Predicting the zoonotic capacity of mammals to transmit SARS-CoV-2

Ilya R. Fischhoff[1]*, Adrian A. Castellanos[1]*, João P.G.L.M. Rodrigues[2], Arvind Varsani[3,4], Barbara A. Han[1] ✝

Affiliations:

[1] Cary Institute of Ecosystem Studies. Box AB Millbrook, NY 12545, USA
[2] Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
[3] The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA
[4] Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Rondebosch, 7700, Cape Town, South Africa

* contributed equally
✝ corresponding author

Email addresses:

Ilya Fischhoff: fischhoff@gmail.com
Adrian A. Castellanos: castellanosa@caryinstitute.org
João P.G.L.M. Rodrigues: joaor@stanford.edu
Arvind Varsani: arvind.varsani@asu.edu
Barbara A. Han: hanb@caryinstitute.org

Keywords: coronavirus, COVID-19, hosts, reservoirs, zoonotic, spillover, spillback, susceptibility, machine learning, homology modelling, ACE2

Summary

Back and forth transmission of SARS-CoV-2 between humans and animals has the potential to create wild reservoirs of virus that can endanger both long-term control of COVID-19 in people, and vulnerable animal populations that are particularly susceptible to lethal disease. In the near term, SARS-CoV-2 virus variants arising in newly established animal hosts could escape immunity conferred by current human vaccines. In the long-term, animal reservoirs of SARS-CoV-2 increase the overall risk of disease resurgence, making global disease control unlikely. Predicting potential animal host species is key to targeting critical surveillance as well as lab experiments testing susceptibility of potential hosts. A major bottleneck to predicting animal hosts is a paucity of molecular information about the structure of ACE2 across species, a key cellular receptor required for viral cell entry that is highly conserved across thousands of animal species. We overcome this bottleneck by combining 3D modeling of virus and host cell protein interactions with machine learning analysis of species' ecological and biological traits, enabling predictions about the zoonotic capacity of SARS-CoV-2 for over 5,000 mammals — an order of magnitude more species than previously possible. High accuracy model predictions are strongly corroborated by available and emerging *in vivo* empirical studies. We also identify numerous common mammal species whose predicted zoonotic capacity and close proximity to humans may facilitate spillover and spillback transmission of SARS-CoV-2. Our results reveal high priority areas of geographic overlap between global COVID-19 hotspots and potential new mammal hosts of SARS-CoV-2. Predictive modeling integrating data across multiple biological scales offers a conceptual advance that may expand our predictive capacity for zoonotic viruses with similarly unknown and potentially broad host ranges.

Introduction

The ongoing COVID-19 pandemic has surpassed 3.4 million deaths globally as of 17 May 2021 [1,2]. Like previous pandemics in recorded history, COVID-19 originated from the spillover of a zoonotic pathogen, SARS-CoV-2, a betacoronavirus originating from an unknown animal host [3–6]. The broad host range of SARS-CoV-2 is due in part to its use of a highly conserved cell surface receptor to enter host cells, the angiotensin-converting enzyme 2 receptor (ACE2) [7] found in all major vertebrate groups [8].

The ubiquity of ACE2 coupled with the high prevalence of SARS-CoV-2 in the global human population explains multiple observed *spillback* infections since the emergence of SARS-CoV-2 in 2019. In spillback infection, human hosts transmit SARS-CoV-2 virus to cause infection in non-human animals. In addition to threatening wildlife and domestic animals, repeated spillback infections may lead to the establishment of new animal hosts from which SARS-CoV-2 can continue to pose a risk of *secondary spillover* infection to humans through bridge hosts (e.g., [9]) or newly established enzootic reservoirs. Indeed, this risk has already been realized in Denmark [10] and The Netherlands, where SARS-CoV-2 spilled back from humans to farmed mink (*Neovison vison*) with secondary spillover of a SARS-CoV-2 variant from mink back to humans [11]. A major concern in such secondary spillover events is the appearance of a mutant strain [11,12] affecting host range [13] or leading to increased transmissibility in humans [14,15] (but see [16,17]). Preliminary evidence shows that the mink-derived variant exhibits moderately reduced sensitivity to neutralizing antibodies [10], raising concerns that humans may eventually experience more virulent infections from spillback variants, and that vaccines may become less efficient at conferring immunity to variants [18]. Conversely, human-derived variants pose additional spillback risks to animals. For example, in contrast to previous infection trials [19], two new human variants are now confirmed to have overcome the species barrier to infect lab mice (*Mus musculus*)[20].

Spillback infections from humans to animals are already occurring worldwide. In addition to secondary spillover infections from mink farms, SARS-CoV-2 has been found for the first time in wild and escaped mink in multiple states in the United States, with viral sequences identical to SARS-CoV-2 in nearby farmed mink [21–23]. A variety of pets, domesticated animals, zoo animals, and wildlife have also been documented as new hosts of SARS-CoV-2 (Table 1). The global scale of human infections and the increasing range of known hosts observed for SARS-CoV-2 suggest that SARS-CoV-2 has the capacity to establish novel enzootic infection cycles in animals. In response, recent computational studies make predictions about the susceptibility of animal species to SARS-CoV-2 [13,24–32]. These studies compare sequences of ACE2 orthologs among species (*sequence-based* studies), or model the structure of the viral spike protein bound to ACE2 orthologs (*structure-based* studies), yielding a wide range of predictions with varying degrees of agreement with laboratory animal experiments (Figure 1).

Table 1. Species with confirmed suitability for SARS-CoV-2 infection from natural infections or *in vivo* experiments. Asterisks reference species with infection status from preprints (not yet peer-reviewed). Some species (e.g, dogs) with natural infection studies also have *in vivo* experimental studies.

Species	Susceptibility	Study type	Location	References
Cow (*Bos taurus*)	Yes	*In vivo* experiment	Lab	[33]
Dog (*Canis lupus familiaris*)	Yes	Natural infection	Multiple countries	[34–38]
African green monkey (*Chlorocebus aethiops*)	Yes	*In vivo* experiment	Lab	[39]
Big brown bat (*Eptesicus fuscus*)	No	*In vivo* experiment	Lab	[40]
Cat (*Felis catus*)	Yes	Natural infection	Multiple countries	[34,36,37,41]
Gorilla (*Gorilla gorilla*)	Yes	Natural infection	USA, Zoo	[42]
Crab-eating macaque (*Macaca fascicularis*)	Yes	*In vivo* experiment	Lab	[43]
Rhesus macaque (*Macaca mulatta*)	Yes	*In vivo* experiment	Lab	[44]
Golden hamster (*Mesocricetus auratus*)	Yes	*In vivo* experiment	Lab	[45]
House mouse (*Mus musculus*)	No	*In vivo* experiment	Lab	[19] (but see [20])
Ferret (*Mustela putorius furo*)	Yes	*In vivo* experiment	Lab	[38]
American mink (*Neovison vison*)	Yes	Natural infection	Multiple countries	[36,37,46]
Raccoon dog (*Nyctereutes procyonoides*)	Yes	*In vivo* experiment	Lab	[47]
European rabbit (*Oryctolagus cuniculus*)	Yes	*In vivo* experiment	Lab	[48]
Lion (*Panthera leo*)	Yes	Natural infection	Multiple countries, Zoos	[37,49]
Tiger (*Panthera tigris*)	Yes	Natural infection	USA and Sweden, Zoos	[36,37,49,50]

Deer mouse (*Peromyscus maniculatus*)*	Yes	*In vivo* experiment	Lab	51,52
Cougar (*Puma concolor*)	Yes	Natural infection	South Africa, Zoo	37
Egyptian fruit bat (*Rousettus aegyptiacus*)	Yes	*In vivo* experiment	Lab	53
Pig (*Sus scrofa*)	No	*In vivo* experiment	Lab	38,53
Northern treeshrew (*Tupaia belangeri*)	Yes	*In vivo* experiment	Lab	54
Snow leopard (*Uncia uncia*)	Yes	Natural infection	USA, Zoo	55
Bank vole (*Clethrionomys glareolus*)	Yes	*In vivo* experiment	Lab	56
Asian small-clawed otter (*Aonyx cinereus*)	Yes	Natural infection	USA, Zoo	37,57
White-tailed deer (Odocoileus virginianus)	Yes	In vivo experiment	Lab	58

