Predicting missing links in global host-parasite networks

Maxwell J. Farrell[1,2,3*], Mohamad Elmasri[4], David Stephens[4] & T. Jonathan Davies[5,6]

[1]Department of Biology, McGill University
[2]Ecology & Evolutionary Biology Department, University of Toronto
[3]Center for the Ecology of Infectious Diseases, University of Georgia
[4]Department of Mathematics & Statistics, McGill University
[5]Botany, Forest & Conservation Sciences, University of British Columbia
[6]African Centre for DNA Barcoding, University of Johannesburg
[*]To whom correspondence should be addressed: e-mail: maxwell.farrell@utoronto.ca

Parasites that infect multiple species cause major health burdens globally, but for many, the full suite of susceptible hosts is unknown. Proactive disease surveillance involves gathering host-parasite association data, predicting missing links, and targeting efforts towards the most likely undocumented interactions. Using the largest global network of mammal host-parasite interactions amalgamated to date (>29,000 interactions), we predict undocumented links and conduct targeted literature searches. We find evidence for many of the top "missing" links, including parasites of humans, domesticated animals, and endangered wildlife, and identify regions such as tropical and central America as likely hotspots of undocumented associations. This approach of iterated prediction and targeted surveillance may help reduce disease burdens by identifying undocumented reservoir species, cryptic transmission pathways, and predicting future host-parasite interactions.

Introduction

Most disease-causing organisms of humans and domesticated animals can infect multiple species (*1, 2*), which has ramifications for biodiversity conservation (*3*) and human health via direct infection, food insecurity, and diminished livelihoods (*4*). The human burden of multi-host diseases falls largely on the world's poor livestock keepers (*5*), many of which live in biodiversity hotspots (*6*) where resources for disease surveillance and reporting are lacking (*7*). Despite the severe burdens they impose, we do not know the full range of susceptible host species for the majority of infectious organisms (*8*). For a given parasite, knowledge of susceptible hosts is critical for understanding disease spread and persistence in multi-host systems (*9*). Predicting the suite of potential host species for known infectious diseases may also allow for more effective disease control and eradication, rapid response following disease emergence (i.e. increases in prevalence, geographic spread, or infection of novel host species), and reducing the risk of disease spillover through limiting cross-species contact of potential hosts (*10, 11, 12, 13*).

While investigation into the specific ecologies, potential reservoirs, and susceptible hosts often takes place after infectious organisms have emerged as public health threats, iterative prediction and verification of likely host species provides one approach to surveillance that can strengthen capacities for disease monitoring and control of multi-host pathogens before emergence (*14*). By predicting undocumented host species for known infectious diseases, we can efficiently gather baseline knowledge of the diversity of host-parasite interactions, ultimately supporting the development of fundamental theory in disease ecology and evolution, and strengthening disease control and surveillance for multi-host pathogens. This form of proactive disease surveillance involves synthesizing current knowledge of host-parasite associations, identifying gaps, predicting missing links, and targeting undocumented interactions underlying contemporary disease burdens, and those which may emerge in the future.

Currently, the host ranges of parasites are best described by global host-parasite interaction databases that are largely compiled from primary academic literature, and used to study macroecological patterns of infectious diseases (*15, 16*) and predict the potential for zoonotic spillover (*17*). However, these global databases are known to be incomplete, with some estimated to be missing up to $\sim$40% of host-parasite interactions among the species sampled (*8*). Missing interactions likely reflect multiple processes, including errors or omissions in database compilation, unpublished or unreported observations, and potential but currently unrealized associations.

Filling in knowledge gaps, and building more comprehensive databases of missing interactions provides substantial insight into the ecological and evolutionary forces shaping host-parasite interactions and disease spread. This includes the role of network structure for transmission (*18, 19*), the nature of "forbidden links" (*20*), and the drivers of parasite richness (*21, 22, 23, 24*) and parasite sharing across hosts (*25, 26, 27, 28*). Futher, determining the diversity and roles of wildlife hosts in disease transmission may help identify factors driving the evolution of highly virulent parasites (*29, 30*). However, updating existing global host-parasite databases requires substantial effort, including conducting systematic and ongoing literature searches across a wide range of taxa, and incorporating any recently published or previously overlooked primary literature. We suggest that an iterative approach to data amalgamation, link prediction, and validation of potential host species offers a cost-effective solution to addressing knowledge gaps in global host-parasite networks, ultimately strengthening capacities for disease monitoring and control of both emerging and neglected multi-host pathogens.

