K/T, 35E/K, 38D/E, 82T/M/N, and 353 K) (Shang et al., 2020; Wan et al., 2020). We performed sequence alignment of ACE2 interacting residues from 54 different species, and revealed that 31K/T, 35E/K, 38D/E, and 353 K residues, despite being highly conserved in high and medium susceptible species, are also found in some of the low susceptible or resistant species. In contrast, 82T/M/N residues are highly conserved and found only in high and medium susceptible species. Based on the sequence alignment of ACE2 interacting residues, our analysis proposes six unique residues that could help in differentiation of susceptible from the resistant species; Susceptible species (27T/I, 30D/E/Q, 82M/T/D/N, 326G/E/R/T, and 352G+353 K + 354G), resistant species (N326 + N330), and 352G+353 K + 354H/R/Q for reduced susceptibility to SARS-CoV-2 infection.

Next, based on the predicted binding affinity (in terms of Kd), similarity score (ACE2 interacting residues similar to that of hACE2), and six unique residues proposed in this study as determinants of SARS-CoV-2 susceptibility, ten bat species (*Rhinolophus sinicus*, *R. affinis*, *R. macrotis*, *R. pearsonii*, *R. ferrumequinum*, *Myotis lucifugus*, *Artibeus jamaicensis*, *Rousettus aegyptiacus*, *Pteropus vampyrus*, and *Myotis myotis*) were predicted as medium susceptibility to SARS-CoV-2 infection, while the remaining 8 bat species were predicted to be resistant to SARS-CoV-2 infection (Supplementary Table S6). Intriguingly, two of the ten ACE2 isoforms identified in *Rhinolophus sinicus* and four ACE2 isoforms identified in *Rhinolophus affinis* were predicted to have a high binding affinity with the spike protein of SARS-CoV-2.

3.4. Structural insights into differential ACE2 isoforms binding affinity

The interaction interfaces of hACE2, dACE2, rsACE2 (two isoforms), and raACE2 with RBD were analyzed separately from the refined ACE2-RBD complexes using PRODIGY (Xue et al., 2016) and then compared with the hACE2-RBD interface to investigate the molecular basis of