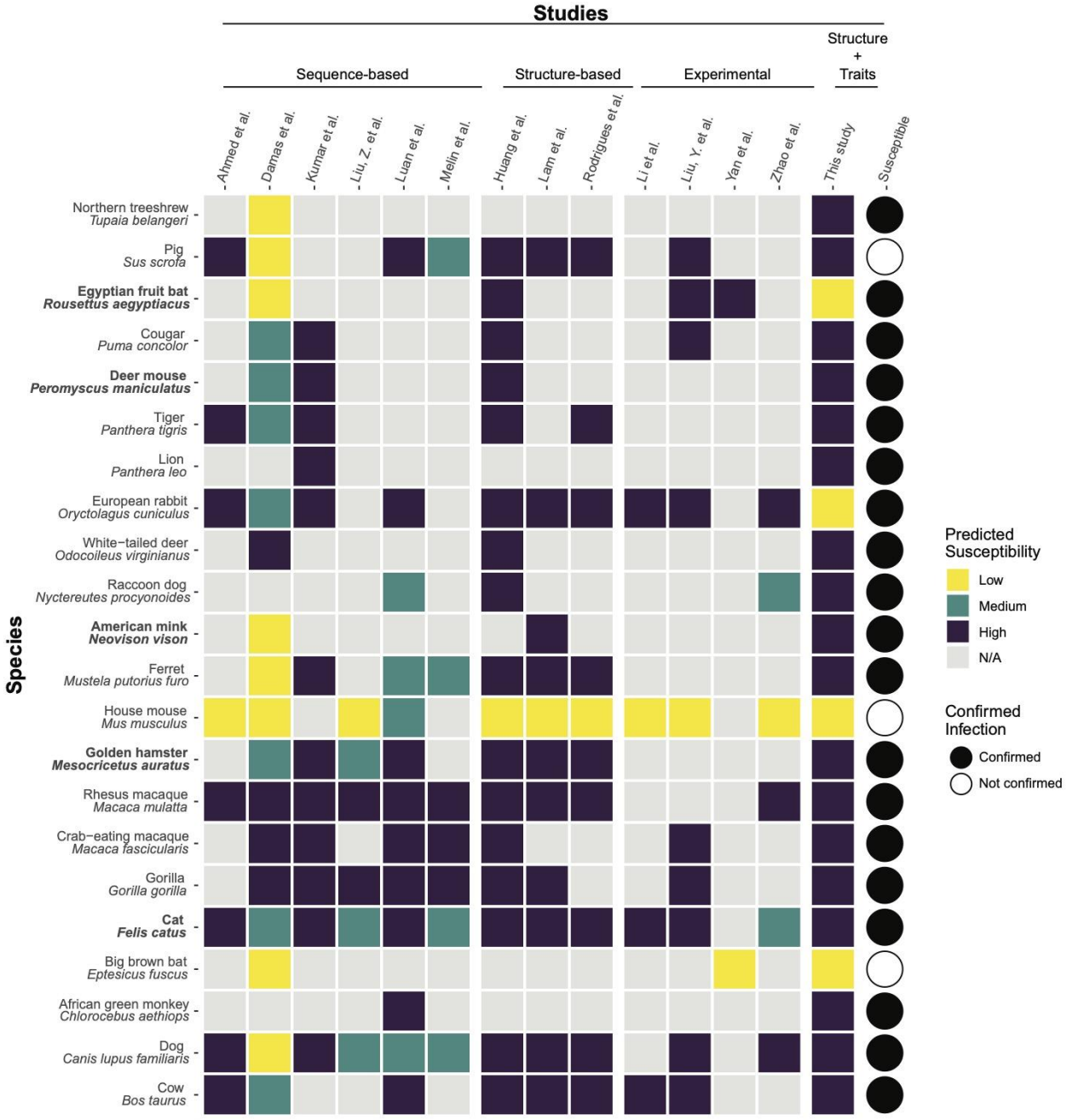

Figure 1. A heatmap summarizing predicted susceptibility to SARS-CoV-2 for species with confirmed infection from *in vivo* experimental studies or from documented natural infections. Studies that make predictions about species susceptibility are shown on the x-axis, organized by method of prediction (those relying on ACE2 sequences, estimating binding strength using three dimensional structures, or laboratory experiments). Predictions about zoonotic capacity from this study are listed in the second to last column, with high and low categories determined by zoonotic capacity observed in *Felis catus*. Confirmed infections for species along the y-axis are summarized in [59] and are depicted as a series of filled or unfilled circles. Bolded species have been experimentally confirmed to transmit SARS-CoV-2 to naive conspecifics. Species predictions range from warmer colors (yellow: low susceptibility or zoonotic capacity for SARS-CoV-2) to cooler colors (purple: high susceptibility or zoonotic capacity). See

Supplementary Methods (https://doi.org/10.25390/caryinstitute.c.5293339) for detailed methods about study categorization.

Sequence-based studies

Studies predicting host susceptibility based on amino acid sequence similarity between human (hACE2) and non-human ACE2 assume that a high degree of similarity correlates with stronger viral binding, especially at amino acid residues where hACE2 interacts with the SARS-CoV-2 spike glycoprotein. For some species, such as rhesus macaques [60], these qualitative predictions are borne out by *in vivo* studies (Figure 1), but predictions from these methods do not consistently match real-world outcomes. For example, sequence similarity predicted weak viral binding for minks and ferrets, which have all been confirmed as highly susceptible, with minks capable of onward transmission [11,32,38] (Figure 1). These mismatches to experimental or real-world outcomes may arise in part because protein three-dimensional structure, the main determinant of protein function, is robust to changes in amino acid sequence [61,62]. As such, details about the interaction between host ACE2 and the viral spike protein are not well captured by sequence-based studies.

Structure-based studies

Modeling the three-dimensional structure of protein-protein complexes addresses some of the limitations of sequence-based approaches, and has proven useful to predict how different ACE2 orthologs bind to the SARS-CoV-2 viral spike protein receptor-binding domain (RBD) [13,24]. These studies can also be useful for identifying particular ACE2 amino acid residues essential for a productive interaction with the viral RBD, thus improving predictive models of susceptibility through structure-based inference [13]. These studies leveraged known structures of the hACE2 receptor bound to the SARS-CoV-2 RBD and used powerful simulation methods to predict how variation across different ACE2 orthologs affects binding with the viral RBD. While these approaches successfully predicted strong binding for species that have been infected (e.g. domestic cat, tiger, dog, and ferret) and weak binding for species in which experimental infections failed (e.g. chicken, duck[38], mouse[19]), the results are also not consistently supported by experiments. For instance, while guinea pig ACE2 scored favorably among susceptible species in one of the studies [13], this ortholog was shown experimentally not to bind to the SARS-CoV-2 RBD [63].

Although structural modeling has produced the most accurate results to date, all currently available approaches for predicting the host range of SARS-CoV-2 are fundamentally constrained by the availability and quality of ACE2 sequences. ACE2 is ubiquitous across chordates, likely because of its role in several highly conserved physiological pathways [64]. Because it is so highly conserved, the majority of mammal species (>6,000 species) are likely to have ACE2 receptors, but there are many fewer sequences available from which to make predictions using existing modeling methods (~300 species). The functional importance of the ACE2 receptor suggests that it has evolved in association with other intrinsic organismal traits that are more easily observed and for which data are available for far more species. These suites of correlated organismal traits may provide a robust statistical proxy that can be leveraged to predict suitable hosts for SARS-CoV-2. Previous trait-based analyses applied

statistical (machine) learning techniques to accurately distinguish the zoonotic capacity of various organisms [65–67], and predict likely hosts for particular groups of related viruses [68,69], predictions which have subsequently been validated through independent laboratory and field investigations (e.g., [70,71]).

Here, we combine molecular structural modeling of viral binding with machine learning of species-level traits to generate predictions about species' zoonotic capacity for SARS-CoV-2 virus across 5,400 mammal species, expanding our predictive capacity by an order of magnitude (Figure 2). Crucially, this integrated approach enables predictions for species whose ACE2 sequences are not available by leveraging information available from viral binding dynamics and biological traits of potential hosts. In our workflow (Figure 2), we first carry out structural modeling to quantify the binding strength of SARS-CoV-2 RBD for vertebrate species using published ACE2 amino acid sequences [72]. We then collate species traits and apply machine learning to predict the zoonotic capacity for 5,400 mammal species, determined by a conservative threshold of susceptibility and onward transmission capacity of SARS-CoV-2 reported by *in vivo* studies and applied to our structural modeling results. Because COVID-19 is, at this time, primarily a disease affecting humans, spillback infection of SARS-CoV-2 from humans to animals is the most likely mode by which new host species will become established. Among mammal species with the highest predicted zoonotic capacity for SARS-CoV-2, we identify a subset of species for which the threat of spillback infection appears greatest due to geographic overlaps and opportunities for contact with humans in areas of high SARS-CoV-2 prevalence globally. Our predictions contribute to a critical interdisciplinary and iterative process between computational modeling, field surveillance, and laboratory experiments that is necessary for improving zoonotic risk quantification, and to better inform next steps toward the prevention of enzootic SARS-CoV-2 transmission and spread (Figure 2).

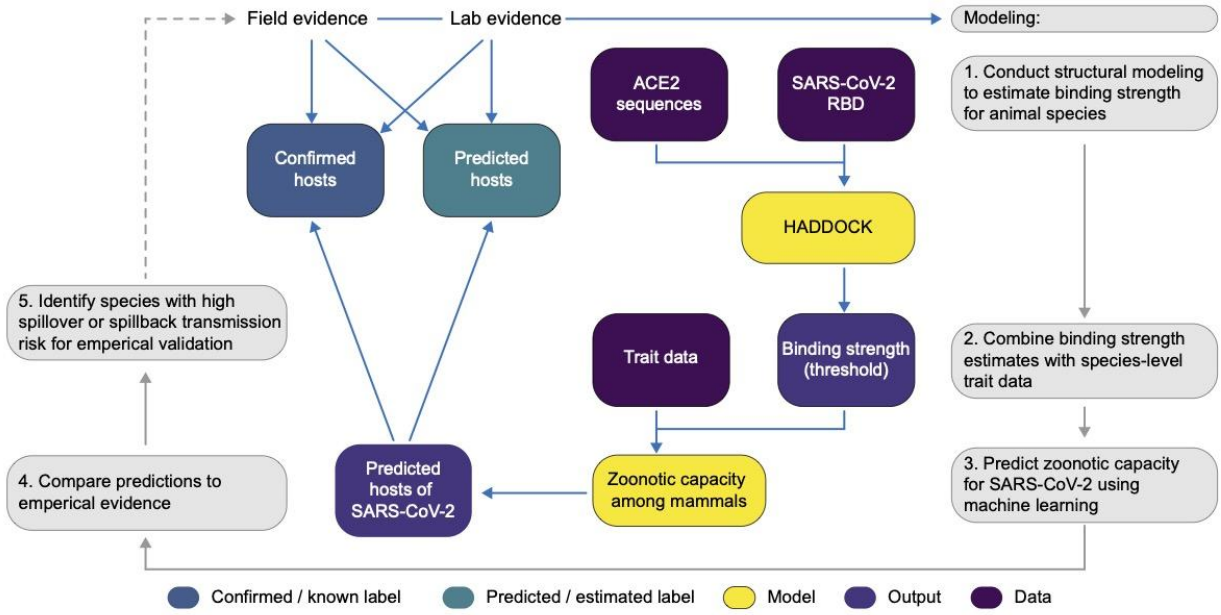

Figure 2. A flowchart showing the progression of our workflow combining evidence from limited lab and field studies with additional data types to predict zoonotic capacity across mammals through multi-scale

statistical modeling (gray boxes, steps 1-5). For all vertebrates with published ACE2 sequences, we modelled the interface of species' ACE2 bound to the viral receptor binding domain using HADDOCK. We then combined the HADDOCK scores, which approximate binding strength, with species' trait data and trained machine learning models for both mammals and vertebrates (yellow boxes). Mammal species predicted to have high zoonotic capacity were then compared to results of *in vivo* experiments and *in silico* studies that applied various computational approaches. Based on predictions from our model, we identified a subset of species with particularly high risk of spillback and secondary spillover potential to prioritize additional lab validation and field surveillance (dashed line).

Methods

Protein sequence and alignment

We assembled a dataset of ACE2 NCBI GenBank accessions that are known human ACE2 orthologs or have high similarity to known orthologs as determined using BLASTx [73]. Using the R package *rentrez* and the accession numbers, we downloaded ACE2 protein sequences [74]. We supplemented these sequences by manually downloading four additional sequences from the MEROPS database [75].

Structural Modeling of ACE2 orthologs bound to SARS-CoV-2 spike

The modeling of all 326 ACE2 orthologs bound to SARS-CoV-2 spike receptor binding domain was carried out as described previously [13], with a few differences. Sequences of ACE2 orthologs were aligned using MAFFT [76] and trimmed to the region resolved in the template crystal structure of hACE2 bound to the SARS-CoV-2 spike (PDB ID: 6m0j, [77]. Ambiguous positions in each sequence, artifacts of the sequencing method, were replaced by Glycine to minimize assumptions about the nature of the amino acid side-chain but still allow for modeling. For each ortholog, we generated 10 homology models using MODELLER 9.24 [78,79], with restricted optimization (*fastest* schedule) and refinement (*very_fast* schedule) settings, and selected a representative model based on the normalized DOPE score. These representative models were then manually inspected and 27 were removed from further analysis due to large insertions/deletions or to the presence of too many ambiguous amino acids at the interface with spike. Each validated model was submitted for refinement to the HADDOCK web server [80], which ran 50 independent short molecular dynamics simulations in explicit solvent to optimize the interface between the two proteins . For each one of the animal species in our study, we assigned an average and standard deviation of the scores of the 10 best refined models, ranked by their HADDOCK score -- a combination of van der Waals, electrostatics, and desolvation energies. A lower (more negative) HADDOCK score predicts stronger binding between the two proteins. We hereafter refer to predicted binding strength, or simply binding strength, to indicate HADDOCK score. The HADDOCK server is freely available, and we provide code to reproduce analyses or to aid in the application of this modeling approach to other similar problems (https://zenodo.org/record/4517509).

