Methods and Applications

Machine Learning Approach Effectively Predicts Binding Between SARS-CoV-2 Spike and ACE2 Across Mammalian Species — Worldwide, 2021

Yue Ma[1,&]; Yu Hu[1,2,&]; Binbin Xia[1]; Pei Du[1]; Lili Wu[1]; Mifang Liang[3]; Qian Chen[1,4]; Huan Yan[5]; George F. Gao[1]; Qihui Wang[1,#]; Jun Wang[1,#]

ABSTRACT

Introduction: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a recently emergent coronavirus of natural origin and caused the coronavirus disease (COVID-19) pandemic. The study of its natural origin and host range is of particular importance for source tracing, monitoring of this virus, and prevention of recurrent infections. One major approach is to test the binding ability of the viral receptor gene ACE2 from various hosts to SARS-CoV-2 spike protein, but it is time-consuming and labor-intensive to cover a large collection of species.

Methods: In this paper, we applied state-of-the-art machine learning approaches and created a pipeline reaching >87% accuracy in predicting binding between different ACE2 and SARS-CoV-2 spike.

Results: We further validated our prediction pipeline using 2 independent test sets involving >50 bat species and achieved >78% accuracy. A large-scale screening of 204 mammal species revealed 144 species (or 61%) were susceptible to SARS-CoV-2 infections, highlighting the importance of intensive monitoring and studies in mammalian species.

Discussion: In short, our study employed machine learning models to create an important tool for predicting potential hosts of SARS-CoV-2 and achieved the highest precision to our knowledge in experimental validation. This study also predicted that a wide range of mammals were capable of being infected by SARS-CoV-2.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused the ongoing pandemic of coronavirus disease (COVID-19) and has led to more than 229 million people infected and 4.7 million fatalities as of September 23, 2021 (https://covid19.who.int). Despite a large number of investigations on the biology and pathology of SARS-CoV-2, as well as treatment of COVID-19, the virus and pandemic still pose a tremendous threat to global health and stability. The natural origin of this virus has gained consensus among scientific communities but available evidence is still short of being conclusive. For instance, bats and pangolins have been proposed but disputes still remain (*1*), leaving room for misinformation and abuse. Identifying the host species susceptible to, including the source and intermediate species of, SARS-CoV-2 is still one of the central scientific objectives for COVID-19 research and will help provide information for monitoring and containing a potential viral reservoir as well as preventing reoccurring zoonosis as in the case of influenza viruses.

The entry of SARS-CoV-2 to host cells requires the binding of its spike protein and host angiotensin I converting enzyme 2 (ACE2), a process that underwent intense investigation. Blocking their binding with a list of neutralizing monoclonal antibodies (mAbs) has been demonstrated to effectively prevent viral entry to cells *in vitro* and *in vivo* (*2*), and several mAbs were approved for clinical treatment of COVID patients (*3*). Short peptide mimicking the structure of ACE2 region binding to the viral spike protein has also been developed, which binds the receptor binding domain (RBD) of spike proteins with picomole-level affinity and effectiveness in cell assays (*4*). Besides serving as a target for treatment, the ability of binding between the SARS-CoV-2 spike and the ACE2 from non-human species indicated the susceptibility of those species towards SARS-CoV-2 and, combined with ecological data and evolutionary evidence, might identify key species as probable origins and/or intermediate hosts of SARS-CoV-2.

Screening the binding between the ACE2 from large-scale collection of species and the SARS-CoV-2 spike protein thus is highly desired; however, in reality,

there are great constraints due to costs and time required for experimental verification. Alternatively, bioinformatic approaches capable of predicting binding between the two proteins with high precision are helpful in prioritizing species of interest and excluding very unlikely species, reducing the cost and time for this purpose. Based on sequence similarity in the ACE2 across species, Damas et al. (*5*) proposed a score predicting binding to the SARS-CoV-2 spikes; since then, many species' ACE2 have been tested, and retrospectively it is clear that the approach is limited in its precision. Namely, ACE2 from all bat species (36 in total in their prediction) were predicted to be "low" or "very low" in binding to the SARS-CoV-2 spike, but later experiments demonstrated that 20 species' ACE2 (55.56%) could bind to the viral spike (*6*). Alongside bats, 17 out of 29 (58.62%) other mammals with ACE2 genes considered unlikely to bind to the SARS-CoV-2 spike actually had ability to bind as well (Supplementary Table S1, available in http://weekly.chinacdc.cn/). Thus, the currently available bioinformatic approach has an extremely high false negative rate and is still short of precisely predicting binding between the SARS-CoV-2 spike protein and the ACE2 across species.

METHODS

We have therefore applied machine learning approaches to address the remaining challenges (see Supplementary Materials, available in http://weekly.chinacdc.cn/). Machine learning methods have the ability to combine diverse and complex data and automatically learn features for prediction, classification, and regressions. In biology, they have been successfully applied in establishing predictive and classification models using genomic features (*7*), metabolic markers (*8*), and many more (*9*). In our study, we selected five representative machine learning methods to perform classification (i.e., prediction of binding *vs.* non-binding), namely Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Adaboost (ADA), and Gradient Boosting Regression Tree (GBRT). For the single estimator we chose SVM and DT because they are suitable for small training sets. However, single estimators have a tendency to cause poor generalizability or robustness. To reduce this issue, we chose three additional ensemble methods (RF, ADA, and GBRT) for the construction of the prediction model.

The five models were further equipped with *a priori* information to establish a combined prediction pipeline. A study on the human ACE2 introduced mutations at 117 amino acid (AA) sites individually, whereas at each site the AA was mutated to all potential alternative AAs and the changes in affinity (relative to the wildtype ACE2) to that of SARS-CoV-2 have been experimentally examined, providing a quantitative reference data (*10*). Further, studies from Wang et al. (*11*) and Liu et al. (*12*) identified subsets of 24 and 20 AAs, respectively, in the human ACE2 as important sites for interaction with SARS-CoV-2 spike protein, which can be used as qualitative information to reduce model complexity and potential over-fitting. Based on reported experimental verifications of the ACE2 protein from 90 species (73 unique species, 27 from Wu et al. (*13*), 49 from Liu et al. (*12*). 14 are from our lab and currently being considered for independent publication), we aligned the ACE2 sequences of those species to the human ACE2 and extracted AAs to replace with log2 enrichment ratios for the 117, 24, and 20 sites as input data format (Figure 1A). We have deposited this pipeline and details of the method at https://github.com/mayuefine/Binding-prediction.

RESULTS

The training and the test set data contained 62 and 11 species, respectively, and the test set was set aside from the training process. In order to screen the models with a stable performance, we trained five models on three groups of site information (group 20, group 24, and group 117, each group containing 5 machine learning approaches). Finally, the predictions of the three groups were combined and a combination of six models with the highest precision was chosen as our prediction pipeline, out of a total of 408 combinations; this pipeline reached an in silico precision of circa 87.5% (Figure 1B) and was used for subsequent analysis. We used this pipeline to generate a prediction score for each ACE2 sequence, which was equal to the number of models predicting that it binded to the viral spike divided by the total number of models.

Bat species of the order Chiroptera were of highest interest for tracing the origin and studying the host range of SARS-CoV-2, as bat species harbor multiple coronavirus species including the SARS virus. One of the closest related strains of coronavirus to SARS-CoV-2, RaTG13, was found in horseshoe bats (*Rhinolophus affinis*) (*14*). Thus, we applied our pipeline and