Species	Predicted Kd (M)	ACE2 interacting residues																				No. of residues matching with hACE2	Percent of residues matching with hACE2														
		S19	E23	Q24	A25	T27	F28	D30	K31	H34	E35	E37	D38	Y41	Q42	L45	L79	A80	M82	Y83	T324			Q325	G326	N330	L351	G352	K353	G354	D355	R357	-	R377			
VERY HIGH																																					
<i>Felis catus</i> (Cat)	9.20E-11	S	-	L	A	T	F	-	E	K	H	E	E	E	Y	Q	L	-	L	T	-	Y	T	Q	G	N	L	G	K	G	D	R	-	R	24	0.80	
HIGH																																					
<i>Physeter catodon</i> (Sperm whale)	1.00E-10	S	-	Q	-	T	F	-	Q	K	H	E	E	D	Y	Q	L	-	L	T	T	Y	-	Q	E	N	L	G	K	G	D	R	A	R	22	0.73	
<i>Odocoileus virginianus</i> (White-tailed deer)	1.40E-10	S	E	Q	-	T	F	-	E	K	H	E	E	D	Y	Q	L	-	M	A	T	Y	T	Q	G	N	L	G	K	G	D	R	A	R	26	0.87	
<i>Homo sapiens</i> (Human)	2.00E-10	S	E	Q	A	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	A	M	Y	T	Q	G	N	L	G	K	G	D	R	-	R	30	1.00	
<i>Rhinopithecus roxellana</i> (Golden snub-nosed monkey)	2.90E-10	S	-	Q	A	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	-	M	Y	T	-	G	N	L	G	K	G	D	R	A	R	27	0.90	
<i>Papio anubis</i> (Olive baboon)	3.00E-10	-	E	Q	A	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	-	M	Y	-	-	G	N	L	G	K	G	D	R	A	R	26	0.87	
<i>Gorilla gorilla gorilla</i> (Western lowland gorilla)	3.60E-10	S	E	Q	A	T	F	-	L	D	K	H	E	E	D	Y	Q	L	-	L	-	M	Y	T	-	G	N	L	G	K	G	D	R	A	R	28	0.93
<i>Cricetulus griseus</i> (Chinese hamster)	3.80E-10	-	E	Q	K	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	A	N	Y	T	Q	G	N	-	G	K	G	D	R	-	R	25	0.83	
<i>Eumetopias jubatus</i> (Steller sea lion)	3.90E-10	-	-	S	L	T	F	-	E	K	S	E	E	E	Y	Q	L	-	L	Q	T	Y	-	Q	G	N	L	G	K	H	D	R	A	R	18	0.60	
<i>Paguma larvata</i> (Masked palm civet)	4.10E-10	S	L	A	K	T	F	-	L	E	T	Y	E	E	Y	Q	V	K	L	A	T	Y	-	-	G	N	L	G	K	G	D	R	-	R	19	0.63	
<i>Mustela putorius furo</i> (Ferret)	4.20E-10	-	-	S	L	T	F	-	E	K	Y	E	E	E	Y	Q	-	-	L	H	T	Y	T	-	G	N	-	G	K	R	D	R	-	R	17	0.57	
<i>Theropithecus gelada</i> (Gelada)	4.70E-10	-	-	Q	K	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	A	M	Y	-	Q	G	N	-	G	K	G	D	R	A	R	25	0.83	
<i>Peromyscus maniculatus bairdi</i> (Prairie deer mouse)	4.80E-10	S	E	Q	A	I	F	-	D	K	H	E	E	D	Y	Q	L	-	L	-	N	Y	T	-	G	R	N	-	G	K	G	D	R	-	R	23	0.77
<i>Globicephala melas</i> (Long finned pilot whale)	4.90E-10	-	-	S	R	T	F	-	Q	K	H	E	E	D	Y	Q	L	-	I	A	T	Y	T	Q	G	N	-	G	K	G	D	R	-	-	20	0.67	
<i>Neophocaena asiakorialis</i> (Yangtze porpoise)	5.90E-10	-	-	S	Q	T	F	-	Q	K	H	E	E	D	Y	Q	L	R	I	A	T	Y	T	E	-	N	L	G	K	G	D	R	-	-	21	0.70	