Trait data collection and cleaning

We gathered ecological and life history trait data from AnAge [81], Amniote Life History Database [82], and EltonTraits [83], among other databases (Supplementary Table 1; for details on data processing, see Supplementary Methods with all supplementary data, figures, methods, and tables available at https://doi.org/10.25390/caryinstitute.c.5293339). Using these data, we also engineered additional traits that have shown importance in predicting host-pathogen associations in other contexts. For example, as a measure of habitat breadth [84], we computed for each species the percentage of ecoregions it occupies. To assess the influence of sampling bias across species, we used the wosr R package [85] to count the number of studies returned in a search in Web of Science for each species' Latin binomial and included this as a proxy for sampling bias in our model.

Following the results of initial structural modeling (described above), we observed that per-residue energy decomposition analysis of HADDOCK scores for 29 species indicated that all species with strong predicted binding had in common a salt bridge between SARS-CoV-2 K417 and a negatively charged amino acid at position 30 in the ACE2 sequence [13]. Given the apparent effect of amino acid 30 on overall binding strength, we constructed an additional feature to denote whether amino acid 30 is negatively charged (and therefore more likely to support strong binding) and included this feature as an additional trait in our models.

Modeling

Quantifying a threshold for zoonotic capacity using HADDOCK. While ACE2 binding is necessary for viral entry into host cells, it is not sufficient for SARS-CoV-2 transmission. Multiple *in vivo* experiments suggest that not all species that are capable of binding SARS-CoV-2 are capable of transmitting active infection to other individuals (e.g., cattle, *Bos taurus* [33]; bank voles, *Myodes glareolus* (Ulrich et al. 2021)). Viral replication, and infectious viral shedding that enables onward transmission, are both required for a species to become a suitable bridge or reservoir species for SARS-CoV-2. In order to constrain our predictions to species with the greatest potential to perpetuate onward transmission, we trained our models on a conservative threshold of binding strength (HADDOCK score = -129). This value is between the scores for two species: the domestic cat (*Felis catus*), which is currently the species with weakest predicted binding with confirmed conspecific transmission [86], and the pig (*Sus scrofa*), which shows the strongest estimated binding for which experimental inoculation failed to cause detectable infection [38]. Binding strength was binarized according to this threshold, above which it is more likely that both infection and onward transmission will occur following the results of multiple empirical studies (Table 1). We note that there are species confirmed to be susceptible whose predicted binding strength is weaker than cats, but conspecific transmission has not been confirmed in these species. While it is likely that intraspecific transmission will be reported for additional species as the pandemic continues, the binding strength selected for this analysis represents an appropriately conservative threshold based on currently available evidence. For additional modeling details, see Supplementary Methods.

Trait-based modeling to predict zoonotic capacity

We applied generalized boosted regression [87] to host trait data to predict species' binding strength to SARS-CoV-2. We applied this approach initially to all of the vertebrate species for which we estimated HADDOCK scores, but these models did not perform well. This was likely due to extensive dissimilarities among traits describing different classes of organisms. For instance, traits that are commonly measured for reptiles are different from those of interest for birds or amphibians. Moreover, currently available ACE2 sequences are dominated by ray-finned fishes and mammals.

Given that only mammals have so far been confirmed as both susceptible and capable of onward transmission of SARS-CoV-2, we created a separate set of models to make zoonotic capacity predictions for mammals only. For this mammal-only dataset, we gathered additional species-level traits from PanTHERIA [88] and added a series of binary fields for taxonomic order (based on [89]; Supplementary Table 2). We then applied boosted regression (BRT; gbm package [90] in R version 4.0.0 [90,91]) to impute missing trait data for mammal species (e.g., [67]; see Supplementary Methods for details on imputation methods and results).

Many of the mammals for which we found the strongest evidence of zoonotic capacity are domesticated to some degree (pets, farmed or traded animals, lab models) [11,38,53]. Relative to their ancestors or wild conspecifics, domesticated animals often have distinctive traits [92] that are likely to influence the number of zoonoses found in these species [93]. To account for trait variation due to domestication in certain species, we modeled mammals in two ways. First, we incorporated a variable indicating whether the source populations from which trait data were collected are wild or non-wild (e.g., farmed, pets, laboratory animals; non-wild status confirmed by the Mammal Diversity Database [94]). Trait data collected from both wild and non-wild individuals were considered to represent non-wild species for the purposes of this model. In a second approach, we used only the wild species for model training and evaluation. For both approaches, pre-imputation trait values were used for all non-wild mammals during model training, evaluation, and prediction.

Boosted regression (BRT) is an ensemble machine learning approach that accommodates non-random patterns of missing data, nonlinear relationships, and interacting effects among predictors. In a BRT model, a sequence of regression models are fit by recursive binary splits, with each additional regression modeling those instances that were poorly accounted for by the previous regression iterations in the tree [87]. We applied grid search to select optimal hyperparameters, and repeated model fitting 50 times using bootstrapped training sets of 80% of labeled data. We measured performance by the area under the receiver operating characteristic curve (AUC) for predictions made on the test dataset (remaining 20%), corrected by comparing with null models created by target shuffling, which employed similar bootstrapping (50 times). Detailed methods can be found in Supplementary Methods. We discuss herein the results of model predictions about zoonotic capacity made by applying this final model to all mammal species. We also report the mean and variation in predicted probabilities across all 50 bootstrapped models in Supplementary File 1.

To visualize geographic patterns, we mapped the geographic ranges of mammal species predicted within the 90th percentile of zoonotic capacity for SARS-CoV-2 using International Union for the Conservation of Nature (IUCN) polygons of species distributions [95]. We subset to the species found in human-associated habitats (e.g., urban areas, crop lands, heavily degraded forests; based on IUCN 2020), and also masked their ranges to areas of high human case counts (using SARS-CoV-2 case data from the COVID-19 Data Repository at Johns Hopkins University [1]).

Additional methods and results of other uninformative model variations are also described in Supplementary Methods and Supplementary Table 3 (e.g., a model in which binding strength is modeled as a continuous rather than a threshold measure, a model predicting the charge at amino acid 30, a model for all vertebrate species) (https://doi.org/10.25390/caryinstitute.c.5293339). We provide code and data files for carrying out boosted regression tree models (https://github.com/HanLabDiseaseEcology/zoonotic_capacity). Details about how the species susceptibility predictions from past studies were standardized into categories (low, medium, high; Figure 1) are also available in Supplementary Methods.

Results

ACE2 host protein sequences and alignment

The ACE2 protein sequence alignment of the orthologs from 326 species spans eight classes and 87 orders (https://zenodo.org/record/4517509). The majority of sequences belonged to the classes Actinopterygii (22.1%), Aves (23.3%), and Mammalia (46.6%). Sequence length ranged from 344 amino acids to 872 with a median length of 805.

Structural modeling of viral binding strength

We predicted binding strength for 299 vertebrates, including 142 mammals. These binding strength scores represented six classes and 80 orders and ranged between -167.816 and -105.615. Across these six vertebrate classes, the strongest predicted binding between ACE2 and SARS-CoV-2 (corresponding to the lowest mean HADDOCK scores) were in ray-finned fishes (Actinopterygii; mean = -137.945) and mammals (Mammalia; mean = -129.193) (Figure 3A). Four of these six classes included at least one species predicted to have stronger binding than *Felis catus* (Figure 3B). Among well-represented mammalian orders (those containing at least 10 species with binding strength predictions), Primates and Carnivora showed predicted mean binding strengths that were stronger than domestic cats (Figure 3C).