Recent efforts have attempted to predict missing links in global host-parasite networks using machine and statistical learning, such as the identification of potential rodent reservoirs of zoonotic diseases from traits and zoonotic pathogen diversity (*31*). However, we frequently lack the detailed knowledge of ecological and functional traits of interacting species required to

make accurate predictions (*8, 32, 33*). Thus, while these approaches work well for smaller networks (*8*), they scale poorly to global-scale ecological datasets in which comparable traits are unavailable for all species (*20*). Algorithms such as recommender systems can offer flexibility, drawing strength from the size of the network, and allowing for the identification of probable links in a variety of large networks (see (*34*) for a review). Recommender systems based on models of real-world networks attempt to capture two behaviours of many physical networks, the scale-free behaviour of interactions (as seen in Fig. SM 1) and local clustering (*35*), but if not properly informed, these models may predict links that are unlikely or impossible given the ecologies of the interacting species (*36*). Here we employ a recently developed hierarchical Bayesian approach for link prediction in bipartite ecological networks inspired by scale-free recommender systems (*37*). This method is particularly well-suited to link prediction in large global host-parasite networks as it does not require trait data, and generates accurate predictions of missing links in extremely sparse networks using only the structure of the observed host-parasite associations, and evolutionary relationships among hosts.

By incorporating host phylogeny into scale-free interaction models we add biological realism, allowing for more accurate predictions with limited data. Phylogenetic trees are representations of species' evolutionary histories, and branch lengths separating species on their phylogeny provide a measure of expected similarities and differences among species (*38*). In host-parasite systems, hosts may be associated with a parasite through inheritance from a common ancestor, or as a result of parasites shifting to use novel host species (*39*). In both cases we expect closely related species will host similar parasite assemblages (*25*). We can therefore generate ranked lists of highly likely, but currently undocumented hosts of known infectious organisms. Such undocumented links may be prime targets for disease surveillance.

Here, we apply three different link prediction models – one based on node affinities (the numbers of interactions per species), one using only phylogenetic information, and a full model

that layers both components (*37*) (Materials & Methods) – to a global host-parasite interaction matrix comprising > 9,000 parasites of mammals (Fig. 1, Materials & Methods), and show that all models display high accuracy in cross-fold validation (Matherials & Methods, Fig. 2, Table SM 1). The strong performance of the models is especially notable given the large size and extreme sparsity of the interaction network (across the whole dataset only ∼0.17% of the ∼16.8 million possible links have recorded evidence).