<i>Zalophus californianus</i> (California sealion)	6.00E-10	S	-	L	K	T	F	-	E	K	S	E	E	E	Y	Q	L	-	-	Q	T	Y	-	Q	G	N	-	G	K	H	D	R	A	R	18	0.60	
<i>Camelus ferus</i> (Wild Bactrian camel)	6.20E-10	S	E	L	A	T	F	-	E	E	H	E	E	D	Y	Q	L	-	-	T	T	Y	-	Q	G	N	-	G	K	G	D	R	-	R	22	0.73	
<i>Mesocricetus auratus</i> (Golden hamster)	6.20E-10	-	-	S	Q	T	F	-	D	K	Q	E	E	D	Y	Q	L	-	L	-	N	Y	G	E	-	N	L	G	K	G	D	R	A	R	20	0.67	
<i>Puma concolor</i> (Mountain lion)	6.40E-10	S	-	L	A	T	F	-	E	K	H	E	E	D	Y	Q	L	-	L	-	T	Y	-	-	G	N	L	G	K	G	D	R	A	R	21	0.70	
<i>Tursiops truncatus</i> (Atlantic bottle-nosed dolphin)	6.70E-10	-	-	E	R	T	F	-	Q	K	R	E	E	D	Y	Q	L	-	I	-	T	Y	T	Q	G	-	-	G	K	G	D	R	A	R	19	0.63	
<i>Neomonachus schauinslandi</i> (Hawaiian monk seal)	6.90E-10	S	D	L	V	T	F	-	E	K	Y	E	E	E	Y	Q	L	-	-	Q	T	Y	-	Q	G	N	L	G	K	H	D	R	A	R	19	0.63	
<i>Ptilocobolus tephrosceles</i> (Ugandan red Colobus)	7.00E-10	-	-	S	Q	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	-	M	Y	-	-	G	N	-	G	K	G	D	R	-	-	22	0.73	
<i>Delphinapterus leucas</i> (Beluga whale)	7.70E-10	-	-	E	Q	-	T	F	-	Q	K	H	E	E	D	Y	Q	L	-	I	-	T	Y	-	-	E	N	-	G	K	G	D	R	-	-	20	0.67
<i>Canis lupus familiaris</i> (Dog)	8.30E-10	-	-	D	L	T	F	-	E	K	Y	E	E	E	Y	Q	L	-	L	A	T	Y	T	Q	E	N	L	G	K	G	D	R	A	R	21	0.70	
<i>Orcinus orca</i> (Killer whale)	8.60E-10	-	-	-	R	T	F	-	Q	K	R	E	E	D	Y	Q	L	R	I	A	T	Y	-	Q	G	N	-	-	K	G	D	R	-	-	19	0.63	
<i>Pongo abelii</i> (Sumatran orangutan)	8.90E-10	S	E	Q	-	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	A	M	Y	-	Q	R	N	L	G	K	G	D	R	-	-	27	0.90	
<i>Manis javanica</i> (Malayan pangolin)	9.30E-10	S	E	-	A	T	F	-	E	K	S	E	E	E	Y	Q	L	-	I	-	N	Y	T	Q	T	E	N	G	K	H	D	R	M	R	19	0.63	
MEDIUM																																					
<i>Ursus maritimus</i> (Polar bear)	1.00E-09	-	-	D	L	A	T	F	L	E	K	Y	E	E	D	Y	Q	L	-	H	T	Y	-	-	E	N	L	G	K	G	D	R	-	R	19	0.63	
<i>Lynx canadensis</i> (Canada lynx)	1.10E-09	S	-	L	A	T	F	-	L	E	K	H	E	E	Y	Q	L	-	L	-	T	Y	-	-	G	N	-	G	K	G	D	R	A	R	21	0.70	
<i>Pan troglodytes</i> (Chimpanzee)	1.20E-09	-	-	-	Q	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	-	M	Y	-	-	G	N	-	G	K	G	D	R	-	-	22	0.73	
<i>Oryctolagus cuniculus</i> (Rabbit)	1.70E-09	-	-	-	L	T	F	-	E	K	Q	E	E	D	Y	Q	L	-	L	T	-	Y	T	-	-	G	N	L	G	K	R	D	R	-	-	20	0.67
<i>Panthera pardus</i> (Leopard)	1.70E-09	-	-	S	L	K	T	F	-	E	K	H	E	E	Y	Q	L	-	L	T	-	Y	-	Q	G	E	L	G	K	G	D	R	-	-	20	0.67	
<i>Vulpes vulpes</i> (Red fox)	1.70E-09	-	-	L	N	T	F	-	E	K	Y	E	E	E	Y	Q	L	-	L	T	-	Y	-	Q	G	N	-	G	K	G	D	R	A	R	19	0.63	