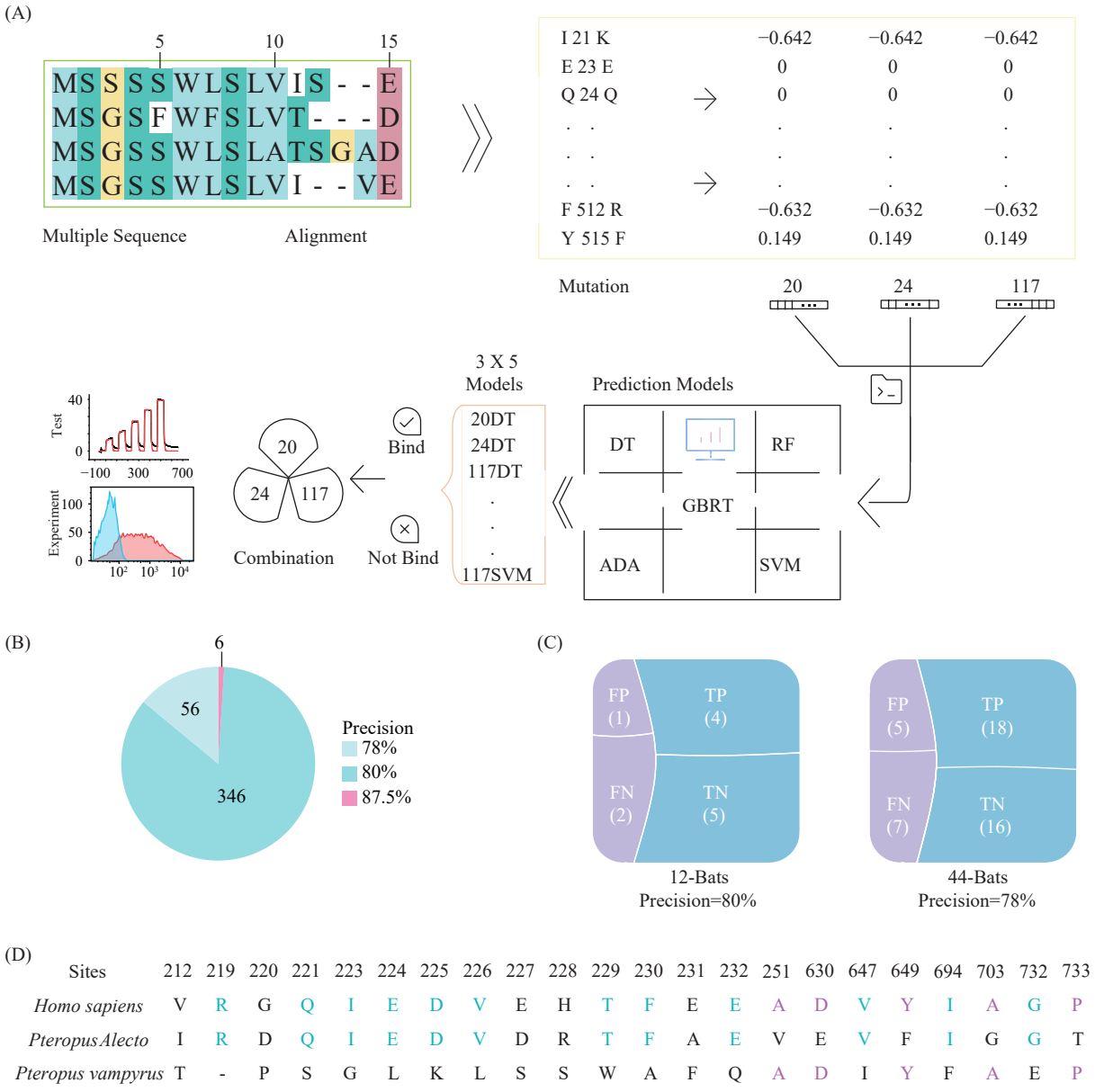


FIGURE 1. Overview of methodology and model performance of this study. (A) Schematic representation of the workflow; (B) The distribution of precision from all 408 potential combinations of models/input data; (C) Distribution of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) in our models' prediction in two experimentally validated datasets; (D) Distribution of different AAs in human (*Homo sapiens*) and two bat species (*P. alecto* and *P. vampyrus*).

Note: After sequencing alignment, information from chosen sites were transformed into vectors and fed to five different models, from which the optimal combination was chosen as pipeline and used to predict available ACE2 sequences. After the prediction, we selected some of the sequences for experimental validation. Figure 1B showed that multiple combinations reached high precision using our testing dataset. that we presume to influence binding between ACE2 and viral spike protein as well, based on the observation that the two bat species' ACE2 have different binding with the viral spike.

Abbreviations: ACE2=angiotensin I converting enzyme 2; DT=decision tree; RF=random forest; GBRT=gradient boosting regression tree; ADA=adaboost; SVM=support vector machine.

examined across bat species with ACE2 sequences available (59 in total), in which we predicted their ability to bind with SARS-CoV-2 spike proteins. We then tested the precision of our prediction in two experimentally validated datasets, in which ACE2 with predictions score >0.5 were considered likely to bind to the viral spike. We selected 12 bats' ACE2 and expressed the proteins, then confirmed with Surface Plasmon Resonance (SPR) and flow cytometry for the ability to bind the viral spike (Supplementary

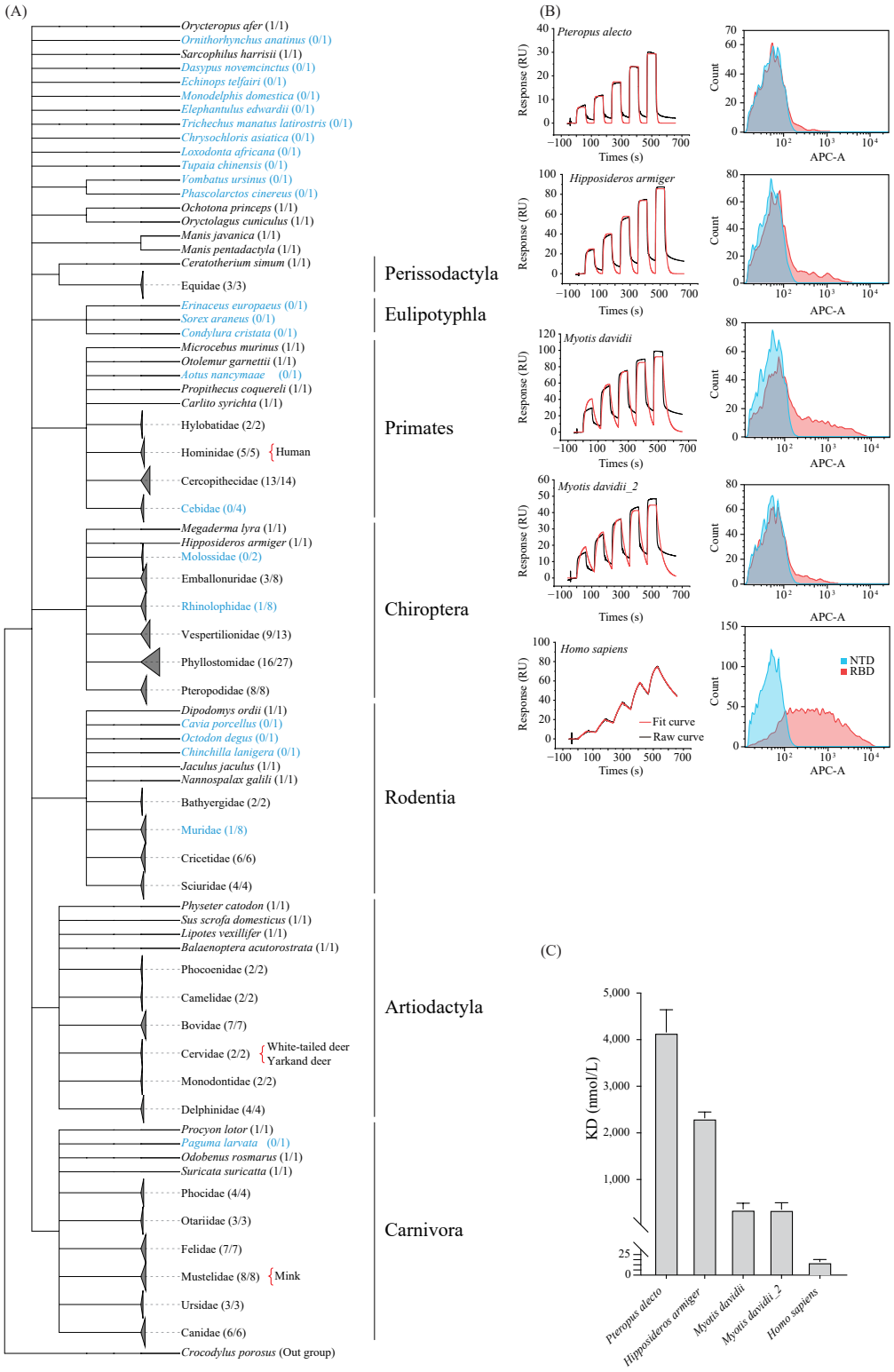


FIGURE 2. Prediction and validations of ACE2 across species in binding to SARS-CoV-2 spike. (A) The predicted range of species with ACE2 capable of binding to SARS-CoV-2; (B) SPR and flow cytometry validation for multiple species' ACE2 in binding to SARS-CoV-2 spike; (C) KD in nmol/L of the species shown in (B).
Note: For families with multiple species, the branch is collapsed and the proportion predicted to bind is shown in Figure 2A. Blue species/families are those predicted not to bind.
Abbreviations: ACE2=angiotensin I converting enzyme 2; SARS-CoV-2=severe acute respiratory syndrome coronavirus 2; SPR=surface plasmon resonance; KD=binding affinity.

Table S2, available in http://weekly.chinacdc.cn/). Overall, 4 of the 6 ACE2s predicted to bind to the SARS-CoV-2 spike were validated to bind to the viral spike (Figure 2B and Supplementary Figure S1, available in http://weekly.chinacdc.cn/), together with 5 ACE2s confirmed not to bind out of 6 ACE2s predicted to be so. Here we achieved a precision of 80% (Figure 1C). Then, using another dataset of 46 bat species by Yan et al. (*6*), after excluding the 2 sequences contained in our training set, we predicted the binding capacity and achieved 78.26% precision as shown in Figure 1C. Thus, our unified pipeline incorporating multiple machine learning models and different sets as input has the ability of confidently predicting binding between bat ACE2s and viral spikes.