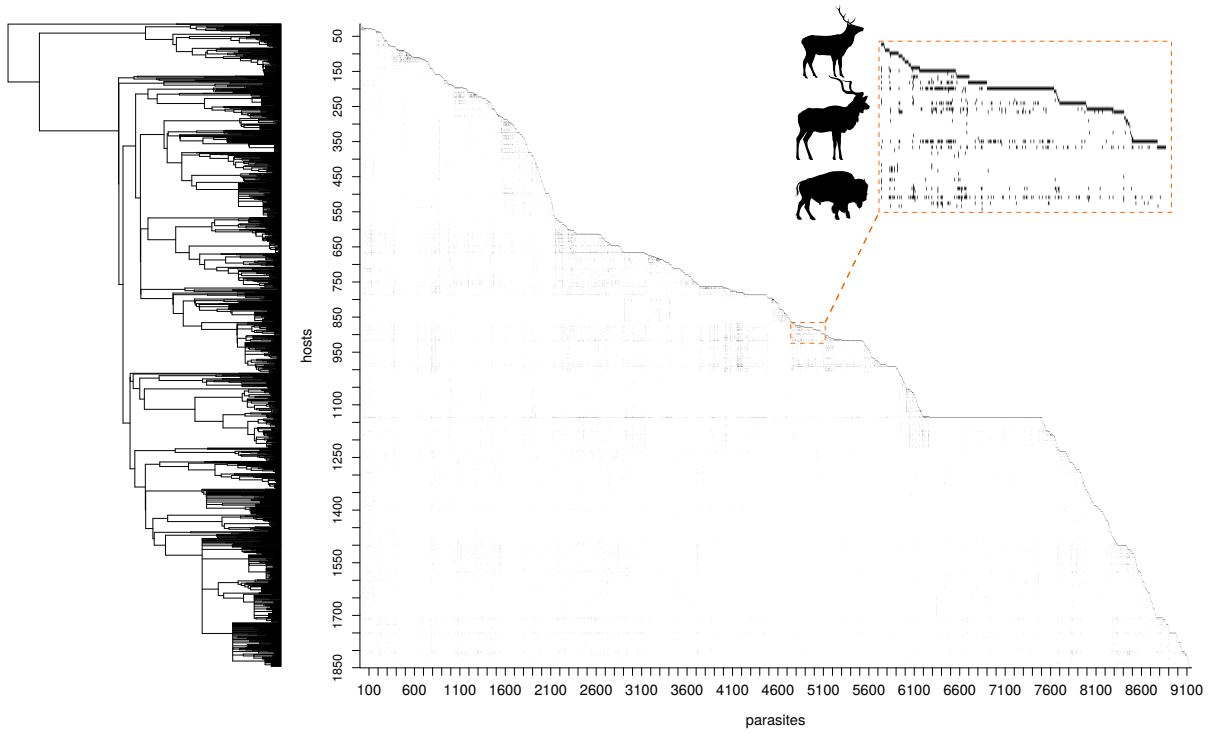

Figure 1: Phylogeny of host species (pruned from the Fritz et al. (*40*) dated supertree for mammals), and host-parasite association matrix showing hosts ordered according to the phylogeny. The size of the matrix does permit Latin binomials to be shown. Instead, axis labels represent indexes for hosts (rows) and parasites. The orange rectangles indicate an expanded subset of the matrix.

Results & Discussion

By conducting targeted literature searches of the top predicted missing links from each of the link prediction models, we show that we are able to successfully predict links with documented support, but are not currently included in major host-parasite databases (See SM for lists of top links and detailed results of literature searches). The top links from the affinity and full models were largely dominated by humans and domesticated animal hosts, while the phylogeny models more often predicted links among wildlife (Fig. 3) including endangered and relatively poorly studied species, some of which are critically endangered (*41*). The full models more often included a larger diversity of parasite species among the top predicted links (Table SM 2), but among all models, parasites infecting large numbers and phylogenetic ranges of hosts were most often included in the top undocumented links (ex. *Rabies lyssavirus*, *Sarcoptes scabiei*, *Toxoplasma gondii*, *Trypanosoma cruzi*). This is not surprising as these parasites are commonly cited as capable of causing disease in a large number of (and sometimes all) mammals; however we were still able to identify many hosts of these parasites that were not included in the original database.

We ran the three models (affinity, phylogeny, and full) on the full dataset (see Fig. 4 for posterior interaction matrices) and identified the top ten links with no evidence in the original database, resulting in 30 links. As humans dominated the top predictions of the affinity model (Fig. 3), we additionally identified the top ten missing links for both domesticated hosts and wild hosts, per model, resulting in an additional 60 links. We also ran the three models for each of the six parasite subsets, again identifying the top ten missing links, resulting in an additional 180 links, with a total of 90 predicted links per model for which to compare model performance based on targeted literature searches. Across the links for which literature searches were conducted, the full model identified a greater number of documented links that were not in the