<i>Equus caballus</i> (Horse)	1.80E-09	S	D	L	A	T	F	-	E	K	S	E	E	E	H	Q	L	-	K	L	T	Y	T	-	G	N	-	G	K	G	D	R	Y	R	19	0.63	
<i>Hylobates moloch</i> (Silvery gibbon)	1.80E-09	-	-	-	Q	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	-	M	Y	T	Q	G	N	-	G	K	G	D	R	A	-	23	0.77	
<i>Macaca mulatta</i> (Rhesus macaque)	1.90E-09	-	-	-	Q	T	F	-	D	K	H	E	E	D	Y	Q	L	-	L	-	M	Y	T	-	G	N	-	G	K	G	D	R	A	R	23	0.77	
<i>Peromyscus leucopus</i> (White-footed mouse)	1.90E-09	S	E	Q	A	I	F	-	D	K	Q	E	E	D	Y	Q	L	-	L	-	N	Y	T	-	G	R	N	-	G	K	G	D	R	-	R	23	0.77
<i>Bos taurus</i> (Domesticated cattle)	2.00E-09	-	-	S	Q	T	F	-	E	K	H	E	E	D	Y	Q	L	-	M	-	T	Y	T	Q	G	N	L	G	K	G	D	R	-	-	22	0.73	
<i>Calithrix jacchus</i> (Common marmoset)	2.00E-09	S	-	Q	K	T	F	-	D	K	H	E	E	D	H	E	L	-	L	-	T	Y	-	-	G	N	L	G	K	Q	D	R	-	-	21	0.70	
<i>Sus scrofa domestica</i> (Domestic pig)	2.00E-09	-	-	L	A	T	F	-	E	K	L	E	E	D	Y	Q	L	-	K	I	T	Y	-	-	G	N	-	G	K	G	D	R	-	-	19	0.63	
<i>Urocitellus parryii</i> (Arctic ground squirrel)	2.00E-09	-	-	A	Q	L	T	F	-	D	K	H	E	E	D	H	Q	L	K	L	A	D	Y	T	Q	G	N	L	Q	K	G	D	R	A	R	23	0.77
<i>Ovis aries</i> (Sheep)	2.10E-09	S	-	Q	-	T	F	-	E	K	H	E	E	D	Y	Q	L	-	M	A	T	Y	-	-	G	N	-	G	K	G	D	R	-	-	22	0.73	
<i>Neovison vison</i> (American mink)	2.40E-09	-	-	D	L	T	F	-	L	E	K	Y	E	E	E	Y	Q	L	-	H	T	Y	-	-	G	N	-	G	K	H	D	R	A	R	16	0.53	
<i>Rhinolophus sinicus</i> (Chinese rufous horseshoe bat)	3.60E-09	-	-	F	L	A	I	F	-	D	E	F	E	E	N	Y	Q	L	-	L	-	N	Y	T	E	G	N	-	G	K	G	D	R	-	-	19	0.63
<i>Rousettus aegyptiacus</i> (Egyptian rousette)	5.30E-09	S	E	-	L	T	F	-	E	K	T	E	E	D	Y	Q	L	K	L	A	T	Y	T	K	-	-	-	G	K	G	D	R	-	-	21	0.70	
<i>Danio rerio</i> (Zebrafish)	5.90E-09	-	-	-	R	E	F	-	N	K	-	-	-	-	-	-	-	E	S	A	Y	-	-	-	-	N	N	N	R	K	D	R	-	-	13	0.43	
LOW																																					
<i>Oryzias latipes</i> (Marine medaka)	1.00E-08	S	E	L	Q	E	F	-	-	K	G	E	Q	E	Y	Q	-	-	L	E	K	Y	F	N	N	N	-	-	R	E	D	R	-	-	13	0.43	
<i>Carcharodon carcharias</i> (Great white shark)	1.10E-08	-	-	E	A	-	F	-	-	K	K	E	N	Y	N	L	-	-	E	K	F	N	Y	M	G	N	-	G	-	A	D	R	A	R	13	0.43	
<i>Labrus bergylla</i> (ballan wrasse)	1.10E-08	-	-	Q	K	E	F	-	-	K	R	E	E	V	Y	-	-	-	L	E	K	Y	F	P	N	N	M	N	R	E	D	R	-	-	11	0.37	
<i>Gadus morhua</i> (Atlantic cod)	1.50E-08	-	-	R	A	A	F	-	E	K	T	E	S	V	D	L	-	-	E	A	F	E	P	D	K	N	-	G	N	G	D	R	-	-	13	0.43	
<i>Pimephales promelas</i> (Fathead minnow)	1.70E-08	S	-	Q	G	E	F	-	-	K	K	E	D	N	Y	Q	L	-	E	A	F	F	D	N	N	-	-	R	E	D	R	-	-	13	0.43		
<i>Rhinatrema bivittatum</i> (Two-lined caecilian)	2.1																																				