It also drew our attention that during our validation, ACE2 sequences from *Pteropus alecto* and *Pteropus vampyrus* have identical AAs at all 117 sites we selected for input; however, *P. alecto* ACE2 could bind to the SARS-CoV-2 spike in our experimental system and *P. vampyrus* ACE2 had no detectable binding, suggesting additional AAs affected the binding capacity. We compared ACE2 sequences of these 2 species and identified in total 22 sites of difference between the 2. Of these sites, 16 are identical to human ACE2 (12 for *P. alecto and 4 for P. vampyrus*) (Figure 1D and Figure 2C). This comparison provided extra information that one or more of the AAs different between *P. alecto* and *P. vampyrus* and humans underly the differences in binding to the viral spike protein but have not been discovered in available studies. Closer investigations revealed that this set of AAs was not involved in binding with viral spike protein, thus their influences were indirect and likely affected by the ACE2 protein structurally or even by post-translation modifications including glycosylation.

Eventually, we refined our models incorporating the modified list of AAs as an input, and performed predictions on available ACE2 sequences from mammalian species (Supplementary Table S3, available in http://weekly.chinacdc.cn/, 204 in total and belonging to 69 families). This has resulted in the ACE2 of interest (likely to bind to the SARS-CoV-2 spike) from a total of 144 species, spread across 47 families (60.87%, Figure 2A). It is worth noting that the wide range of potential mammalian hosts agree with the emerging evidences of SARS-CoV-2 virus presence across mammals. Aside from 5 species of Hominidae (primates), ACE2s were predicted to bind to the viral spike protein in: 13 species of Cercopithecidae (old world monkeys), 8 species of Pteropodidae (old world fruit bats), 7 species of Felidae (cats), 7 species of Bovidae (ruminants), 7 species of Mustelidae (containing minks), 6 species of Canidae (dogs), 3 species of Equidae (horses), 6 species of Cricetidae (muroid rodents), 4 species of Sciuridae (squirrels), and 3 species of Ursidae (bears). Even in all 3 families of marine mammal, their ACE2s had high likelihood to bind to the SARS-CoV-2 spike (in all 4 species of Phocidae, 4 of Delphinidae and 3 of Otariidae, Figure 2B). Our prediction was supported by emerging reports that white-tailed deer (family Cervidae) were positive in antibodies against SARS-CoV-2 in 2021, which came in addition to reports of dogs, cats, and minks being viable hosts for this virus. In summary, based on ACE2 sequence features, our study suggested that SARS-CoV-2 has an extremely large range of potential hosts and indicates the importance of investigating wild animals for viral existence and monitoring its spread.

DISCUSSION

In conclusion, our study employed machine learning models suitable for analyzing sequence data, incorporated established functional data with multiple features extracted from sequences, and achieved high precision in predicting binding between ACE2s from difference species to the spike protein of SARS-CoV-2. The precision within the test data set was 87.5%, and in a total of 44 bat species, the group of mammals that attracted most concern, we achieved >78% precision as well, indicating that the model can be further expanded to predict susceptibility of more bat species once genomic sequences or ACE2 sequences become available (Supplementary Table S4, available in http://weekly.chinacdc.cn/). With the same approach we have also screened the available ACE2 sequences across a large range of mammals, in which we found that a large range of mammals requires attention. Our pipeline is capable of determining species of interest for tracing and analyzing species of interest to understand the potential origin of and transmission routes of SARS-CoV-2.

Our pipeline, in terms of performance, remains to be improved upon, provided that more accurate machine-learning models and/or more a priori information continues to emerge. First, limited by the number of experimentally validated sets and understanding on ACE2-spike interactions, we had to limit the total AAs in the ACE2 sequences for training and prediction, in which our result already indicated

contained critical information that is currently unavailable with regard to AAs in other part of the sequence, as in the case of *P. alecto* and *P. vampyrus*. In addition, the growing concerns amid the COVID-19 pandemic lie in the fast-emerging variants of SARS-CoV-2 strains, especially when mutations in ACE2-interacting AAs in the spike protein have already demonstrated changes in binding affinity to human ACE2s, whether they lead to host range changes and even broader transmission remain to be investigated.

In summary, our approach has the potential and will need to be expanded to analyze binding abilities of different SARS-CoV-2 variants and ACE2s to forecast the potential spread of this virus and identify priority species for monitoring.

Funding: The Strategic Priority Research Programs of the Chinese Academy of Sciences (XDB29020000), the National Natural Science Foundation of China (32041009) and Key R&D Program of Shandong Province (2020CXGC011305).

doi: 10.46234/ccdcw2021.235

[#] Corresponding authors: Jun Wang, junwang@im.ac.cn; Qihui Wang, wangqihui@im.ac.cn.

[1] CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; [2] School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China; [3] State Key Laboratory for Molecular Virology and Genetic Engineering, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China; [4] Institute of Physical Science and Information, Anhui University, Hefei, Anhui, China; [5] State Key Laboratory of Virology, Modern Virology Research Center, College of Life Sciences, Wuhan University, Wuhan, Hubei, China.
[&] Joint first authors.

Submitted: October 23, 2021; Accepted: November 01, 2021

REFERENCES

1. Wacharapluesadee S, Tan CW, Maneeorn P, Duengkae P, Zhu F, Joyjinda Y, et al. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. Nat Commun 2021;12(1):972. http://dx.doi.org/10.1038/s41467-021-21240-1.
2. Kreye J, Reincke SM, Kornau HC, Sánchez-Sendin E, Corman VM, Liu HJ, et al. A therapeutic non-self-reactive SARS-CoV-2 antibody protects from lung pathology in a COVID-19 hamster model. Cell 2020;183(4):1058 – 69.e19. http://dx.doi.org/10.1016/j.cell.2020.09.049.
3. U.S. Food & Drug Administration. Coronavirus (COVID-19) update: FDA authorizes monoclonal antibody for treatment of COVID-19. 2020. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-monoclonal-antibody-treatment-covid-19. [2021-8-22].
4. Cao LX, Goreshnik I, Coventry B, Case JB, Miller L, Kozodoy L, et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. Science 2020;370(6515):426 – 31. http://dx.doi.org/10.1126/science.abd9909.
5. Damas J, Hughes GM, Keough KC, Painter CA, Persky NS, Corbo M, et al. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. Proc Natl Acad Sci USA 2020;117(36):22311 – 22. http://dx.doi.org/10.1073/pnas.2010146117.
6. Yan H, Jiao HW, Liu QY, Zhang Z, Xiong Q, Wang BJ, et al. ACE2 receptor usage reveals variation in susceptibility to SARS-CoV and SARS-CoV-2 infection among bat species. Nat Ecol Evol 2021;5(5):600 – 8. http://dx.doi.org/10.1038/s41559-021-01407-1.
7. Huang SJ, Cai NG, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics 2018;15(1):41 – 51. http://dx.doi.org/10.21873/cgp.20063.
8. Liang L, Rasmussen MLH, Piening B, Shen XT, Chen SJ, Röst H, et al. Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women. Cell 2020;181(7):1680 – 92.e15. http://dx.doi.org/10.1016/j.cell.2020.05.002.
9. Toth R, Schiffmann H, Hube-Magg C, Büscheck F, Höflmayer D, Weidemann S, et al. Random forest-based modelling to detect biomarkers for prostate cancer progression. Clin Epigenetics 2019;11(1):148. http://dx.doi.org/10.1186/s13148-019-0736-8.
10. Chan KK, Dorosky D, Sharma P, Abbasi SA, Dye JM, Kranz DM, et al. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. Science 2020;369(6508):1261 – 5. http://dx.doi.org/10.1126/science.abc0870.
11. Wang QH, Zhang YF, Wu LL, Niu S, Song CL, Zhang ZY, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell 2020;181(4):894 – 904.e9. http://dx.doi.org/10.1016/j.cell.2020.03.045.
12. Liu YH, Hu GW, Wang YY, Ren WL, Zhao XM, Ji FS, et al. Functional and genetic analysis of viral receptor ACE2 orthologs reveals a broad potential host range of SARS-CoV-2. Proc Natl Acad Sci USA 2021;118(12):e2025373118. http://dx.doi.org/10.1073/pnas.2025373118.
13. Wu LL, Chen Q, Liu KF, Wang J, Han PC, Zhang YF, et al. Broad host range of SARS-CoV-2 and the molecular basis for SARS-CoV-2 binding to cat ACE2. Cell Discov 2020;6:68. http://dx.doi.org/10.1038/s41421-020-00210-9.
14. Liu KF, Pan XQ, Li LJ, Yu F, Zheng AQ, Du P, et al. Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in humans and other species. Cell 2021;184(13):3438 – 51.e10. http://dx.doi.org/10.1016/j.cell.2021.05.031.