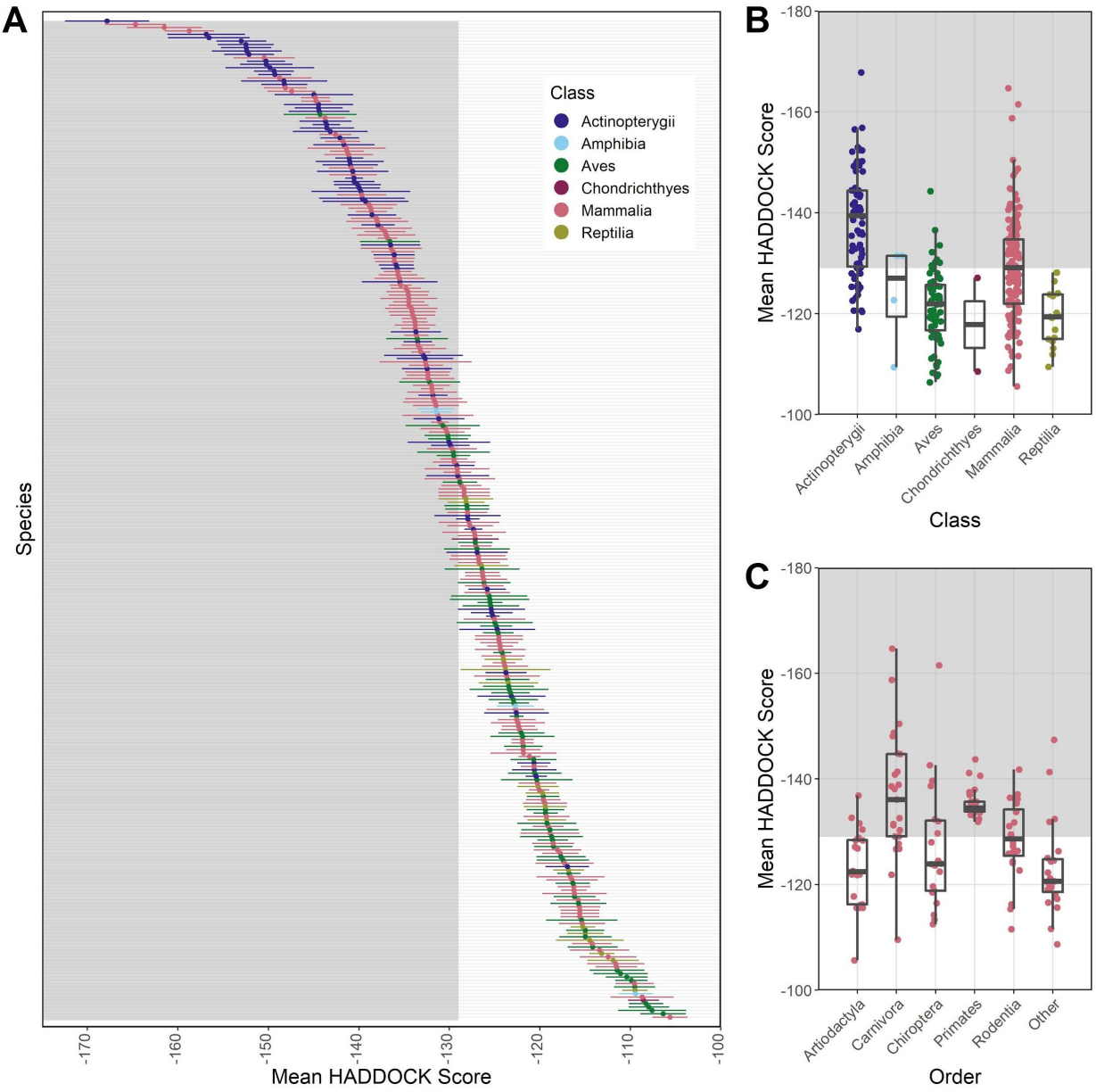

Figure 3. Plots showing results from modeling species' ACE2 interaction with SARS-CoV-2 RBD using HADDOCK to predict binding strength (measured as arbitrary units). HADDOCK scores that predict stronger binding are more negative. The mean and standard deviation of the HADDOCK score for vertebrate species **(A)** for which ACE2 orthologs are available. Binding strengths vary across vertebrate classes **(B)** and across the five most speciose mammalian orders **(C)**. The "Other" category contains species across multiple orders for which ACE2 sequences were available, each with fewer than 10 representative species in the order. The shaded regions of all panels represent predicted binding that is as strong or stronger than (more negative values than) the domestic cat (*Felis catus*), which represents our conservative zoonotic capacity threshold based on currently available empirical evidence.

Species predictions of zoonotic capacity from trait-based machine learning models

The best performing model was trained on a mammal-only dataset with trait imputation and showed corrected test AUC of 0.72 (for results of all other model variations, see Supplementary Table 3). We used this model to generate predictions of zoonotic capacity among mammal species. Citation count, as a proxy for study effort, had ~1% relative importance, suggesting that sampling bias across species had little influence on the model.

This zoonotic capacity model identified 540 species within the 90th percentile probability (0.826 or higher, compared to a total of 2,401 mammal species with prediction scores above 0.5; see Supplementary File 1 for predictions on all 5,400 mammal species, https://doi.org/10.25390/caryinstitute.c.5293339) .

The top 10% of species with the highest predicted probabilities includes representatives from 13 orders. Most primates were predicted to have high zoonotic capacity and collectively showed stronger viral binding compared to other mammal groups (Figure 4). Additional orders with numerous species predicted to have high zoonotic capacity (at least 75% of species above 0.5) include Hyracoidea (hyraxes), Perissodactyla (odd-toed ungulates), Scandentia (treeshrews), Pilosa (sloths and anteaters), Pholidota (pangolins), and non-cetacean Artiodactyla (even-toed ungulates) (Figure 4).

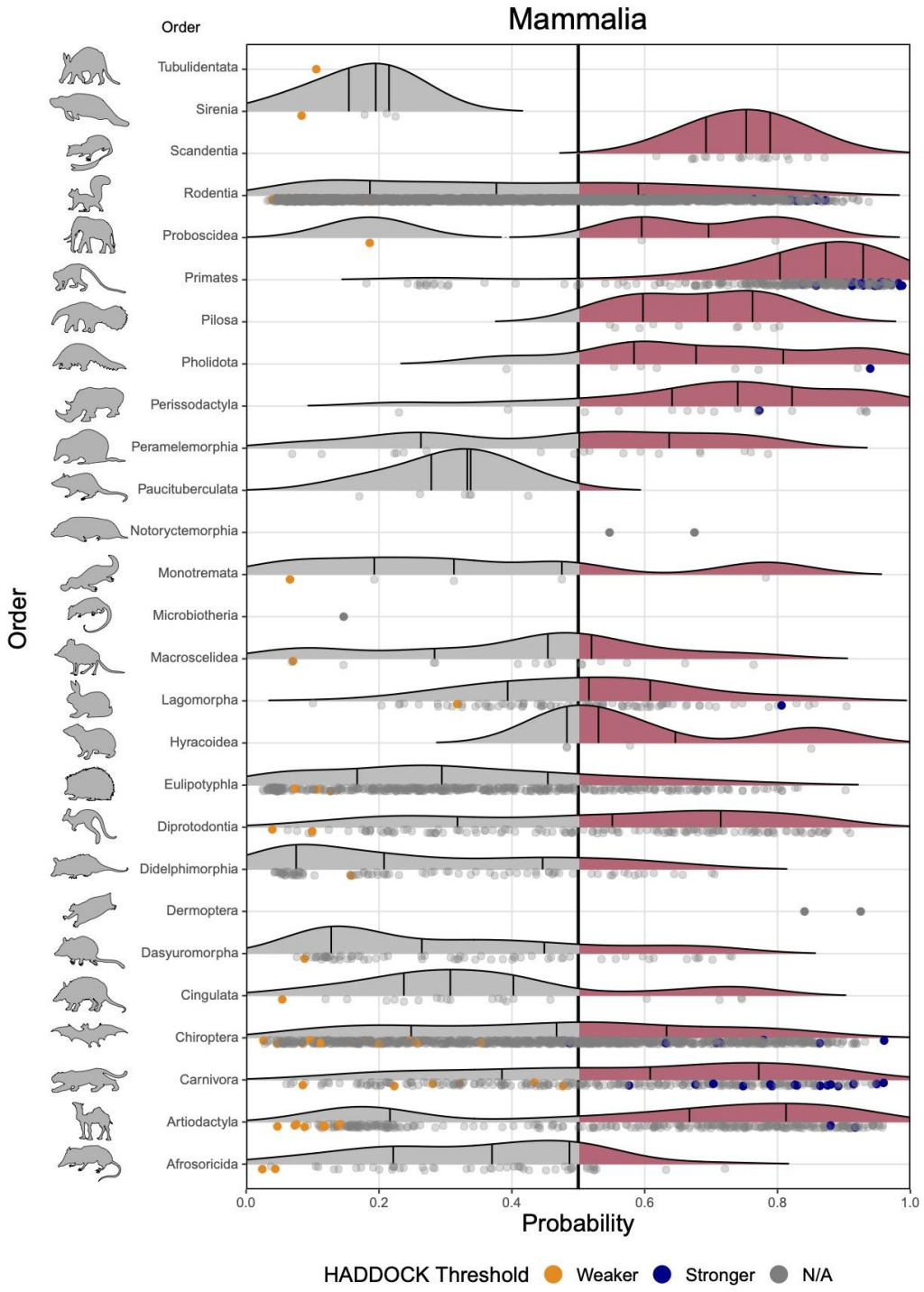

Figure 4. Ridgeline plots showing the distribution of predicted zoonotic capacity across mammals. Predicted probabilities for zoonotic capacity across the x-axis range from 0 (likely not susceptible) to 1 (zoonotic capacity predicted to be the same or greater than *Felis catus*), with the vertical line representing 0.5. The y-axis depicts all mammalian orders represented by our predictions. Density curves represent the distribution of the predictions, with those parts of the curve over 0.5 colored pink and lines representing distribution quartiles. The predicted values for each order are shown as points below the

density curves. Points that were used to train the model are colored: orange represents species with weaker predicted binding, blue represents species with stronger predicted binding. Selected family-level distributions are shown in the Supplemental Figures 5-6 (https://doi.org/10.25390/caryinstitute.c.5293339).

Comparison of species predictions

Comparing species predictions across multiple computational approaches

Our model combined species traits and viral binding strength to predict zoonotic capacity (susceptibility and onward transmission), which was defined as a threshold value based on experimental studies confirming intraspecific transmission among animals, and is therefore more conservative than thresholds adopted by other studies (e.g., based on binding strength, [30]). In addition, our modeling approach (machine learning) and prediction targets (zoonotic capacity) differed compared to existing computational approaches, which applied sequence-based or structure-based analyses constrained by the small number of published ACE2 sequences. Despite these differences, comparing species predictions generated by multiple approaches can be useful for gauging consensus, and for comparing how predictions change from one method to another. Across approaches, there was general agreement in the predictions for primates and for a select group of artiodactyls and carnivores (Figure 5). Our model results also agree with low susceptibility predictions made by several previous studies using sequence-based approaches (e.g., in certain bats and rodents). In general, we note that structure-based models predicted a smaller proportion of species to have low susceptibility compared to sequence-based studies.

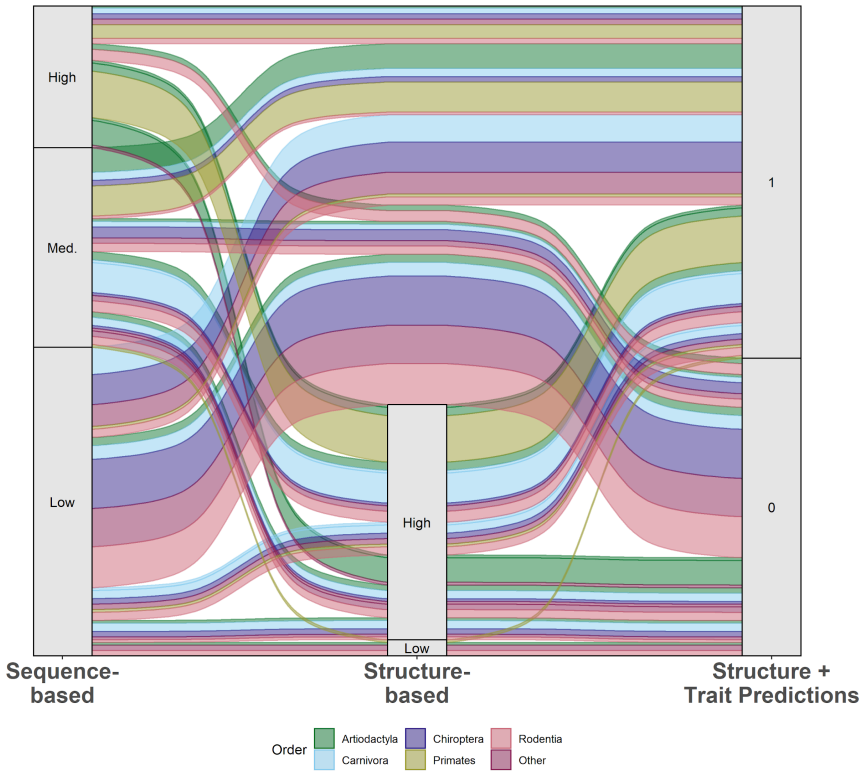

Figure 5. An alluvial plot comparing predictions of species susceptibility from multiple methods. Existing studies (listed in Supplementary Methods) are categorized as either sequence-based or structure-based. Predictions from our zoonotic capacity model result from combining structure-based modeling of viral binding with organismal traits using machine learning to distinguish species with zoonotic capacity above (1) or below (0) a conservative threshold value set by domestic cats (*Felis catus*). Colors represent unique mammalian orders, and the width of colored bands represent the relative number of species with that combination of predictions across methods. See Supplementary Methods (https://doi.org/10.25390/caryinstitute.c.5293339) for details on how species across multiple studies were assigned to categories (high, medium, low).