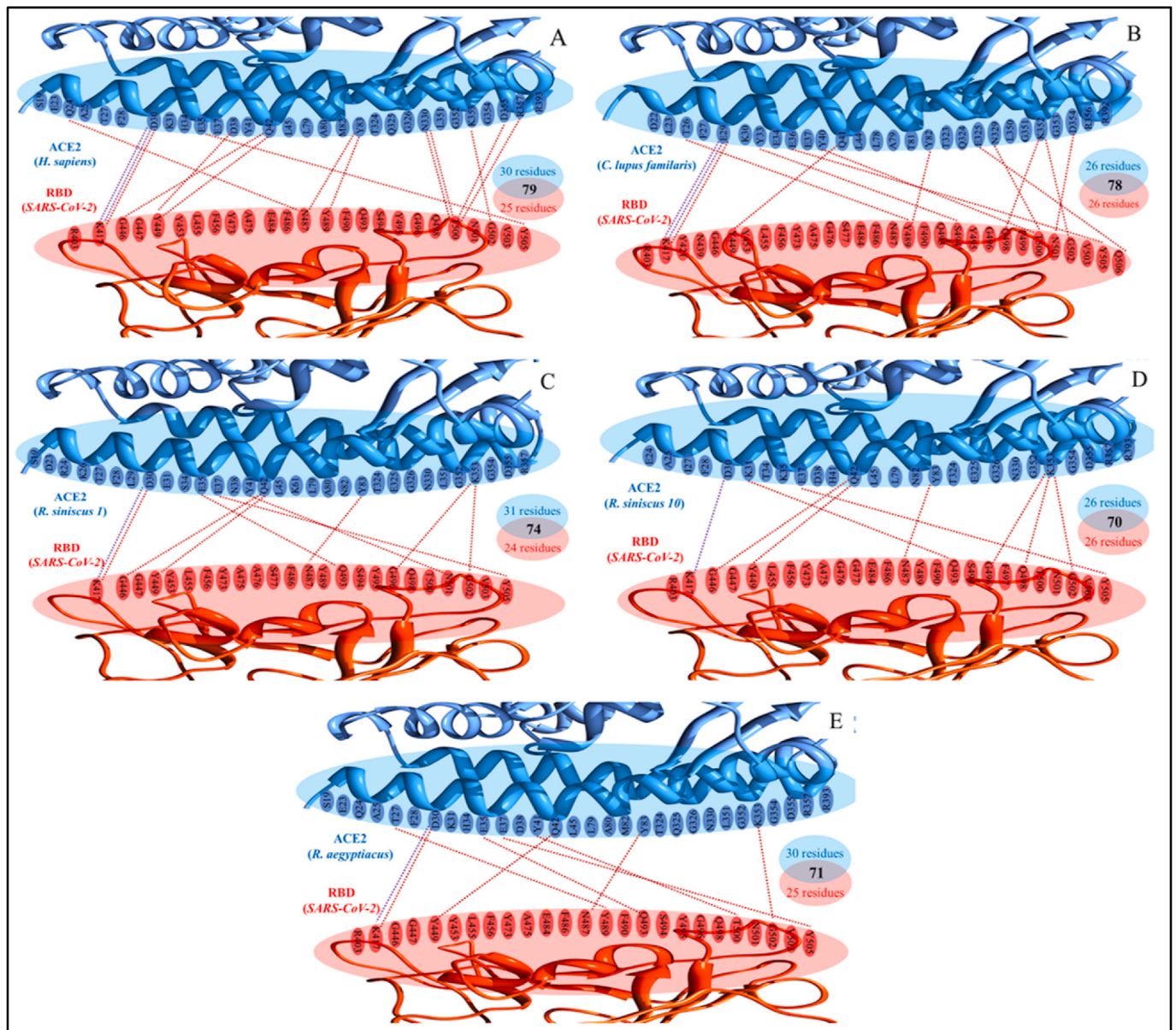


Fig. 5. Structural insights into differential ACE2 binding affinities of selected vertebrate species to SARS-CoV-2 spike protein. The molecular interactions at the ACE2-RBD interface for (A) *Homo sapiens*, (B) *Canis lupus familiaris*, (C) *Rhinolophus sinicus* isoform 1 (GenBank accession no. 1883189465), (D) *Rhinolophus sinicus* isoform 10 (GenBank accession no. 1883189431), and (E) *Rousettus aegyptiacus* are shown. The red and blue dotted lines at the ACE2-RBD interface represent hydrogen bonds and salt bridges, respectively. The total number of interacting residues of ACE2 (blue) and RBD (red), including the total intermolecular contacts at ACE2-RBD interface are also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

ACE2 receptor on host cells, allowing the virus to enter the host cells (Shang et al., 2020a, 2020b). To explore the possible origin of SARS-CoV-2, species at risk, and species that could potentially serve as the intermediate hosts, we first presented a deeper understanding of ACE2 evolutionary diversity, followed by structural insights at the ACE2-RBD interface in a variety of vertebrate species.

Computational tools are the method of choice for examining the protein-protein interactions in a protein complex, especially when studying a large dataset, because many protein structures have yet to be solved experimentally. To forecast different species susceptibility to SARS-CoV-2, previous studies relied on either the comparison of twenty-five amino acids corresponding to the known SARS-CoV-2 Spike protein receptor binding residues for their similarity to the residues in human ACE2 or the prediction of binding energies (Damas et al., 2020; Lan et al., 2020; Rodrigues et al., 2020; Liu et al., 2021b; Shang et al., 2020a, 2020b; Sun et al., 2020). Indeed, different approaches may yield

different metrics for assessing interaction strength, and the resulting mispredictions may affect the reliability of the ACE2 interactions with the spike protein of SARS-CoV-2. Therefore, to find the best metric for assessing ACE2-RBD interactions, we devised and implemented a rigorous computational approach to generate ACE2-RBD protein complex models for 299 ACE2 orthologs, and then benchmarked the predicted dissociation constant (Kd), and change in Gibbs free energy (ΔG) against experimental binding affinity values of dimeric protein complexes ($n = 400$) retrieved from the PDBbind database. The results revealed that predicted dissociation constants are highly correlated with experimental values ($r = 0.858$, $P < 0.0001$), and predicted binding affinity of dACE2/hACE2 to RBD is comparable to the experimental binding affinity (Zhang et al., 2021; Kumar et al., 2021a). Together, these findings support the robustness and reliability of the adopted approach used in this study.