Supplementary Material

Method

Data Collection

The 73 species angiotensin I converting enzyme 2 (ACE2) sequences for constructing predictive models and evaluation were collected from published articles (*1*–*2*) and unpublished data. Overall, 11 sequences from these 73 were randomly selected as test dataset for model evaluation and were not involved in model training.

The sequences of mammalian ACE2 for prediction were downloaded as of September 22, 2020 with a total of 294 ACE2 sequences of mammalian species from 23 orders being gathered. We performed multiple sequence alignment on collection of 294 sequences with human ACE2 sequence, using software CLUSTAL (version 2.1, Conway Institute, UCD, Dublin, Ireland, parameter "complete multiple alignment") (*3*), in which sequences with more than 10 consecutive amino acid missing in the head 100 sites were excluded from the subsequent analysis, resulting in 272 ACE2 sequences (204 unique species).

Model Construction and Evaluation

We selected key amino acid sites and used the log2 enrichment ratios values from Chan et al. to label the amino acids for each ACE2 sequence (*4*), with 20, 24, and 117 sites selected from Liu et al. (*1*), Wang et al. (*2*), and Chan et al. (*4*), respectively. The sequences screened for these three sites were divided into a training dataset and a test dataset with an 8∶2 ratio and used for training and testing of the model, respectively. As for prediction models, we used five different methods to train three different collections of sites, including support vector machine (SVM), Decision Tree, Random Forest, AdaBoost and Gradient Boosting, resulting in 15 models of input data/methods. After hundreds of epochs of training, random combinations of the 15 models were evaluated based on precision (Precision=TP/(TP+FP), where TP: True Positive, FP: False Positive). We selected six model combinations for ACE2 sequences prediction in the subsequent analysis and set the prediction score (Prediction Score=Pn/Mn), where Pn indicated the number of one sequence that was predicted to have binding ability and Mn was the total number of models used for prediction. The threshold value for the prediction score was set to 0.5, i.e., a prediction score ≥0.5 was considered to have the ability to bind with Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The 272 sequences were also screened for sites for binding ability prediction.

Model construction and prediction were carried out based on the scikit-learn module in the Python3 (version 0.22.2, Python Software Foundation, Fredericksburg, VA, USA). The functions used for model training were "svm," "DecisionTreeClassifier," "RandomForestClassifier," "AdaBoostClassifier," and "GradientBoostingClassifier." The parameters used for SVM were: gamma='scale'; class_weight={0:2}; for decision tree classifier were default parameters; for random forest classifier were the following: n_estimators=600, oob_score=True, n_jobs=-1, class_weight={0:2}; for Ada boost classifier were the following: base_estimator=DecisionTreeClassifier (max_depth=2), n_estimators=500; and for gradient boosting classifier were the following: n_estimators=100, learning_rate=1.0, max_depth=1, random_state=0. All details were also available in our github depository.

ACE2 Sequence Acquisition and Gene cloning

Twelve bat orthologs were randomly selected from the test sets. The full-length coding sequences (accession numbers are shown in Supplementary Table S2) of these orthologs were synthesized and cloned into the pEGFP-N1 vector for flow cytometry (FACS). The extracellular domain of these ACE2 orthologs was fused with the Fc domain of mouse IgG (mFc) and cloned into the pCAGGS expression vector for surface plasmon resonance (SPR).

Protein Expression and Purification

The SARS-CoV-2 receptor-binding domain (RBD) and SARS-CoV-2 N-terminal domain (NTD) proteins used for flow cytometry and SPR were expressed and purified from the supernatants of HEK293F cells culture as described in our previous work (*5*). Proteins were stored in a PBS buffer [1.8 mmol/L KH_2PO_4, 10 mmol/L Na_2HPO_4 (pH 7.4), 137 mmol/L NaCl, 2.7 mmol/L KCl] buffer. The indicated pCAGGS plasmids were transiently transfected into HEK293T cells (ATCC CRL-3216). Supernatants containing mFc-tagged ACE2 proteins were collected and concentrated at 48 h post-transfection.

SUPPLEMENTARY TABLE S1. Binding ability of various mammalian ACE2, including published experimental results, prediction from Damas et al. (*7*) using our method.

Mammals	Species	Common name	Experiment results	Previous prediction results	Our prediction	Accession Number
Bats	*Anoura caudifer*	Tailed tailless bat	Binding	Very low	Binding	GCA_004027475.1
	Artibeus jamaicensis	Jamaican fruit-eating bat	Binding	Very low	Binding	GCA_004027435.1
	Carollia perspicillata	Seba's short-tailed bat	Binding	Very low	Not bind	GCA_004027735.1
	Desmodus rotundus	Common vampire bat	Binding	Very low	Not bind	XP_024425698.1
	Eidolon helvum	Straw-colored fruit bat	Binding	Low	Binding	GCA_000465285.1
	Eonycteris spelaea	Lesser dawn bat	Binding	Low	Binding	GCA_003508835.1
	Macroglossus sobrinus	Long-tongued fruit bat	Binding	Very low	Binding	GCA_004027375.1
	Megaderma lyra	Indian false vampire	Binding	Low	Binding	MT515624
	Micronycteris hirsuta	Hairy big-eared bat	Binding	Very low	Not bind	GCA_004026765.1
	Miniopterus schreibersii	Schreibers' long-fingered bat	Binding	Very low	Binding	GCA_004026525.1
	Mormoops blainvillei	Antillean ghost-faced bat	Binding	Very low	Not bind	GCA_004026545.1
	Myotis brandtii	Brandt's bat	Binding	Very low	Binding	XP_014399780.1
	Myotis davidii	David's myotis	Binding	Very low	Binding	XP_006775273.1
	Myotis lucifugus	Little brown bat	Binding	Very low	Binding	XP_023609437.1
	Myotis myotis	Greater mouse-eared bat	Binding	Very low	Binding	https://vgp.github.io/genomeark/Myotis_myotis
	Noctilio leporinus	Greater bulldog bat	Binding	Very low	Binding	GCA_004026585.1
	Pipistrellus pipistrellus	Common pipistrelle	Binding	Very low	Not bind	GCA_004026625.1
	Pteropus alecto	Black flying fox	Binding	Low	Binding	XP_006911709.1
	Rousettus aegyptiacus	Egyptian rousette	Binding	Low	Binding	XP_015974412.1
	Tadarida brasiliensis	Brazilian free-tailed bat	Binding	Very low	Not bind	GCA_004025005.1
Other Mammals	*Ailuropoda melanoleuca*	Giant panda	Binding	Low	Binding	XP_002930657.1
	Camelus ferus	Wild Bactrian camel	Binding	Low	Binding	XP_006194263.1
	Ceratotherium simum simum	Southern white rhinoceros	Binding	Low	Binding	XP_004435206.1
	Equus caballus	Horse	Binding	Low	Binding	XP_001490241.1
	Peromyscus leucopus	White-footed mouse	Binding	Low	Binding	XP_028743609.1
	Rousettus aegyptiacus	Egyptian rousette	Binding	Low	Binding	XP_015974412.1
	Sus scrofa	Pig	Binding	Low	Binding	NP_001116542.1
	Ursus arctos horribilis	Grizzly bear	Binding	Low	Binding	XP_026333865.1
	Vulpes vulpes	Red fox	Binding	Low	Binding	XP_025842512.1
	Callorhinus ursinus	Northern fur seal	Binding	Very low	Binding	XP_025713397.1
	Eumetopias jubatus	Steller sea lion	Binding	Very low	Binding	XP_027970822.1
	Jaculus jaculus	lesser Egyptian jerboa	Binding	Very low	Binding	XP_004671523.1
	Manis javanica	Malayan pangolin	Binding	Very low	Binding	XP_017505746.1
	Mustela erminea	Stoat	Binding	Very low	Binding	XP_032187677.1
	Myotis lucifugus	Little brown bat	Binding	Very low	Binding	XP_023609437.1
	Neomonachus schauinslandi	Hawaiian monk seal	Binding	Very low	Binding	XP_021536480.1
	Zalophus californianus	California sea lion	Binding	Very low	Binding	XP_027465353.1

Abbreviations: ACE2=angiotensin I converting enzyme 2.

SUPPLEMENTARY TABLE S2. Results of binding between ACE2 from 12 bat species and SARS-CoV-2 spike performed in our study.