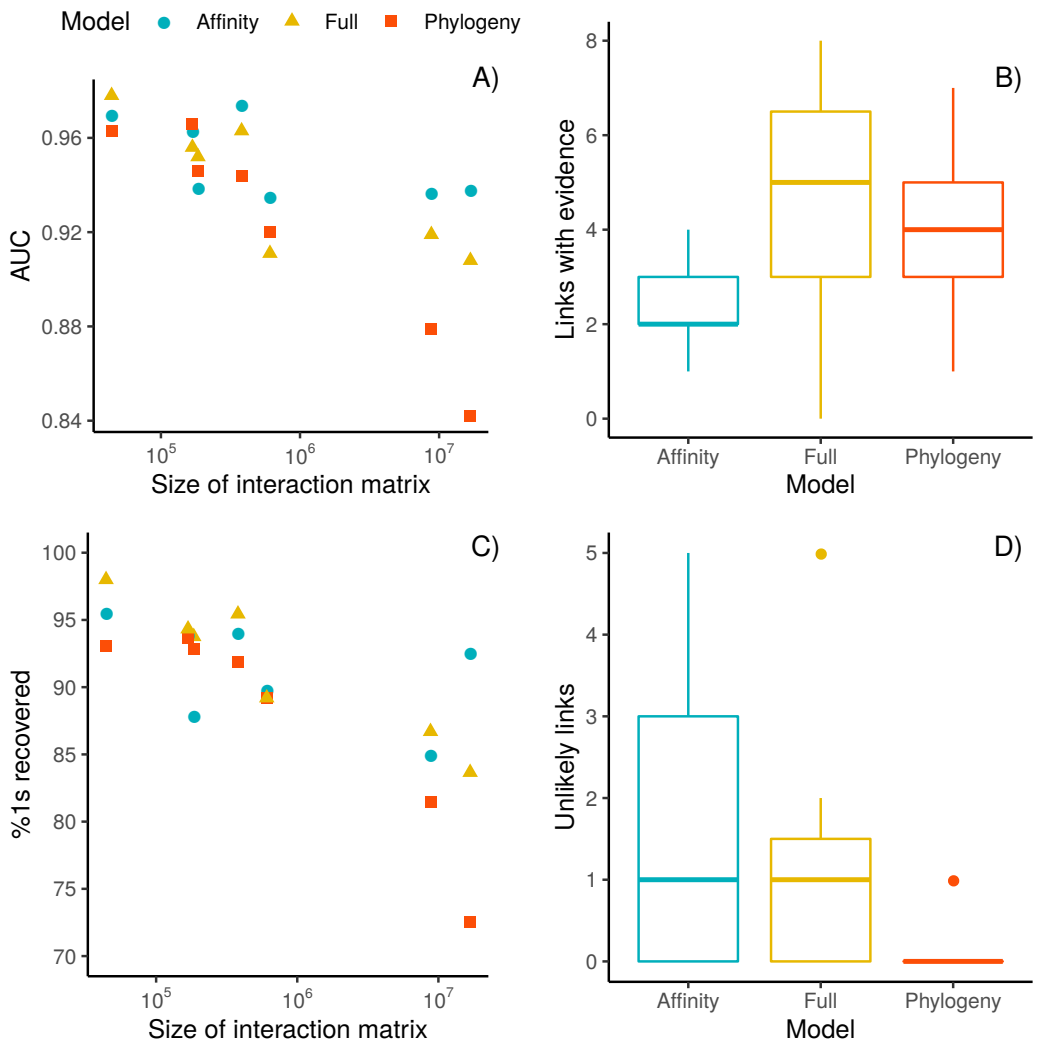

Figure 2: Model diagnostic plots, including predictive performance after 10-fold cross-validation (A,C), and the results of targeted literature searches for the top 10 undocumented links per model (B,D). Panel A shows area under the receiver operating characteristic curve (AUC) in relation to the size of interaction matrix (total number of potential host-parasite links). Panel C shows the percent documented interactions (1s) correctly recovered from the held-out portion in relation to the size of interaction matrix. Panels B and D show boxplots of the number of links with published evidence (B), and the number of unlikely links (D) out of the top ten most probable yet undocumented interactions for each of the three model types (affinity, full, phylogeny) run across the full dataset and each of the models subset by parasite type (Arthropods, Bacteria, Fungi, Helminths, Protozoa, Viruses).