It is worth noting that the findings are based on predicted propensity

(dissociation constants) to bind to SARS-CoV-2 for categorizing vertebrate species into very high, high, medium, and low propensity groups. Intriguingly, the results demonstrated that predicted binding affinity of ACE2 with RBD of SARS-CoV-2 based on dissociation constants is a better descriptor of species susceptibility to SARS-CoV-2 because all 46 vertebrate species known to support SARS-CoV-2 infection based on natural and experimental infections were correctly predicted in our study (WHO, 2021). In previous studies, new world monkey, such as marmosets, were predicted to be less susceptible or resistant to SARS-CoV-2 infection (Damas et al., 2020; Liu et al., 2021a; Shi et al., 2020). In contrast, the current study defined marmosets as belonging to medium propensity category (10 times lower predicted binding affinity than that of hACE2). This finding is supported by a recent experimental study in which older marmosets developed mild infection after being exposed to SARS-CoV-2 (Singh et al., 2021). Furthermore, the high susceptibility of white-tailed deer predicted in our study is in line with recent studies demonstrating that white-tailed deer are highly susceptible to SARS-CoV-2 infection naturally (Cool et al., 2021; Kuchipudi, 2022). Additionally, the current study found a medium propensity for cattle and pigs, which is consistent with a previous study that showed efficient entry of SARS-CoV-2 in A549 cells expressing cattle and pig ACE2 (Liu et al., 2021a).

The previous studies that proposed five critical ACE2 interacting residues based on their conservation in known susceptible species (31K/T, 35E/K, 38D/E, 82T/M/N, and 353 K) were inconsistent with the comparatively large and diverse ACE2 dataset presented in this study (Shang et al., 2020a, 2020b; Wan et al., 2020). Therefore, based on the sequence alignment of ACE2 interacting residues, we propose six unique residues that could help in differentiation of susceptible from resistant species; Susceptible species (27T/I, 30D/E/Q, 82M/T/D/N, 326G/E/R/T, and 352G+353 K + 354G), resistant species (N326 + N330), and 352G+353 K + 354H/R/Q for reduced susceptibility to SARS-CoV-2 infection. Furthermore, the current approach differs substantially from previous *in silico* approaches in several aspects: 1) a phylogenetic analysis of 299 ACE2 orthologs from vertebrate species was conducted to demonstrate the existence of considerably high evolutionary diversity across the six vertebrate classes; 2) the ACE2-RBD protein complex models for 299 ACE2 orthologs from vertebrate species were generated by implementing a robust computational modeling approach and then selecting the best model for downstream processing; 3) the best metric (dissociation constant) was chosen for predicting the binding affinity of ACE2 with the spike protein of SARS-CoV-2 based on the benchmarking with the experimental data; 4) the different ACE2 isoforms were analyzed in a specific bat species to reveal their varied binding affinity to the spike protein of SARS-CoV-2. As a result, we believe that using these approaches allowed us to generate a more realistic representation of species at risk, and species that may serve as intermediate hosts. Though the findings revealed that a variety of vertebrate species could be potential SARS-CoV-2 intermediate hosts, this does not mean that the true intermediate host is one of them. The list can be whittled down even more by taking in account the animals' living conditions, particularly their proximity to human dwellings.

Increasing evidence suggests that the binding affinity of ACE2 orthologs from different bat species to the RBD of SARS-CoV-2 differed significantly, implying the existence of diversified bat species' susceptibility to SARS-CoV-2 (Boni et al., 2020; Latinne et al., 2020). The current study's results are moderately consistent with previous *in silico* studies for the susceptibility prediction of bat species to SARS-CoV-2, because of implementation of different approaches for the prediction of binding affinity. For instance, unlike previous *in silico* studies, this study's predictions of bat species susceptibility to SARS-CoV-2 are partially consistent with a recent functional experiment study that used 293 T cells expressing bat ACE2 orthologues to assess bat species susceptibility using pseudo-typed virus entry assay (Yan et al., 2021). However, this functional experiment contradicts another functional study (Zhou et al., 2021), which demonstrated that HeLa cells

expressing *Rhinolophus siniscus* ACE2 could serve as an entry receptor for SARS-CoV-2, whereas the first study did not. In comparison to hACE2, *Rhinolophus siniscus*, *Rhinolophus affinis*, and *Rhinolophus macrotis* have moderate binding affinity to the spike protein of SARS-CoV-2, according to the current study. These findings are in line with earlier functional experiments (Li et al., 2021; Liu et al., 2021b; Zhou et al., 2020). In summary, findings of the present study show a considerable consistency with the functional experiments than that of previous *in silico* approaches. Moreover, there were notable inconsistencies in RBD binding to ACE2 of *R. siniscus* in different studies (Liu et al., 2021a; Wu et al., 2020; Yan et al., 2021; Zhou et al., 2020). These inconsistencies are possible due to the presence of at least ten ACE2 isoforms in *R. siniscus*, which have varying predicted binding affinity to the spike protein of SARS-CoV-2. These findings suggested that the ACE2 binding affinity varied due to different isoforms in a specific bat species, which could have direct implications in the spillover events based on the distribution of these isoforms in different tissues. It would be of interest to delineate the tissue-specific expression of these ACE2 isoforms in future studies.