Species	KD (nmol/L)	Prediction score	Accession number
Pteropus alecto	4,163.47±479.62	1.00	XP_006911709.1
Pteropus vampyrus	–*	1.00	XP_011361275.1
Hipposideros armiger	2,323.89±124.60	0.70	XP_019522936.1
Myotis davidii	369.03±126.37	0.79	XP_015426919.1
Myotis davidii	361.33±144.51	0.79	XP_006775273.1
Rhinolophus pearsonii	–	0.20	ABU54053.1
Megaderma lyra	735.58±121.91	0.47	QKE49998.1
Molossus molossus	–	0.33	KAF6491643.1
Pipistrellus abramus	–	0.11	ACT66266.1
Rhinolophus landeri	3,635.83±156.31	0.01	ALJ94034.1
Scotophilus dinganii	–	0.22	QJF77809.1
Tadarida brasiliensis	–	0.17	QLF98520.1
Homo sapiens	13.28±2.06	1.00	NP_00135844.1

Note: Prediction score of >0.5 is considered to be able to bind SARS-CoV-2 spike.
Abbreviations: ACE2=angiotensin I converting enzyme 2; SARS-CoV-2=severe acute respiratory syndrome coronavirus 2; KD=binding affinity.
* No detected affinity.

Flow Cytometry Analysis

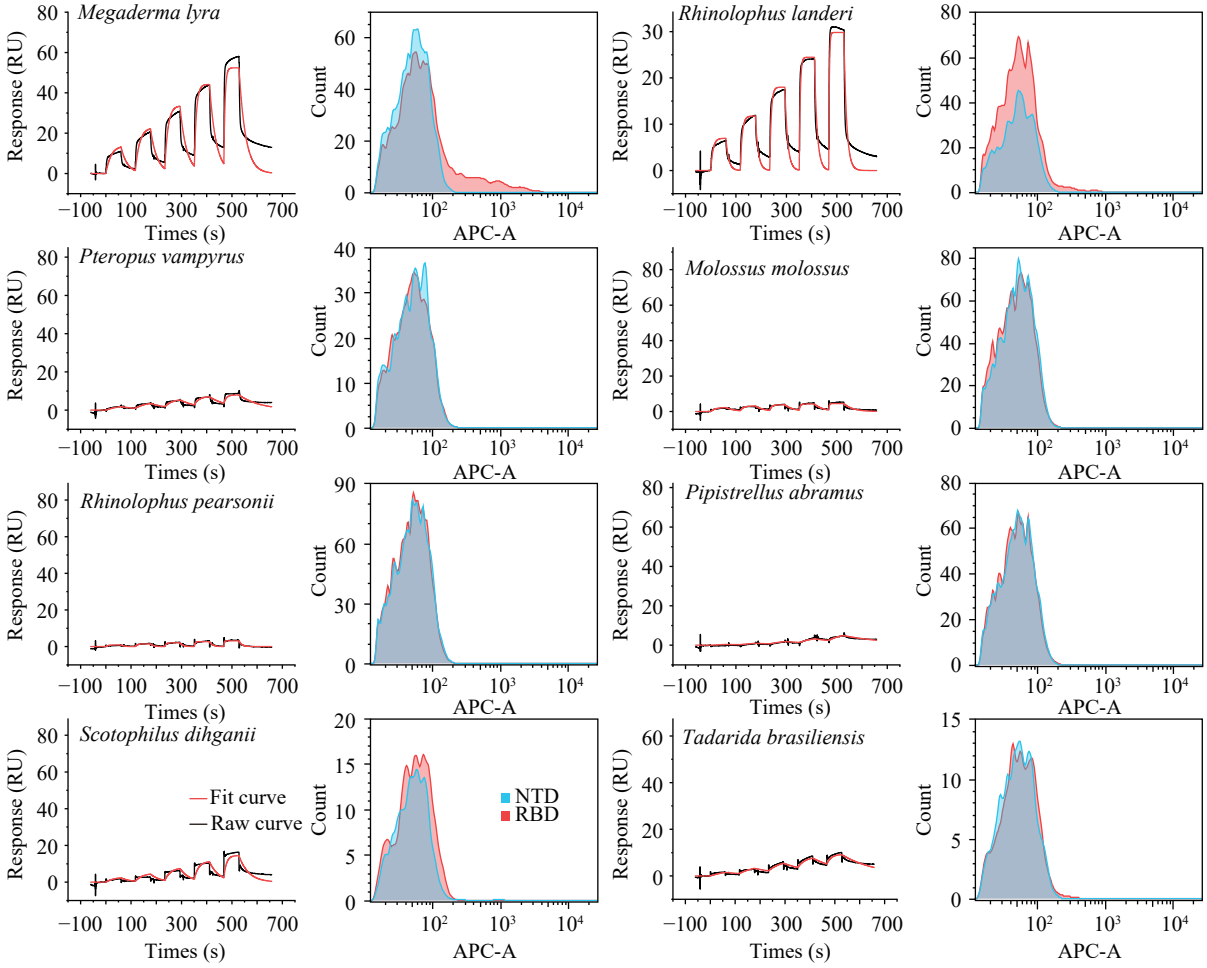
To test the binding between each of the 12 ACE2s and SARS-CoV-2 RBD, the 12 bat ACE2s fused with eGFP were expressed on the cell surface by transfecting each of the 12 pEGFP-N1-ACE2s plasmids into BHK21 cells (ATCC, ATCC CCL-10) using PEI (Alfa). Cell culture was replaced with fresh media (DMEM with 10% FBS, Gibco) 4–6 h post-transfection. After 48 h, cells were collected and resuspended in PBS. Then, 2×10^5 cells were incubated with the histidine tagged test proteins (SARS-CoV-2 RBD, SARS-CoV-2 NTD) at a concentration of 10 μg/mL at 37 °C for 30 min. Cells were then washed three times in PBS and stained with anti-His/APC antibodies (1:500, Miltenyi Biotec, AB_2751870) for 30 min at 37 °C. Flow cytometry (FACS) data were acquired on a BD FACSCanto (BD Biosciences, Franklin Lakes, NJ, USA) and analyzed using FlowJo V10 software (TreeStar Inc., Ashland, OR, USA), with results shown in Supplementary Figure S1.

SPR Analysis

We tested the binding affinities between the mFc-tagged ACE2s and SARS-CoV-2 RBD or SARS-CoV RBD proteins by SPR using a BIAcore 8K (GE Healthcare) carried out at 25 °C in single-cycle mode. The PBST buffer (1.8 mmol/L KH_2PO_4, 10 mmol/L Na_2HPO_4 (pH 7.4), 137 mmol/L NaCl, 2.7 mmol/L KCl, and 0.05% (v/v) Tween 20) was used as the running buffer. The CM5 biosensor chip was first immobilized with anti-mIgG antibody (ZSGB-BIO, ZF-0513) as previously described. (*1*) The supernatants containing mFc-tagged ACE2s were injected and captured by the antibody immobilized on the CM5 chip at approximately 300-600 response units. The serially diluted SARS-CoV-2 RBD protein flowed over the chip surface, with another channel set as control. The chip was regenerated using pH 1.7 glycine after each reaction. The equilibrium dissociation constants (binding affinity, KD) for each pair of interaction were calculated with BIAcore_8K evaluation software (GE Healthcare, Chicago, IL, USA) by fitting to a 1∶1 Langmuir binding model. Data were analyzed using OriginLab (Origin 2018, OriginLab Corporation, Northampton, MA, USA).

Phylogenetic Tree

The phylogenetic tree was constructed by uploading the species names from 272 sequences into NCBI Taxonomy Common Tree (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/). The visualization of the phylogenetic tree was based on iTol (*6*).



SUPPLEMENTARY FIGURE S1. SPR and flow cytometry validation for multiple species' ACE2.
Abbreviations: ACE2=angiotensin I converting enzyme 2; SPR=surface plasmon resonance; RU=response unit; NTD=N-terminal domain; RBD=receptor-binding domain.

SUPPLEMENTARY TABLE S3. Prediction of the binding capacity of collected mammalian ACE2 to SARS-CoV-2.