Comparing model predictions to in vivo *outcomes*

While there were comparatively many fewer *in vivo* studies exploring susceptibility and transmission in animals, our model predictions matched the results of most of these studies (Figure 1). For instance, experiments on deer mice (*Peromyscus maniculatus*; [51,52]) and raccoon dogs (*Nyctereutes procyonoides*; [47]) confirmed SARS-CoV-2 infection and transmission to naive conspecifics. Our model also estimated a high probability of zoonotic capacity of American mink for SARS-CoV-2 (*Neovison vison*, probability=0.83, 90th percentile), in which farmed individuals present severe infection from human spillback, and demonstrate the capacity to transmit to conspecifics as well as to humans [11,46]. Our model also correctly predicted relatively low zoonotic capacity for big brown bats (*Eptesicus fuscus*; [40]).

Some model predictions differed from the results of experimental studies. For instance, our model estimated a moderately high probability of zoonotic capacity for pigs (*Sus scrofa*, probability = 0.72, ~80th percentile). While some computational and cell-based studies have predicted strong viral binding in this species (Liu et al. 2021; Luan et al. 2020), *in vivo* studies report no detectable infection or onward transmission of SARS-CoV-2 [38,53]. Similarly for cattle (*Bos taurus*), our model also estimated a moderately high probability for zoonotic capacity (0.72, ~80th percentile), but in a live animal experiment, cattle were confirmed to be susceptible to infection, although no transmission was observed to virus-naive conspecifics [33].

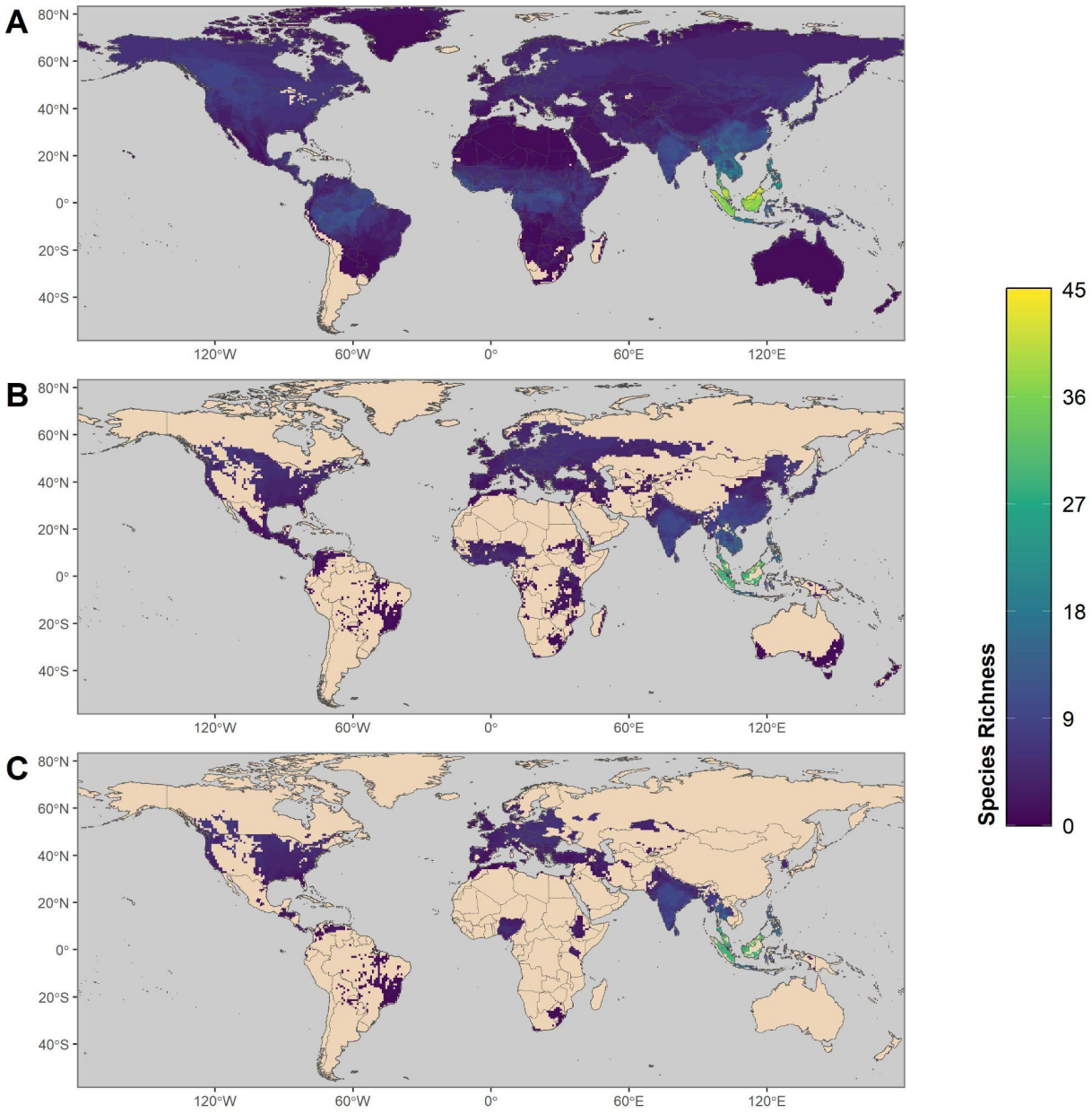

Figure 6: Maps showing the global distribution of species with predicted capacity to transmit SARS-CoV-2. **(A)** depicts global species richness of the top 10 percent of model-predicted zoonotic capacity. Geographic ranges of this subset of species were filtered to those associated with human-dominated or human-altered habitats **(B)**, and further filtered to show the subset of species that overlaps with areas of high human SARS-CoV-2 positive case counts (over 100,000 cumulative cases as of 17 May 2021) **(C)**. For a full list of model-predicted zoonotic capacity of species by country, see Supplementary File 2 (https://doi.org/10.25390/caryinstitute.c.5293339).

Discussion

We combined structure-based inference about viral binding with species-level trait data to make predictions about the capacity of animal species to become zoonotic hosts of SARS-CoV-2 (*zoonotic capacity*). Our definition of zoonotic capacity includes critical elements necessary for an animal host to serve as a zoonotic host, either as a new enzootic reservoir or as a bridge host capable of seeding secondary transmission to humans following an initial spillback event. First, species susceptibility to SARS-CoV-2 is a necessary condition, which we assumed to depend on the strength of binding between SARS-CoV-2 RBD and host ACE2. Second, zoonotic capacity includes the capacity for onward transmission, which we model as a threshold quantity based on available empirical evidence that confirms SARS-CoV-2 transmission to naive conspecific hosts. To extend predictive capacity beyond the small number of species for which ACE2 sequences are currently available, we leveraged data on intrinsic biological traits of 5,400 mammal species. This combined modeling approach predicted zoonotic capacity with 72% accuracy, and identified numerous mammal species whose predicted zoonotic capacity meets or exceeds the viral susceptibility and transmissibility observed in experimental infections with SARS-CoV-2. In addition to wide agreement with *in vivo* study results produced to date (Table 1), these model predictions corroborate the predictions of previous studies generated using the limited number of available ACE2 sequences (Figure 1). Below we discuss predictions of zoonotic capacity for a number of ecologically and epidemiologically relevant categories of mammalian hosts.

Captive, farmed, or domesticated species. Given that the type and frequency of contact with humans fundamentally underlies transmission risk, it is notable that our model predicted high zoonotic capacity for multiple captive species that have also been confirmed as susceptible to SARS-CoV-2 via experiments or natural infections. These include numerous carnivore species, such as large cats from multiple zoos and pet dogs and cats. Our model also predicted high SARS-CoV-2 zoonotic capacity for many farmed, domesticated, and live traded species. The water buffalo (*Bubalus bubalis*), widely bred for dairy production and farming, had the highest probability of zoonotic capacity among livestock (0.91). The 90th percentile of model predictions also included American mink (*Neovison vison*), red fox (*Vulpes vulpes*), sika deer (*Cervus nippon*), white-lipped peccary (*Tayassu pecari*), nilgai (*Boselaphus tragocamelus*), and raccoon dogs (*Nyctereutes procyonoides*), all of which are farmed, with the latter two considered invasive species in some areas [96,97]. In addition to the risks of secondary spillover to humans and the potential for large economic losses from culling infected animals [98], the escape of

farmed individuals into wild populations has implications for the spread and enzootic establishment of SARS-CoV-2 [21]. These findings also have implications for vaccination strategies, for instance, prioritizing people in regular contact with potential bridge species (e.g., veterinarians, abattoir-workers, farmers, etc).

Live traded or hunted wildlife species. The majority of the legal live mammal trade consists of primates and carnivores [99], and model predictions included several species from these groups. Our model predicted high zoonotic capacity in 20 (out of 21) species in the primate genus *Macaca* which comprise the majority of all live-traded primates. Several live-traded carnivores and pangolins were also assigned high zoonotic capacity, including the Asiatic black bear (*Ursus thibetanus*), grey wolf (*Canis lupus*), and jaguar (*Panthera onca*), the Philippine pangolin (*Manis culionensis*) and Sunda pangolin (*M. javanica*). Pangolins are notable because one of the betacoronaviruses with the highest sequence similarity to SARS-CoV-2 was isolated from Sunda pangolins [100,101].

Commonly hunted species in the top 10% of predictions include duiker (*Cephalophus zebra*, West Africa), warty pig (*Sus celebes*, Southeast Asia), and two species of deer (*Odocoileus hemionus* and *O. virginianus*) that are widespread across the Americas. The white-tailed deer (*O. virginianus*) was recently confirmed capable of transmitting SARS-CoV-2 to conspecifics via indirect contact (aerosolized virus particles) [58].