In conclusion, current findings support the bat origin of SARS-CoV-2 and the involvement of intermediate hosts in virus transmission to humans based on the predicted binding affinity of 299 ACE2 orthologs from various vertebrate species to the spike protein of SARS-CoV-2. It also elucidates the broad host range of SARS-CoV-2 and possibility of frequent cross-species transmission events. Furthermore, SARS-CoV-2 may be far more widespread than previously thought, highlighting the importance of intensive surveillance programmes aimed at identifying susceptible hosts, particularly those with the ability to cause zoonosis, in order to prevent future outbreaks.

Funding

This study was supported in part by Japan Society for the Promotion of Science (JSPS) KAKENHI grant (18H02395 to R.K. and K.Y.J.Z.), by the ICAR-National Agricultural Science Fund (NASF/ABA-8028/2020-21 to N.K.).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank RIKEN Advanced Center for Computing and Communication (ACCC) for computing resources on the Hokusai Big-Waterfall supercomputer. Finally, all of the authors express their appreciation to their respective Institutes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envres.2022.113303>.

References

- Battle, G.M., 2016. PDBePISA : Identifying and Interpreting the Likely Biological Assemblies of a Protein Structure.
- Bitencourt-Ferreira, G., de Azevedo Jr., W.F., 2019. Homology modeling of protein targets with MODELLER. *Methods Mol. Biol.* 2053, 231–249.
- Boni, M.F., et al., 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* 5, 1408–1417.
- Cool, K., et al., 2021. Infection and Transmission of SARS-CoV-2 and its Alpha Variant in Pregnant White-Tailed Deer. *bioRxiv*.
- Damas, J., et al., 2020. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* 117, 22311–22322.
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9 (8), 772.