Species	Common name	Prediction scores	Data availability
Hylobates moloch	Silvery gibbon	1.00	XP_032612508.1
Phocoena sinus	Vaquita	1.00	XP_032476001.1
Globicephala melas	Long-finned pilot whale	1.00	XP_030703991.1
Lynx canadensis	Canada lynx	1.00	XP_030160839.1
Monodon monoceros	Narwhal	1.00	XP_029095804.1
Peromyscus leucopus	White-footed mouse	1.00	XP_028743609.1
Balaenoptera acutorostrata scammoni	Common minke whale	1.00	XP_028020351.1
Eumetopias jubatus	Steller sea lion	1.00	XP_027970822.1
Marmota flaviventris	Yellow-bellied marmot	1.00	XP_027802308.1
Zalophus californianus	California sea lion	1.00	XP_027465353.1
Bos indicus x Bos taurus	Hybrid cattle	1.00	XP_027389729.1
Bos indicus x Bos taurus	Hybrid cattle	1.00	XP_027389727.1
Cricetulus griseus	Chinese hamster	1.00	XP_027288607.1
Lagenorhynchus obliquidens	Pacific white-sided dolphin	1.00	XP_026951598.1
Acinonyx jubatus	Cheetah	1.00	XP_026910297.1
Ursus arctos horribilis	Grizzly bear	1.00	XP_026333865.1
Vulpes vulpes	Red fox	1.00	XP_025842512.1
Puma concolor	Puma	1.00	XP_025790417.1
Callorhinus ursinus	Northern fur seal	1.00	XP_025713397.1
Canis lupus dingo	Dingo	1.00	XP_025292925.1
Theropithecus gelada	Gelada	1.00	XP_025227847.1
Neophocaena asiaeorientalis asiaeorientalis	Yangtze finless	1.00	XP_024599894.1
Pongo abelii	Sumatran orangutan	1.00	XP_024096013.1
Physeter catodon	Sperm whale	1.00	XP_023971279.1
Felis catus	Domestic cat	1.00	XP_023104564.1
Piliocolobus tephrosceles	Ugandan red colobus	1.00	XP_023054821.1
Delphinapterus leucas	Beluga whale	1.00	XP_022418360.1
Papio anubis	Olive baboon	1.00	XP_021788732.1
Neomonachus schauinslandi	Hawaiian monk seal	1.00	XP_021536486.1
Neomonachus schauinslandi	Hawaiian monk seal	1.00	XP_021536480.1
Sus scrofa	Pig	1.00	XP_020935034.1
Sus scrofa	Pig	1.00	XP_020935033.1
Odocoileus virginianus texanus	White-tailed deer	1.00	XP_020768965.1
Bos indicus	Bos taurus indicus	1.00	XP_019811720.1
Bos indicus	Bos taurus indicus	1.00	XP_019811719.1
Tursiops truncatus	Common bottlenose dolphin	1.00	XP_019781177.1
Panthera pardus	Leopard	1.00	XP_019273508.1
Gorilla gorilla gorilla	Western lowland gorilla	1.00	XP_018874749.1
Manis javanica	Malayan pangolin	1.00	XP_017505746.1
Pan troglodytes	Chimpanzee	1.00	XP_016798469.1
Pan troglodytes	Chimpanzee	1.00	XP_016798468.1
Rousettus aegyptiacus	Egyptian rousette	1.00	XP_015974412.1

Continued

Species	Common name	Prediction scores	Data availability
Marmota marmota marmota	Alpine marmot	1.00	XP_015343540.1
Propithecus coquereli	Coquerel's sifaka	1.00	XP_012494185.1
Ovis aries	Sheep	1.00	XP_011961657.1
Cercocebus atys	Sooty mangabey	1.00	XP_011891198.1
Mandrillus leucophaeus	Drill	1.00	XP_011850923.1
Colobus angolensis palliatus	Angola colobus	1.00	XP_011795654.1
Macaca nemestrina	Pig-tailed macaque	1.00	XP_011733505.1
Homo sapiens	Human	1.00	XP_011543854.1
Homo sapiens	Human	1.00	XP_011543853.1
Homo sapiens	Human	1.00	XP_011543851.1
Pteropus vampyrus	Large flying fox	1.00	XP_011361275.1
Rhinopithecus roxellana	Golden snub-nosed monkey	1.00	XP_010364367.2
Pan paniscus	Pygmy chimpanzee	1.00	XP_008972437.1
Pan paniscus	Pygmy chimpanzee	1.00	XP_008972428.1
Nannospalax galili	Upper galilee mountains blind mole rat	1.00	XP_008839098.1
Ursus maritimus	Polar bear	1.00	XP_008694637.1
Chlorocebus sabaeus	Green monkey	1.00	XP_007989304.1
Lipotes vexillifer	Yangtze River dolphin	1.00	XP_007466389.1
Panthera tigris altaica	Amur tiger	1.00	XP_007090142.1
Peromyscus maniculatus bairdii	Prairie deer mouse	1.00	XP_006973269.1
Pteropus alecto	Black flying fox	1.00	XP_006911709.1
Bubalus bubalis	Water buffalo	1.00	XP_006041602.1
Bos mutus	Wild yak	1.00	XP_005903173.1
Capra hircus	Goat	1.00	XP_005701129.2
Canis lupus familiaris	Dog	1.00	XP_005641049.1
Macaca fascicularis	Crab-eating macaque	1.00	XP_005593094.1
Ictidomys tridecemlineatus	Thirteen-lined ground squirrel	1.00	XP_005316051.3
Bos taurus	Cattle	1.00	XP_005228486.1
Bos taurus	Cattle	1.00	XP_005228485.1
Mesocricetus auratus	Golden hamster	1.00	XP_005074266.1
Heterocephalus glaber	Naked mole-rat	1.00	XP_004866157.1
Ochotona princeps	American pika	1.00	XP_004597549.2
Ceratotherium simum simum	Southern white rhinoceros	1.00	XP_004435206.1
Odobenus rosmarus divergens	Pacific walrus	1.00	XP_004415448.1
Orcinus orca	Killer whale	1.00	XP_004269705.1
Cricetulus griseus	Chinese hamster	1.00	XP_003503283.1
Nomascus leucogenys	Northern white-cheeked gibbon	1.00	XP_003261132.2
Ailuropoda melanoleuca	Giant panda	1.00	XP_002930657.1
Oryctolagus cuniculus	Rabbit	1.00	XP_002719891.1
Chrysocyon brachyurus	Maned wolf	1.00	QNC68917.1
Neofelis diardi	Sunda clouded leopard	1.00	QNC68916.1
Speothos venaticus	Bush dog	1.00	QNC68915.1
Manis pentadactyla	Chinese pangolin	1.00	QLH93383.1

Continued

Species	Common name	Prediction scores	Data availability
Dobsonia viridis	Greenish naked-backed fruit bat	1.00	QJF77815.1
Syconycteris australis	Southern blossom bat	1.00	QJF77811.1
Epomophorus wahlbergi	Wahlberg's epauletted fruit bat	1.00	QJF77792.1
Homo sapiens	Human	1.00	NP_068576.1
Homo sapiens	Human	1.00	NP_001358344.1
Capra hircus	Goat	1.00	NP_001277036.1
Canis lupus familiaris	Dog	1.00	NP_001158732.1
Macaca mulatta	Rhesus monkey	1.00	NP_001129168.1
Pongo abelii	Sumatran orangutan	1.00	NP_001124604.1
Sus scrofa	Pig	1.00	NP_001116542.1
Felis catus	Domestic cat	1.00	NP_001034545.1
Bos taurus	Cattle	1.00	NP_001019673.2
Rousettus leschenaultii	Leschenault's rousette	1.00	BAF50705.1
Rousettus leschenaultii	Leschenault's rousette	1.00	ADJ19219.1
Mesocricetus auratus	Golden hamster	1.00	ACT66278.1
Felis catus	Domestic cat	1.00	ACT66276.1
Oryctolagus cuniculus	Rabbit	1.00	ACT66271.1
Sus scrofa domesticus	Domestic pig	1.00	ACT66265.1
Rhinolophus ferrumequinum	Greater horseshoe bat	1.00	ACM45790.1
Macaca mulatta	Rhesus monkey	1.00	ACI04576.1
Macaca mulatta	Rhesus monkey	1.00	ACI04571.1
Macaca mulatta	Rhesus monkey	1.00	ACI04570.1
Macaca mulatta	Rhesus monkey	1.00	ACI04569.1
Macaca mulatta	Rhesus monkey	1.00	ACI04568.1
Macaca mulatta	Rhesus monkey	1.00	ACI04567.1
Macaca mulatta	Rhesus monkey	1.00	ACI04566.1
Macaca mulatta	Rhesus monkey	1.00	ACI04564.1
Macaca mulatta	Rhesus monkey	1.00	ACI04563.1
Macaca mulatta	Rhesus monkey	1.00	ACI04562.1
Macaca mulatta	Rhesus monkey	1.00	ACI04560.1
Macaca mulatta	Rhesus monkey	1.00	ACI04559.1
Macaca mulatta	Rhesus monkey	1.00	ACI04557.1
Macaca mulatta	Rhesus monkey	1.00	ACI04556.1
Macaca mulatta	Rhesus monkey	1.00	ACI04555.1
Macaca mulatta	Rhesus monkey	1.00	ACI04554.1
Macaca mulatta	Rhesus monkey	1.00	ACI04553.1
Macaca mulatta	Rhesus monkey	1.00	ACI04552.1
Nyctereutes procyonoides	Raccoon dog	1.00	ABW16956.1
Chlorocebus aethiops	Grivet	1.00	AAY57872.1
Camelus ferus	Wild bactrian camel	0.99	XP_006194263.1
Jaculus jaculus	Lesser Egyptian jerboa	0.99	XP_004671523.1
Mirounga leonina	Southern elephant seal	0.97	XP_034852450.1
Trachypithecus francoisi	Francois's langur	0.97	XP_033056809.1