Bats. Similarly, bats are of special interest because of the high diversity of betacoronaviruses found in *Rhinolophus spp.* and other bat species [102–105]. Our model identified 35 bat species within the 90th percentile of zoonotic capacity for SARS-CoV-2. Within the genus *Rhinolophus*, our model identified the large rufous horseshoe bat (*Rhinolophus rufus*), a known natural host for bat betacoronaviruses [102] and a congener to three other horseshoe bats harboring betacoronaviruses with high nucleotide sequence similarity to SARS-CoV-2 (~92-96%) [6,106,107]. For these three species, our model assigned a range of probabilities for SARS-CoV-2 zoonotic capacity (*Rhinolophus affinis* (0.58), *R. malayanus* (0.70), and *R. shameli* (0.71)) and also predicted relatively high probabilities for two congeners, *Rhinolophus acuminatus* (0.84) and *R. macrotis* (0.70). These predictions are in agreement with recent experiments demonstrating efficient viral binding of SARS-CoV-2 RBD for *R. macrotis* [108] and confirmation of SARS-CoV-2-neutralizing antibodies in field-caught *R. acuminatus* harboring a closely related betacoronavirus [109].

Within the genus *Pteropus* (flying foxes), our model identified 17 species with high probabilities of zoonotic capacity for SARS-CoV-2. Some of these species are confirmed reservoirs of other zoonotic viruses in Southeast Asia (e.g., henipaviruses in *P. lylei*, *P. vampyrus*, *P. conspicillatus*, and *P. alecto*). While contact patterns between bats and humans may be somewhat less direct compared with captive or farmed species, annual outbreaks attributed to viral spillover transmission from bats illustrate a persistent epizootic risk to humans [110–112] and confirm that gaps in systematic surveillance of zoonotic viruses, including betacoronaviruses, remain an urgent priority (e.g., [113]).

Rodents. Our model identified 76 rodent species with high zoonotic capacity for SARS-CoV-2, some of which thrive in human-altered settings. Among these, our model predicted high probabilities for the deer mouse (*Peromyscus maniculatus*) and the white-footed mouse (*P. leucopus*). These are among the most well-studied mammals in North America, in part due to their status as zoonotic reservoirs for multiple zoonotic pathogens and parasites [114–116]. Experimental infection, viral shedding, and sustained intraspecific transmission of SARS-CoV-2 were recently confirmed for *P. maniculatus* [51,52].

Our model predicted low zoonotic capacity for *Mus musculus* (0.11), corresponding with *in vivo* experiments suggesting this species is not susceptible to infection by the initial human variant of SARS-CoV-2[19], although notably, more recent experiments have confirmed the susceptibility of *M. musculus* to two newer human-derived variants[20]. Also in the top 10% were two rodent species considered to be human commensals whose geographic ranges are expanding due to human activities: *Rattus argentiventer* (0.84) and *R. tiomanicus* (0.79) (Supplementary File 1) [117–119]. Additional common rodent species with relatively high probabilities of zoonotic capacity include domesticated guinea pigs (*Cavia porcellus*), gerbils (*Gerbillus gerbillus*, *Meriones tristrami*), and several common mouse species (*Apodemus peninsulae*, *A. flavicollis*, and *A. sylvaticus*), all of which are known reservoirs for other zoonotic diseases. It is notable that many of these rodent species are regularly preyed upon by carnivore species, such as the red fox (*Vulpes vulpes*) or domestic cats (*Felis catus*) who themselves were predicted to have high zoonotic capacity for SARS-CoV-2.

Species with large geographic ranges. With sufficient opportunity for infectious contact, the risk of zoonotic spillback transmission increases with SARS-CoV-2 prevalence in human populations. Among species with high model-predicted zoonotic capacity, there were several relatively common species with very large geographic ranges or synanthropic tendencies that overlap with global hotspots of COVID-19 in people (Figure 6, Supplementary File 2). Notable species that are widely distributed across much of the northern hemisphere include the red fox (*Vulpes vulpes*, ~50 countries), the European polecat (*Mustela putorius*), the raccoon dog (*Nyctereutes procyonoides*), stoat (*Mustela erminea*) and wolf (*Canis lupus*). White-tailed deer (*Odocoileus virginianus*) are among the most geographically widespread species across Latin American countries with high SARS-CoV-2 prevalence. Globally, South and Southeast Asia had the highest diversity of mammal species with high predicted zoonotic capacity for SARS-CoV-2 (~90 species). Notable examples in the 90th percentile probability in this region include both rodents and bats. For example, Finlayson's squirrel (*Callosciurus finlaysonii*) is native to Mainland Southeast Asia, but introductions via the pet trade in Europe have led to invasive populations in multiple countries [120]. Hunting has been documented for numerous bat species with geographic ranges across Southeast Asia (e.g., *Cheiromeles torquatus*, *Cynopterus brachyotis*, *Rousettus amplexicaudatus*, *Macroglossus minimus*) [121,122], and there were multiple additional bat species in the 90th percentile from Asia and Africa where bats are subject to hunting pressure and from which other betacoronaviruses have been identified [105,123]. There were also several wide-ranging species whose contact with humans are limited to specialized settings. For instance, biologists and wildlife managers handle live individuals for research

purposes, including grizzly bear (*Ursus arctos*), polar bear (*Ursus maritimus*), and wolf (*Canis lupus*), all of which are in the 89th percentile or above for predicted zoonotic capacity.

Other high priority mammal species. Species with more equivocal predictions about zoonotic capacity that are in frequent contact with humans warrant further investigation. For instance, while species such as horses (*Equus caballus*), goats (*Capra hircus*), and guinea pigs (*Cavia porcellus*) are not in the top 10% of predicted zoonotic capacity, due to the nature of their contact with humans they may experience greater risks of spillback infection, or pose a greater risk to humans for secondary spillover infection compared to many wild species. Conversely, while certain endangered or nearly extinct species are predicted to have relatively high zoonotic capacity, they may have fewer opportunities for human contact. For species of conservation concern, spillback transmission of SARS-CoV-2 from humans presents an important source of risk[28,124], particularly for populations that are under active management, including *ex situ* management such as captive breeding. These species include the scimitar-horned oryx (*Oryx dammah*), addax (*Addax nasomaculatus*), some Antarctic fauna and mountain gorillas (*Gorilla beringei*) in which SARS-CoV-2 spillback infection may occur through close-proximity eco-tourism activities [125,126]. Indeed, spillback transmission of SARS-CoV-2 has already been confirmed in a closely related species, the Western lowland gorilla (*Gorilla gorilla*) in captivity [127], leading to the vaccination of bonobos and orangutans with an experimental COVID-19 vaccine [128]. These species may benefit from focused risk mitigation efforts, such as those enacted recently to protect endangered black-footed ferrets (*Mustela nigripes*) from potential SARS-CoV-2 spillback [129].

All fifteen species of *Tupaia* treeshrews were predicted by our model to have medium to high probability (ranging from 0.62 to 0.87). One species, *T. belangeri*, has been explored as a potential lab model for several human infectious diseases including SARS-CoV-2 [130] but relative to other treeshrews, our model assigned only medium probability for SARS-CoV-2 zoonotic capacity in this species (0.67). This result matches lab studies reporting asymptomatic infection and low viral shedding in *T. belangeri* [54]. In contrast, the common treeshrew (*T. glis*) was in the 94th percentile of zoonotic capacity (0.87 probability). These two species are sympatric in parts of their range, exist in close proximity to humans, and also overlap geographically with COVID-19 hotspots in Southeast Asia, suggesting the possibility of spillover transmission among congeners if spillback transmission occurs from humans to these species.

Strengthening predictive capacity for zoonoses. While there was wide agreement between our model predictions and empirical studies, examining biases and mismatches between experimental results and model-generated predictions will focus research attention on characterizing what factors underlie the disconnects between predicted and observed zoonotic capacity. For instance, in pigs (*Sus scrofa*) this study along with multiple other computational and experimental studies predicted susceptibility to SARS-CoV-2 (Figure 1), but this prediction has not been supported by results from whole animal inoculations, which so far have showed unproductive infection [38,53]. As an example of how methods influence predictions, our model incorporating both molecular structure and species traits generally estimated weaker binding

strengths for cetaceans, and high probabilities for catarrhine primates compared to other studies that employed different structural modeling methods [27,28,30].

Disconnects between real-world observations and *in silico* predictions of zoonotic capacity may arise because host susceptibility and transmission capacity are necessary but not sufficient for zoonotic risk to be realized in natural settings. These processes depend strongly on the cellular environments in which cell entry and viral replication take place (e.g., the presence of suitable receptors and key proteases, [7]), and on host immunogenicity [131]. These processes are therefore embedded in a broader ecological context impacting intra-host infection dynamics (latency, recrudescence, tolerance), and environmental drivers of host susceptibility and viral persistence that collectively determine where and when spillover may occur [131–134]. Insofar as data limitations preclude perfect computational predictions of zoonotic capacity (e.g., limited ACE2 sequences or species trait data, structural information over or under-predicting susceptibility in species), laboratory experiments are also limited in assessing true zoonotic capacity. For SARS-CoV-2 and other host-pathogen systems, animals that are readily infected in the lab appear to be less susceptible in non-lab settings (ferrets in the lab vs. mixed results in ferrets as pets [37,53,135]; rabbits in the lab vs. rabbits as pets [48,136]). Moreover, wildlife hosts confirmed to shed multiple zoonotic viruses in natural settings (e.g., bats, [137]) can be much less tractable for laboratory investigations (for instance, requiring high biosecurity containment and very limited sample sizes in unnatural settings). While laboratory experiments are critical for understanding mechanisms of pathogenesis and disease, without field surveillance and population-level studies they offer imperfect reflections of zoonotic capacity in the natural world. These examples illustrate that there is no single methodology sufficient to understand and predict zoonotic transmission, for SARS-CoV-2 or any zoonotic pathogen. They also demonstrate the need for improved coordination among theoretical and statistical models, lab work, and field work to improve zoonotic predictive capacity [138]. Integration of multiple methodologies and disciplines, as done here, and more efficient iteration between computational predictions, laboratory experiments, and targeted animal surveillance will better link transmission mechanisms to the broader conditions underpinning zoonotic disease emergence in nature.

Acknowledgments

We are grateful for discussions with Drs. Alexandre Bonvin, Colin Parrish, Dennis Bente, Hyunwook Lee, Kathryn Hanley, Susan Hafenstein, and John Paul Schmidt about various components of this project. This work was supported by the NSF EEID program (DEB 1717282), DARPA PREEMPT program (D18AC00031), CREATE-NEO, a member of the NIH NIAID CREID program (1U01 AI151807-01), and the NVIDIA Corporation GPU grant program (BAH); by the NSF Polar program (OPP 1935870, 1947040) (AV); and by NIH NIGMS (R35GM122543) (JPGLMR).

Competing interests

The authors declare no competing interests.