- Heo, L., et al., 2013. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 41, W384–W388.
- Heo, L., et al., 2016. GalaxyRefineComplex: refinement of protein-protein complex model structures driven by interface repacking. *Sci. Rep.* 6, 32153.
- Hossain, M.G., Javed, A., Akter, S., Saha, S., 2021. SARS-CoV-2 host diversity: an update of natural infections and experimental evidence. *J. Microbiol. Immunol. Infect.* 54 (2), 175–181.
- Jumper, J., et al., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589.
- Katoh, K., et al., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20, 1160–1166.
- Kaushik, R., Jayaram, B., 2016. Structural difficulty index: a reliable measure for modelability of protein tertiary structures. *Protein Eng. Des. Sel.* 29 (9), 391–397.
- Kaushik, R., Zhang, K.Y.J., 2020. A protein sequence fitness function for identifying natural and nonnatural proteins. *Proteins* 88, 1271–1284.
- Kaushik, R., Singh, A., Jayaram, B., 2018. Where informatics lags chemistry leads. *Biochemistry* 57 (5), 503–505.
- Kaushik, R., Zhang, K.Y.J., 2022. ProFitFun: a protein tertiary structure fitness function for quantifying the accuracies of model structures. *Bioinformatics* 38 (2), 369–376. <https://doi.org/10.1093/bioinformatics/btab666>.
- Kuchipudi, S.V., et al., 2022. Multiple spillovers and onward transmission of SARS-CoV-2 in free-living and captive White-tailed deer. *PNAS* 119 (6), e2121644119. <https://doi.org/10.1073/pnas.2121644119>.
- Kumar, N., et al., 2021b. A novel consensus-based computational pipeline for rapid screening of antibody therapeutics for efficacy against SARS-CoV-2 variants of concern including omicron variant. *bioRxiv*, 2022.02.11.480177.
- Kumar, N., et al., 2021a. Evolutionary signatures governing the codon usage bias in coronaviruses and their implications for viruses infecting various bat species. *Viruses* 13, 1847. <https://doi.org/10.3390/v13091847>.
- Lan, J., et al., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220.
- Latinne, A., et al., 2020. Origin and cross-species transmission of bat coronaviruses in China. *Nat. Commun.* 11, 4235.
- Li, P., et al., 2021. The *Rhinolophus affinis* bat ACE2 and multiple animal orthologs are functional receptors for bat coronavirus RaTG13 and SARS-CoV-2. *Sci. Bull.* 66, 1215–1227.
- Liu, Z., et al., 2015. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics* 31, 405–412.
- Liu, K., et al., 2021a. Cross-species recognition of SARS-CoV-2 to bat ACE2. *Proc. Natl. Acad. Sci. U. S. A.* 118.
- Liu, Y., et al., 2021b. Functional and Genetic Analysis of Viral Receptor ACE2 Orthologs Reveals a Broad Potential Host Range of SARS-CoV-2, vol. 118. *Proc Natl Acad Sci U S A*.
- Lytras, S., et al., 2021. Exploring the Natural Origins of SARS-CoV-2 in the Light of Recombination. *bioRxiv*, 2021.01.22.427830.
- Malik, Y.S., et al., 2020. Coronavirus Disease Pandemic (COVID-19): Challenges and a Global Perspective, vol. 9. *Pathogens*.
- Malik, Y.S., et al., 2021a. How artificial intelligence may help the Covid-19 pandemic: pitfalls and lessons for the future. *Rev. Med. Virol.* 31, 1–11.
- Malik, Y.S., et al., 2021b. Evolutionary and codon usage preference insights into spike glycoprotein of SARS-CoV-2. *Briefings Bioinf.* 22 (2), 1006–1022.
- McGuffin, L.J., et al., 2021. ModFOLD8: accurate global and local quality estimates for 3D protein models. *Nucleic Acids Res.* 49, W425–W430.
- Mishra, A., et al., 2021. SARS-CoV-2 delta variant among asiatic lions, India. *Emerg. Infect. Dis.* 27, 2723–2725.
- O’Leary, N.A., et al., 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745.
- Olechnovič, K., Venclovas, Č., 2019. VoroMQA web server for assessing three-dimensional structures of proteins and protein complexes. *Nucleic Acids Res.* 47, W437–W442.
- Rodrigues, J.P., et al., 2020. Insights on cross-species transmission of SARS-CoV-2 from structural modeling. *PLoS Comput. Biol.* 16 (12), e1008449.
- Shang, J., et al., 2020a. Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* 117, 11727–11734.
- Shang, J., et al., 2020b. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221–224.
- Shi, J., et al., 2020. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* 368, 1016–1020.
- Singh, A., et al., 2016. ProTSAV: a protein tertiary structure analysis and validation server. *Biochim. Biophys. Acta* 1864, 11–19.
- Singh, D.K., et al., 2021. Responses to acute infection with SARS-CoV-2 in the lungs of rhesus macaques, baboons and marmosets. *Nat Microbiol* 6, 73–86.
- Stamatakis, A., 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Sun, J., et al., 2020. COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. *Trends Mol. Med.* 26, 483–495.
- Szilagyi, A., Zhang, Y., 2014. Template-based structure modeling of protein-protein interactions. *Curr. Opin. Struct. Biol.* 24, 10–23.
- Tamura, K., et al., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11030–11035.
- van Aart, A.E., et al., 2021. SARS-CoV-2 infection in cats and dogs in infected mink farms. *Transbound Emerg. Dis.* 00, 1–7. <https://doi.org/10.1111/tbed.14173>.
- Wan, Y., et al., 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* 94.
- WHO, 2021. W. H. O., Infection with SARS-CoV-2 in Animals, vol. 2021. WHO.
- Wu, L., et al., 2020. Broad host range of SARS-CoV-2 and the molecular basis for SARS-CoV-2 binding to cat ACE2. *Cell Discov.* 6, 68.
- Xue, L.C., et al., 2016. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* 32, 3676–3678.
- Yan, H., et al., 2021. ACE2 receptor usage reveals variation in susceptibility to SARS-CoV and SARS-CoV-2 infection among bat species. *Nat. Ecol. Evol.* 5, 600–608.
- Yang, J., et al., 2015. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* 12, 7–8.
- Yang, J., et al., 2020. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U. S. A.* 117, 1496–1503.
- Zhang, T., et al., 2020. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr. Biol.* 30, 1346–1351.e2.
- Zhang, Z., et al., 2021. The molecular basis for SARS-CoV-2 binding to dog ACE2. *Nat. Commun.* 12, 4195.
- Zhou, P., et al., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273.
- Zhou, H., et al., 2021. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell* 184, 4380–4391 e14.