Continued

Species	Common name	Prediction scores	Data availability
Macaca mulatta	Rhesus monkey	0.97	ACI04573.1
Equus asinus	African wild ass	0.93	XP_014713133.1
Equus przewalskii	Przewalski's horse	0.93	XP_008542995.1
Orycteropus afer afer	Aardvark	0.93	XP_007951028.1
Microtus ochrogaster	Prairie vole	0.93	XP_005358818.1
Equus caballus	Horse	0.93	XP_001490241.1
Neovison vison	American mink	0.93	QPL12211.1
Arctonyx collaris	Hog badger	0.93	QLF98526.1
Cynopterus sphinx	Indian short-nosed fruit bat	0.93	QKE49997.1
Uroderma bilobatum	Tent-building bat	0.93	QJF77842.1
Platyrrhinus vittatus	Greater broad-nosed bat	0.93	QJF77835.1
Platyrrhinus helleri	Heller's broad-nosed bat	0.93	QJF77834.1
Cynopterus sphinx	Indian short-nosed fruit bat	0.93	QJF77831.1
Chiroderma villosum	Hairy big-eyed bat	0.93	QJF77830.1
Chiroderma salvini	Salvin's big-eyed bat	0.93	QJF77829.1
Artibeus phaeotis	Dwarf fruit-eating bat	0.93	QJF77823.1
Artibeus lituratus	Great fruit-eating bat	0.93	QJF77822.1
Artibeus jamaicensis	Jamaican fruit-eating bat	0.93	QJF77821.1
Phodopus campbelli	Campbell's desert hamster	0.93	ACT66274.1
Fukomys damarensis	Damara mole-rat	0.91	XP_010643477.1
Cervus hanglu yarkandensis	Yarkand deer	0.88	KAF4027296.1
Urocitellus parryii	Arctic ground squirrel	0.87	XP_026252506.1
Urocitellus parryii	Arctic ground squirrel	0.87	XP_026252505.1
Myotis lucifugus	Little brown bat	0.87	XP_023609439.1
Myotis lucifugus	Little brown bat	0.87	XP_023609437.1
Myotis brandtii	Brandt's bat	0.87	XP_014399783.1
Myotis brandtii	Brandt's bat	0.87	XP_014399782.1
Myotis brandtii	Brandt's bat	0.87	XP_014399780.1
Taphozous melanopogon	Black-bearded Tomb Bat	0.87	QJF77841.1
Taphozous theobaldi	Theobald's tomb bat	0.87	QJF77840.1
Artibeus glaucus watsoni	–*	0.87	QJF77824.1
Artibeus hartii	Little fruit-eating bat	0.86	QJF77832.1
Scotophilus kuhlii	Lesser asiatic yellow house bat	0.82	QJF77810.1
Scotophilus dinganii	Yellow-bellied house bat	0.82	QJF77809.1
Procyon lotor	Raccoon	0.80	BAE72462.1
Myotis davidii	David's myotis	0.79	XP_015426919.1
Myotis davidii	David's myotis	0.79	XP_006775273.1
Tylonycteris robustula	Greater bamboo bat	0.79	QJF77813.1
Sarcophilus harrisii	Tasmanian devil	0.77	XP_031814825.1
Dipodomys ordii	Ord's kangaroo rat	0.77	XP_012887573.1
Dipodomys ordii	Ord's kangaroo rat	0.77	XP_012887572.1
Vicugna pacos	Alpaca	0.77	XP_006212709.1
Phoca vitulina	Harbor seal	0.73	XP_032245506.1

Continued

Species	Common name	Prediction scores	Data availability
Eptesicus fuscus	Big brown bat	0.73	XP_027986092.1
Eptesicus fuscus	Big brown bat	0.73	XP_008153150.1
Megaderma lyra	Indian false vampire	0.73	QKE49998.1
Hipposideros armiger	Great roundleaf bat	0.70	XP_019522936.1
Glossophaga commissarisi	Commissaris's long-tongued bat	0.70	QJF77793.1
Microcebus murinus	Gray mouse lemur	0.69	XP_020140826.1
Carlito syrichta	Philippine tarsier	0.69	XP_008062810.1
Anoura geoffroyi	Geoffroy's tailless bat	0.67	QJF77820.1
Suricata suricatta	Meerkat	0.66	XP_029786256.1
Anoura cultrata	Handley's tailless bat	0.66	QJF77819.1
Kerivoula pellucida	Clear-winged woolly bat	0.63	QJF77795.1
Grammomys surdaster	Grammomys	0.60	XP_028617961.1
Coleura afra	African sheath-tailed bat	0.59	QJF77826.1
Neoromicia nanus	Banana bat	0.58	QJF77804.1
Otolemur garnettii	Small-eared galago	0.56	XP_003791912.1
Hylonycteris underwoodi	Underwood's long-tongued bat	0.54	QJF77833.1
Lontra canadensis	Northern American river otter	0.53	XP_032736028.1
Enhydra lutris kenyoni	Sea otter	0.53	XP_022374079.1
Enhydra lutris kenyoni	Sea otter	0.53	XP_022374078.1
Mustela lutreola	European mink	0.53	QNC68911.1
Melogale moschata	Chinese ferret-badger	0.53	QLF98521.1
Mustela putorius furo	Domestic ferret	0.53	NP_001297119.1
Mustela erminea	Stoat	0.53	XP_032187677.1
Halichoerus grypus	Gray seal	0.52	XP_035963182.1
Sturnira parvidens	−	0.51	QJF77839.1
Sturnira ludovici	Highland Yellow-shouldered Bat	0.51	QJF77838.1
Sturnira hondurensis	−	0.51	QJF77837.1
Arvicanthis niloticus	African grass rat	0.49	XP_034341939.1
Mastomys coucha	Southern multimammate mouse	0.47	XP_031226742.1
Mus pahari	Shrew mouse	0.47	XP_021043935.1
Antrozous pallidus	Pallid bat	0.47	QJF77789.1
Carollia perspicillata	Seba's short-tailed bat	0.44	QJF77828.1
Carollia castanea	Chestnut short-tailed bat	0.44	QJF77827.1
Chinchilla lanigera	Long-tailed chinchilla	0.42	XP_013362428.1
Chinchilla lanigera	Long-tailed chinchilla	0.42	NP_001269290.1
Rhinolophus sinicus	Chinese rufous horseshoe bat	0.41	ACT66275.1
Mus caroli	Ryukyu mouse	0.40	XP_021009138.1
Vampyrum spectrum	Spectral bat	0.38	QJF77843.1
Carollia sowelli	Sowell's short-tailed bat	0.38	QJF77814.1
Loxodonta africana	African savanna elephant	0.37	XP_023410960.1
Tadarida brasiliensis	Brazilian free-tailed bat	0.37	QLF98520.1
Sorex araneus	European shrew	0.36	XP_004612266.1
Elephantulus edwardii	Cape elephant shrew	0.34	XP_006892457.1

Continued

Species	Common name	Prediction scores	Data availability
Rattus norvegicus	Norway rat	0.34	NP_001012006.1
Molossus molossus	Pallas's mastiff bat	0.33	KAF6491643.1
Aeorestes cinereus	Hoary bat	0.33	QJF77796.1
Cavia porcellus	Domestic guinea pig	0.32	ACT66270.1
Rhinolophus sinicus	Chinese rufous horseshoe bat	0.31	ADN93475.1
Micronycteris schmidtorum	Schmidts's big-eared bat	0.26	QJF77799.1
Lonchophylla robusta	Orange nectar bat	0.24	QJF77797.1
Glossophaga soricina	Pallas's long-tongued bat	0.24	QJF77794.1
Miniopterus natalensis	Natal long-fingered bat	0.22	XP_016058453.1
Tupaia chinensis	Chinese tree shrew	0.22	XP_006164754.1
Dasypus novemcinctus	Nine-banded armadillo	0.22	XP_004449124.1
Rhinolophus macrotis	Big-eared horseshoe bat	0.22	ADN93471.1
Sapajus apella	Tufted capuchin	0.20	XP_032141854.1
Cebus capucinus imitator	White headed capuchin	0.20	XP_017367865.1
Condylura cristata	Star-nosed mole	0.20	XP_012585871.1
Aotus nancymaae	Ma's night monkey	0.20	XP_012290105.1
Saimiri boliviensis boliviensis	Bolivian squirrel monkey	0.20	XP_010334925.1
Callithrix jacchus	White-tufted-ear marmoset	0.20	XP_008987241.1
Emballonura alecto	Small Asian sheath-tailed bat	0.20	QJF77816.1
Mus musculus	House mouse	0.20	NP_001123985.1
Mus musculus	House mouse	0.20	ACT66269.1
Rhinolophus pearsonii	Pearson's horseshoe bat	0.20	ABU54053.1
Vombatus ursinus	Common wombat	0.19	XP_027691156.1
Phascolarctos cinereus	Koala	0.19	XP_020863153.1
Grammomys surdaster	Grammomys	0.18	XP_028636273.1
Mirounga leonina	Southern elephant seal	0.17	XP_034882212.1
Phoca vitulina	Harbor seal	0.17	XP_032285427.1
Octodon degus	Degu	0.17	XP_023569950.1
Monodelphis domestica	Gray short-tailed opossum	0.17	XP_007500942.1
Monodelphis domestica	Gray short-tailed opossum	0.17	XP_007500941.1
Monodelphis domestica	Gray short-tailed opossum	0.17	XP_007500935.1
Paguma larvata	Masked palm civet	0.17	Q56NL1.1
Phyllostomus discolor	Pale spear-nosed bat	0.16	XP_028378317.1
Desmodus rotundus	Common vampire bat	0.16	XP_024425698.1
Rhynchonycteris naso	Proboscis bat	0.16	QJF77807.1
Octodon degus	Degu	0.14	XP_023575315.1
Trichechus manatus latirostris	Florida manatee	0.13	XP_004386381.1
Rhinolophus alcyone	Halcyon horseshoe bat	0.13	ALJ94035.1
Cavia porcellus	Domestic guinea pig	0.12	XP_023417808.1
Pipistrellus abramus	Japanese house bat	0.11	ACT66266.1
Theropithecus gelada	Gelada	0.10	XP_025218729.1
Chrysochloris asiatica	Cape golden mole	0.10	XP_006833624.1
Micronycteris hirsuta	Hairy big-eared bat	0.10	QJF77798.1