References

1. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
2. WHO. WHO coronavirus disease (COVID-19) dashboard. (2021).
3. Keele, B. F. *et al.* Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523–526 (2006).
4. Gage, K. L. & Kosoy, M. Y. Natural history of plague: perspectives from more than a century of research. *Annu. Rev. Entomol.* **50**, 505–528 (2005).
5. Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889 (2005).
6. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
7. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **5**, 562–569 (2020).
8. Chou, C.-F. *et al.* ACE2 orthologues in non-mammalian vertebrates (Danio, Gallus, Fugu, Tetraodon and Xenopus). *Gene* **377**, 46–55 (2006).
9. Guth, S., Visher, E., Boots, M. & Brook, C. E. Host phylogenetic distance drives trends in virus virulence and transmissibility across the animal-human interface. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190296 (2019).
10. WHO. *SARS-CoV-2 mink-associated variant strain – Denmark*. https://www.who.int/csr/don/06-november-2020-mink-associated-sars-cov2-denmark/en/ (2020).
11. Oude Munnink, B. B. *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* (2020) doi:10.1126/science.abe5901.

12. Garry, R. F. Mutations arising in SARS-CoV-2 spike on sustained human-to-human transmission and human-to-animal passage. *Virological* https://virological.org/t/mutations-arising-in-sars-cov-2-spike-on-sustained-human-to-human-transmission-and-human-to-animal-passage/578 (2021).
13. Rodrigues, J. P. G. L. M. *et al.* Insights on cross-species transmission of SARS-CoV-2 from structural modeling. *PLoS Comput. Biol.* **16**, e1008449 (2020).
14. Davies, N. G. *et al.* Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. *medRxiv* 2020.12.24.20248822 (2020) doi:10.1101/2020.12.24.20248822.
15. Volz, E. *et al.* Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv* 2020.12.30.20249034 (2021) doi:10.1101/2020.12.30.20249034.
16. Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virological* https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (2020).
17. Tegally, H. *et al.* Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* 2020.12.21.20248640 (2020) doi:10.1101/2020.12.21.20248640.
18. Van Egeren, D. *et al.* Risk of evolutionary escape from neutralizing antibodies targeting SARS-CoV-2 spike protein. *bioRxiv* (2020) doi:10.1101/2020.11.17.20233726.
19. Bao, L. *et al.* The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature* **583**, 830–833 (2020).
20. Montagutelli, X. *et al.* The B1.351 and P.1 variants extend SARS-CoV-2 host range to mice. *Cold Spring Harbor Laboratory* 2021.03.18.436013 (2021) doi:10.1101/2021.03.18.436013.
21. DeLiberto, T. & Shriner, S. *ProMED*. https://promedmail.org/promed-post/?id=8015608

(2020).

22. ODA. Mink at affected Oregon farm negative for SARS-CoV-2, wildlife surveillance continues. (2020).
23. Shriner, S. *et al.* SARS-CoV-2 Exposure in Escaped Mink, Utah, USA. *Emerging Infectious Disease journal* **27**, (2021).
24. Lam, S. D. *et al.* SARS-CoV-2 spike protein predicted to form complexes with host receptor protein orthologues from a broad range of mammals. *Sci. Rep.* **10**, 16471 (2020).
25. Liu, Z. *et al.* Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J. Med. Virol.* **92**, 595–601 (2020).
26. Luan, J., Jin, X., Lu, Y. & Zhang, L. SARS-CoV-2 spike protein favors ACE2 from Bovidae and Cricetidae. *J. Med. Virol.* (2020).
27. Mathavarajah, S., Stoddart, A. K., Gagnon, G. A. & Dellaire, G. Pandemic danger to the deep: the risk of marine mammals contracting SARS-CoV-2 from wastewater. 2020.08.13.249904 (2020) doi:10.1101/2020.08.13.249904.
28. Melin, A. D., Janiak, M. C., Marrone, F., 3rd, Arora, P. S. & Higham, J. P. Comparative ACE2 variation and primate COVID-19 risk. *Commun Biol* **3**, 641 (2020).
29. Kumar, A. *et al.* Predicting susceptibility for SARS-CoV-2 infection in domestic and wildlife animals using ACE2 protein sequence homology. *Zoo Biol.* (2020) doi:10.1002/zoo.21576.
30. Huang, X., Zhang, C., Pearce, R., Omenn, G. S. & Zhang, Y. Identifying the Zoonotic Origin of SARS-CoV-2 by Modeling the Binding Affinity between the Spike Receptor-Binding Domain and Host ACE2. *J. Proteome Res.* **19**, 4844–4856 (2020).
31. Ahmed, R., Hasan, R., Siddiki, A. M. A. M. Z. & Islam, M. S. Host range projection of SARS-CoV-2: South Asia perspective. *Infect. Genet. Evol.* **87**, 104670 (2021).
32. Damas, J. *et al.* 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 (2020).

33. Ulrich, L., Wernike, K., Hoffmann, D., Mettenleiter, T. C. & Beer, M. Experimental Infection of Cattle with SARS-CoV-2. *Emerg. Infect. Dis.* **26**, 2979–2981 (2020).

34. Hamer, S. A. *et al.* Natural SARS-CoV-2 infections, including virus isolation, among serially tested cats and dogs in households with confirmed human COVID-19 cases in Texas, USA. *Cold Spring Harbor Laboratory* 2020.12.08.416339 (2020) doi:10.1101/2020.12.08.416339.

35. Sit, T. H. C. *et al.* Infection of dogs with SARS-CoV-2. *Nature* **586**, 776–778 (2020).

36. USDA. Cases of SARS-CoV-2 in animals in the United States. https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/sa_one_health/sars-cov-2-animals-us (2020).

37. OIE. Events in animals: OIE - World Organisation for Animal Health. https://www.oie.int/en/scientific-expertise/specific-information-and-recommendations/questions-and-answers-on-2019novel-coronavirus/events-in-animals/ (2021).

38. Shi, J. *et al.* Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* **368**, 1016–1020 (2020).

39. Woolsey, C. *et al.* Establishment of an African green monkey model for COVID-19. *bioRxiv* 2020.05.17.100289 (2020) doi:10.1101/2020.05.17.100289.

40. Hall, J. S. *et al.* Experimental challenge of a North American bat species, big brown bat (Eptesicus fuscus), with SARS-CoV-2. *Transbound. Emerg. Dis.* (2020) doi:10.1111/tbed.13949.

41. Zhang, Q. *et al.* SARS-CoV-2 neutralizing serum antibodies in cats: a serological investigation. 2020.04.01.021196 (2020) doi:10.1101/2020.04.01.021196.

42. San Diego Zoo. Gorilla Troop at the San Diego Zoo Safari Park Test Positive for COVID-19. https://zoo.sandiegozoo.org/pressroom/news-releases/gorilla-troop-san-diego-zoo-safari-park-test-positive-covid-19 (2021).

43. Rockx, B. *et al.* Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. *Science* **368**, 1012–1015 (2020).

44. Munster, V. J. *et al.* Respiratory disease in rhesus macaques inoculated with SARS-CoV-2. *Nature* **585**, 268–272 (2020).

45. Sia, S. F. *et al.* Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. *Nature* **583**, 834–838 (2020).

46. Oreshkova, N. *et al.* SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. *Euro Surveill.* **25**, (2020).

47. Freuling, C. M. *et al.* Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection. *Emerg. Infect. Dis.* **26**, 2982–2985 (2020).

48. Mykytyn, A. Z. *et al.* Susceptibility of rabbits to SARS-CoV-2. *Emerg. Microbes Infect.* **10**, 1–7 (2021).

49. Bartlett, S. L. *et al.* SARS-COV-2 INFECTION AND LONGITUDINAL FECAL SCREENING IN MALAYAN TIGERS (PANTHERA TIGRIS JACKSONI), AMUR TIGERS (PANTHERA TIGRIS ALTAICA), AND AFRICAN LIONS (PANTHERA LEO KRUGERI) AT THE BRONX ZOO, NEW YORK, USA. *J. Zoo Wildl. Med.* **51**, 733–744 (2021).

50. Wang, L. *et al.* Complete Genome Sequence of SARS-CoV-2 in a Tiger from a U.S. Zoological Collection. *Microbiol Resour Announc* **9**, (2020).

51. Fagre, A. *et al.* SARS-CoV-2 infection, neuropathogenesis and transmission among deer mice: Implications for reverse zoonosis to New World rodents. *Cold Spring Harbor Laboratory* 2020.08.07.241810 (2020) doi:10.1101/2020.08.07.241810.

52. Griffin, B. D. *et al.* North American deer mice are susceptible to SARS-CoV-2. *bioRxiv* 2020.07.25.221291 (2020) doi:10.1101/2020.07.25.221291.

53. Schlottau, K. *et al.* SARS-CoV-2 in fruit bats, ferrets, pigs, and chickens: an experimental transmission study. *Lancet Microbe* **1**, e218–e225 (2020).

54. Zhao, Y. *et al.* Susceptibility of tree shrew to SARS-CoV-2 infection. *Sci. Rep.* **10**, 16007 (2020).

55. Louisville Zoo. Louisville Zoo Female Snow Leopard Tests Positive for SARS-CoV-2.

https://louisvillezoo.org/louisville-zoo-female-snow-leopard-tests-positive-for-sars-cov-2-media-release/ (2020).

56. Ulrich, L. *et al.* Experimental SARS-CoV-2 Infection of Bank Voles. *Emerg. Infect. Dis.* **27**, 1193–1195 (2021).
57. Georgia Aquarium. Asian Small-Clawed Otters at Georgia Aquarium Test Positive for COVID-19. http://news.georgiaaquarium.org/stories/releases-20210418.
58. Palmer, M. V. *et al.* Susceptibility of white-tailed deer (Odocoileus virginianus) to SARS-CoV-2. *Cold Spring Harbor Laboratory* 2021.01.13.426628 (2021) doi:10.1101/2021.01.13.426628.
59. Gryseels, S. *et al.* Risk of human-to-wildlife transmission of SARS-CoV-2. *Mamm. Rev.* **8**, e00373-17 (2020).
60. Deng, W. *et al.* Rhesus macaques can be effectively infected with SARS-CoV-2 via ocular conjunctival route. *Cold Spring Harbor Laboratory* 2020.03.13.990036 (2020) doi:10.1101/2020.03.13.990036.
61. Rodrigues, J. P. G. L. M. *et al.* Defining the limits of homology modeling in information-driven protein docking. *Proteins* **81**, 2119–2128 (2013).
62. Sander, C. & Schneider, R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* **9**, 56–68 (1991).
63. Li, Y. *et al.* SARS-CoV-2 and Three Related Coronaviruses Utilize Multiple ACE2 Orthologs and Are Potently Blocked by an Improved ACE2-Ig. *J. Virol.* **94**, (2020).
64. Fournier, D., Luft, F. C., Bader, M., Ganten, D. & Andrade-Navarro, M. A. Emergence and evolution of the renin-angiotensin-aldosterone system. *J. Mol. Med.* **90**, 495–508 (2012).
65. Han, B. A., Schmidt, J. P., Bowden, S. E. & Drake, J. M. Rodent reservoirs of future zoonotic diseases. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7039–7044 (2015).
66. Yang, L. H. & Han, B. A. Data-driven predictions and novel hypotheses about zoonotic tick vectors from the genus Ixodes. *BMC Ecol.* **18**, 7 (2018).