Continued

Species	Common name	Prediction scores	Data availability
Crocodylus porosus	Australian saltwater crocodile	0.09	XP_019384827.1
Crocodylus porosus	Australian saltwater crocodile	0.09	XP_019384826.1
Ornithorhynchus anatinus	Platypus	0.08	XP_001515597.2
Micronycteris microtis	Common big-eared bat	0.08	QJF77800.1
Chrysochloris asiatica	Cape golden mole	0.07	XP_006835673.1
Centronycteris centralis	Thomas's shaggy bat	0.07	QJF77790.1
Rhinolophus sinicus	Chinese rufous horseshoe bat	0.07	ADN93472.1
Balantiopteryx plicata	Gray sac-winged rat	0.03	QJF77825.1
Echinops telfairi	Small madagascar hedgehog	0.01	XP_004710002.1
Rhinolophus landeri	Lander's horseshoe bat	0.01	ALJ94034.1
Rhinolophus pusillus	Least horseshoe bat	0.01	ADN93477.1
Erinaceus europaeus	Western European hedgehog	0.00	XP_007538670.1
Saccopteryx bilineata	Greater sac-winged bat	0.00	QJF77808.1
Rhinolophus ferrumequinum	Greater horseshoe bat	0.00	BAH02663.1
Rhinolophus sinicus	Chinese rufous horseshoe bat	0.00	AGZ48803.1
Rhinolophus ferrumequinum	Greater horseshoe bat	0.00	ADN93470.1

Note: >0.5 prediction score in our analysis indicate bindiSilvery gibbon2 and SARS-CoV-2 spike.
Abbreviations: ACE2=angiotensin I converting enzyme 2; SARS-CoV-2=severe acute respiratory syndrome coronavirus 2.
* No common name.

SUPPLEMENTARY TABLE S4. Results of our predictions and the results of the experimental validation from Yan et al. (*8*).

Sequnece_name	Experiment results	Our prediction sore	Accession number
Rousettus_aegyptiacus	Binding	1.00	XM_016118926.1
Pteropus_alecto	Binding	1.00	XM_006911647.1
Pteropus_giganteus	Binding	1.00	GCA_902729225.1
Eidolon_helvum	Binding	1.00	GCA_000465285.1
Eonycteris_spelaea	Binding	1.00	GCA_003508835.1
Macroglossus_sobrinus	Binding	1.00	GCA_004027375.1
Cynopterus_sphinx	Not bind	0.93	MT515623
Cynopterus_brachyotis	Not bind	0.93	GCA_009793145.1
Rhinolophus_pearsonii	Not bind	0.09	MT515622
Hipposideros_armiger	Binding	0.70	XM_019667391.1
Hipposideros_galeritus	Not bind	0.72	GCA_004027415.1
Hipposideros_pratti	Not bind	0.70	MT515621
Megaderma_lyra	Binding	0.73	MT515624
Noctilio_leporinus	Binding	0.77	GCA_004026585.1
Taphozous_melanopogon	Binding	0.87	MT952961
Anoura_caudifer	Binding	0.72	GCA_004027475.1
Trachops_cirrhosus	Binding	0.26	MT952962
Vampyram_spectrum	Not bind	0.31	MT952963
Tonatia_saurophila	Not bind	0.14	GCA_004024845.1
Phyllostomus_discolor	Not bind	0.16	XM_028522516.1
Carollia_perspicillata	Binding	0.44	GCA_004027735.1
Micronycteris_hirsuta	Binding	0.11	GCA_004026765.1
Sturnira_hondurensis	Binding	0.44	GWHAAZA00000000
Artibeus_jamaicensis	Binding	0.93	GCA_004027435.1
Desmodus_rotundus	Binding	0.16	XM_024569930.1
Pteronotus_parnellii	Not bind	0.36	GCA_000465405.1
Mormoops_blainvillei	Binding	0.38	GCA_004026545.1
Pteronotus_davyi	Not bind	0.33	MT952964
Tadarida_brasiliensis	Not bind	0.37	GCA_004025005.1
Molossus_molossus	Not bind	0.49	https://vgp.github.io/genomeark/Molossus_molossus
Miniopterus_schreibersii	Binding	0.76	GCA_004026525.1
Miniopterus_natalensis	Not bind	0.22	GCA_001595765.1
Eptesicus_fuscus	Not bind	0.73	XM_008154928.2
Aeorestes_cinereus	Not bind	0.33	GCA_011751065.1
Pipistrellus_pipistrellus	Binding	0.36	GCA_004026625.1
Lasiurus_borealis	Not bind	0.29	GCA_004026805.1
Pipistrellus_kuhlii	Not bind	0.32	https://vgp.github.io/genomeark/Pipistrellus_kuhlii
Antrozous_pallidus	Binding	0.86	GCA_007922775.1
Nycticeius_humeralis	Not bind	0.47	GCA_007922795.1
Murina_feae	Not bind	0.48	GCA_004026665.1
Myotis_myotis	Binding	0.72	https://vgp.github.io/genomeark/Myotis_myotis
Myotis_davidii	Binding	0.79	XM_006775210.2
Myotis_brandtii	Binding	0.87	XM_014544294.1
Myotis_lucifugus	Binding	0.87	XM_023753669.1

Note: >0.5 prediction score in our analysis indicate binding between ACE2 and SARS-CoV-2 spike.
Abbreviations: ACE2=angiotensin I converting enzyme 2; SARS-CoV-2=severe acute respiratory syndrome coronavirus 2.

REFERENCES

1. Wang QH, Zhang YF, Wu LL, Niu S, Song CL, Zhang ZY, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell 2020;181(4):894 – 904.e9. http://dx.doi.org/10.1016/j.cell.2020.03.045.
2. Liu YH, Hu GW, Wang YY, Ren WL, Zhao XM, Ji FS, et al. Functional and genetic analysis of viral receptor ACE2 orthologs reveals a broad potential host range of SARS-CoV-2. Proc Natl Acad Sci USA 2021;118(12):e2025373118. http://dx.doi.org/10.1073/pnas.2025373118.
3. Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 1988;73(1):237 – 44. http://dx.doi.org/10.1016/0378-1119(88)90330-7.
4. Chan KK, Dorosky D, Sharma P, Abbasi SA, Dye JM, Kranz DM, et al. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. Science 2020;369(6508):1261 – 5. http://dx.doi.org/10.1126/science.abc0870.
5. Niu S, Wang J, Bai B, Wu LL, Zheng AQ, Chen Q, et al. Molecular basis of cross-species ACE2 interactions with SARS-CoV-2-like viruses of pangolin origin. EMBO J 2021;40(16):e107786. http://dx.doi.org/10.15252/embj.2021107786.
6. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 2021;49(W1):W293 – 6. http://dx.doi.org/10.1093/nar/gkab301.
7. Damas J, Hughes GM, Keough KC, Painter CA, Persky NS, Corbo M, et al. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. Proc Natl Acad Sci USA 2020;117(36):22311 – 22. http://dx.doi.org/10.1073/pnas.2010146117.
8. Yan H, Jiao HW, Liu QY, Zhang Z, Xiong Q, Wang BJ, et al. ACE2 receptor usage reveals variation in susceptibility to SARS-CoV and SARS-CoV-2 infection among bat species. Nat Ecol Evol 2021;5(5):600 – 8. http://dx.doi.org/10.1038/s41559-021-01407-1.