67. Han, B. A., O'Regan, S. M., Paul Schmidt, J. & Drake, J. M. Integrating data mining and transmission theory in the ecology of infectious diseases. *Ecol. Lett.* **23**, 1178–1188 (2020).

68. Han, B. A. *et al.* Undiscovered Bat Hosts of Filoviruses. *PLoS Negl. Trop. Dis.* **10**, e0004815 (2016).

69. Han, B. A. *et al.* Confronting data sparsity to identify potential sources of Zika virus spillover infection among primates. *Epidemics* **27**, 59–65 (2019).

70. Yang, X.-L. *et al.* Genetically Diverse Filoviruses in Rousettus and Eonycteris spp. Bats, China, 2009 and 2015. *Emerg. Infect. Dis.* **23**, 482–486 (2017).

71. Goldstein, T. *et al.* The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. *Nat Microbiol* **3**, 1084–1089 (2018).

72. Sorokina, M. *et al.* Structural models of human ACE2 variants with SARS-CoV-2 Spike protein for structure-based drug design. *Sci Data* **7**, 309 (2020).

73. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

74. Winter, D. rentrez: An R package for the NCBI eUtils API. *R J.* **9**, 520 (2017).

75. Rawlings, N. D. *et al.* The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624–D632 (2018).

76. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

77. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).

78. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5.6.1–5.6.37 (2016).

79. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints.

J. Mol. Biol. **234**, 779–815 (1993).

80. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
81. de Magalhães, J. P. & Costa, J. A database of vertebrate longevity records and their relation to other life-history traits. *J. Evol. Biol.* **22**, 1770–1774 (2009).
82. Myhrvold, N. P. *et al.* An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles: Ecological ArchivesE096-269. *Ecology* **96**, 3109–3000 (2015).
83. Wilman, H. *et al.* EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027–2027 (2014).
84. Dallas, T., Park, A. W. & Drake, J. M. Predicting cryptic links in host-parasite networks. *PLoS Comput. Biol.* **13**, e1005557 (2017).
85. Baker, C. *wosr: Clients to the 'Web of Science' and 'InCites' APIs*. (2018).
86. Bosco-Lauth, A. M. *et al.* Experimental infection of domestic dogs and cats with SARS-CoV-2: Pathogenesis, transmission, and response to reexposure in cats. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 26382–26388 (2020).
87. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
88. Jones, K. E. *et al.* PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184. *Ecology* **90**, 2648–2648 (2009).
89. Wilson, D. E. & Reeder, D. M. *Mammal Species of the World: A Taxonomic and Geographic Reference*. (JHU Press, 2005).
90. Greenwell, B., Boehmke, B., Cunningham, J. & Developers, G. B. M. *Generalized Boosted Regression Models*. (Comprehensive R Archive Network (CRAN), 2020).
91. R Core Team. *R: A language and environment for statistical computing*. (2020).

92. Wilkins, A. S., Wrangham, R. W. & Fitch, W. T. The 'domestication syndrome' in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics* **197**, 795–808 (2014).

93. Cleaveland, S., Laurenson, M. K. & Taylor, L. H. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 991–999 (2001).

94. Database, M. D. *Mammal Diversity Database*. (2020). doi:10.5281/zenodo.4139818.

95. IUCN. The IUCN Red List of Threatened Species. (2020).

96. Pitra, C., Schwarz, S. & Fickel, J. Going west—invasion genetics of the alien raccoon dog Nyctereutes procynoides in Europe. *Eur. J. Wildl. Res.* **56**, 117–129 (2010).

97. Milla, R. *et al.* Phylogenetic patterns and phenotypic profiles of the species of plants and mammals farmed for food. *Nat Ecol Evol* **2**, 1808–1817 (2018).

98. Kevany, S. Danish Covid mink cull and future disease fears will kill fur trade, say farmers. *The Guardian* (2020).

99. Can, Ö. E., D'Cruze, N. & Macdonald, D. W. Dealing in deadly pathogens: Taking stock of the legal trade in live wildlife and potential risks to human health. *Glob Ecol Conserv* **17**, e00515 (2019).

100. Lam, T. T.-Y. *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).

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

102. Tsuda, S. *et al.* Genomic and serological detection of bat coronavirus from bats in the Philippines. *Arch. Virol.* **157**, 2349–2355 (2012).

103. Olival, K. J. *et al.* Possibility for reverse zoonotic transmission of SARS-CoV-2 to free-ranging wildlife: A case study of bats. *PLoS Pathog.* **16**, e1008758 (2020).

104. Anthony, S. J. *et al.* Coronaviruses in bats from Mexico. *J. Gen. Virol.* **94**, 1028–1038

(2013).

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

106. Zhou, H. *et al.* A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr. Biol.* **30**, 2196–2203.e3 (2020).

107. Hul, V. *et al.* A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *bioRxiv* 2021.01.26.428212 (2021) doi:10.1101/2021.01.26.428212.

108. Mou, H. *et al.* Mutations from bat ACE2 orthologs markedly enhance ACE2-Fc neutralization of SARS-CoV-2. *Cold Spring Harbor Laboratory* 2020.06.29.178459 (2020) doi:10.1101/2020.06.29.178459.

109. Wacharapluesadee, S. *et al.* Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* **12**, 972 (2021).

110. Pulliam, J. R. C. *et al.* Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. *J. R. Soc. Interface* **9**, 89–101 (2012).

111. Plowright, R. K. *et al.* Ecological dynamics of emerging bat virus spillover. *Proceedings of the Royal Society B.* **282**, 20142124 (2015).

112. Kessler, M. K. *et al.* Changing resource landscapes and spillover of henipaviruses. *Ann. N. Y. Acad. Sci.* **1429**, 78–99 (2018).

113. Peel, A. J. *et al.* Coronaviruses and Australian bats: a review in the midst of a pandemic. *Aust. J. Zool.* (2020) doi:10.1071/ZO20046.

114. Bordes, F., Blasdell, K. & Morand, S. Transmission ecology of rodent-borne diseases: New frontiers. *Integr. Zool.* **10**, 424–435 (2015).

115. Ostfeld, R. S., Canham, C. D., Oggenfuss, K., Winchcombe, R. J. & Keesing, F. Climate, deer, rodents, and acorns as determinants of variation in lyme-disease risk. *PLoS Biol.* **4**, e145 (2006).

116. Machtinger, E. T. & Williams, S. C. Practical Guide to Trapping Peromyscus leucopus (Rodentia: Cricetidae) and Peromyscus maniculatus for Vector and Vector-Borne Pathogen Surveillance and Ecology. *J. Insect Sci.* **20**, (2020).

117. Morand, S. *et al.* Global parasite and Rattus rodent invasions: The consequences for rodent-borne diseases. *Integr. Zool.* **10**, 409–423 (2015).

118. Hamdan, N. E. S. *et al.* Rodent Species Distribution and Hantavirus Seroprevalence in Residential and Forested areas of Sarawak, Malaysia. *Trop Life Sci Res* **28**, 151–159 (2017).

119. Louys, J. *et al.* Expanding population edge craniometrics and genetics provide insights into dispersal of commensal rats through Nusa Tenggara, Indonesia. *Rec. Aust. Mus.* **72**, 287–302 (2020).

120. Bertolino, S. & Lurz, P. W. W. Callosciurussquirrels: worldwide introductions, ecological impacts and recommendations to prevent the establishment of new invasive populations: Worldwide introductions ofCallosciurussquirrels. *Mamm. Rev.* **43**, 22–33 (2013).

121. Mildenstein, T., Tanshi, I. & Racey, P. A. Exploitation of Bats for Bushmeat and Medicine. in *Bats in the Anthropocene: Conservation of Bats in a Changing World* (eds. Voigt, C. C. & Kingston, T.) 325–375 (Springer International Publishing, 2016). doi:10.1007/978-3-319-25220-9_12.

122. Ransaleleh, T. A. *et al.* Identification of bats on traditional market in dumoga district, North Sulawesi. *IOP Conf. Ser.: Earth Environ. Sci.* **473**, 012067 (2020).

123. Tampon, N. V. T. *et al.* First molecular evidence for bat betacoronaviruses in Mindanao. *Philipp J. Sci.* **149**, 91–94 (2020).

124. Logeot, M. *et al.* Risk assessment of SARS-CoV-2 infection in free-ranging wild animals in Belgium. *Transbound. Emerg. Dis.* (2021) doi:10.1111/tbed.14131.

125. Weber, A., Kalema-Zikusoka, G. & Stevens, N. J. Lack of Rule-Adherence During Mountain Gorilla Tourism Encounters in Bwindi Impenetrable National Park, Uganda, Places Gorillas

at Risk From Human Disease. *Front Public Health* **8**, 1 (2020).

126. Barbosa, A. *et al.* Risk assessment of SARS-CoV-2 in Antarctic wildlife. *Sci. Total Environ.* **755**, 143352 (2021).

127. Gibbons, A. Captive gorillas test positive for coronavirus. *Science* (2021) doi:10.1126/science.abg5458.

128. Daly, N. First great apes at U.S. zoo receive COVID-19 vaccine made for animals. *National Geographic* (2021).

129. Aleccia, J. 'The Biggest Nemesis': Black-Footed Ferrets Get Experimental Coronavirus Vaccine. *Kaiser Health News* (2020).

130. Xu, L. *et al.* COVID-19-like symptoms observed in Chinese tree shrews infected with SARS-CoV-2. *Zool Res* **41**, 517–526 (2020).

131. Bean, A. G. D. *et al.* Studying immunity to zoonotic diseases in the natural host - keeping it real. *Nat. Rev. Immunol.* **13**, 851–861 (2013).

132. Becker, D. J. *et al.* Dynamic and integrative approaches to understanding pathogen spillover. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190014 (2019).

133. Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* (2017) doi:10.1038/nrmicro.2017.45.

134. Morris, D. H. *et al.* The effect of temperature and humidity on the stability of SARS-CoV-2 and other enveloped viruses. *bioRxiv* (2020) doi:10.1101/2020.10.16.341883.

135. Sawatzki, K. *et al.* Ferrets not infected by SARS-CoV-2 in a high-exposure domestic setting. *Cold Spring Harbor Laboratory* 2020.08.21.254995 (2020) doi:10.1101/2020.08.21.254995.

136. Ruiz-Arrondo, I. *et al.* Detection of SARS-CoV-2 in pets living with COVID-19 owners diagnosed during the COVID-19 lockdown in Spain: A case of an asymptomatic cat with SARS-CoV-2 in Europe. *bioRxiv* (2020) doi:10.1101/2020.05.14.20101444.

137. Peel, A. J. *et al.* Synchronous shedding of multiple bat paramyxoviruses coincides with peak periods of Hendra virus spillover. *Emerg. Microbes Infect.* **8**, 1314–1323 (2019).

138. Restif, O. *et al.* Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. *Ecol. Lett.* (2012) doi:10.1111/j.1461-0248.2012.01836.x.