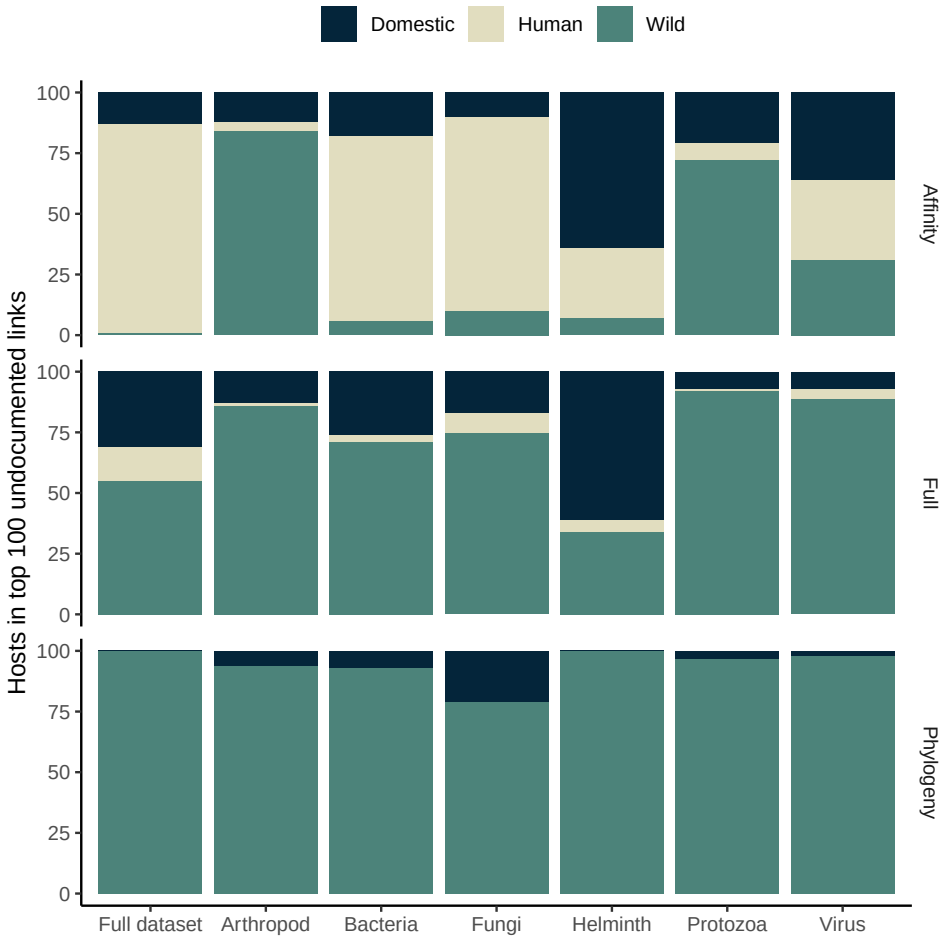

Figure 3: The frequencies of host types (human, domestic, and wildlife) included in the top 100 predicted links which were not documented in the original database. Plots are grouped by data subset (the full dataset, and subsets by parasite type), and model (affinitly, full, and phylogeny).

original database (45/90) compared to the phylogeny (39/90) and affinity (28/90) models (Table SM 2). The affinity model assumes that species with many recorded interactions are more likely to interact with species that also have a large number of recorded interactions. The number of interactions a species is involved in will vary due to ecological or evolutionary reasons, but their documentation may additionally be influenced by research effort (*24*), implying that predictions from the affinity and full models are more likely to reflect study effort when compared to the phylogeny models. While predictive performance (as quantified by AUC and % 1s recov-

ered) was regularly higher for the full and affinity models compared to the phylogeny models (Fig. 2A C, Table SM 1), the phylogeny model was much less prone to predicting ecologically unlikely links (Fig. 2D), indicating that it may better capture the underlying ecological and evolutionary determinants of host-parasite interactions. Further, the top undocumented links for the phylogeny model have substantially lower ranks than those of the affinity and full models (Fig. SM 3), and unlike the other models, all off the top 1000 links predicted by the phylogeny model have already been documented in the original dataset (Fig. SM 4), indicating that a large portion of the top predictions by the phylogeny model accurately reflect previously documented associations.

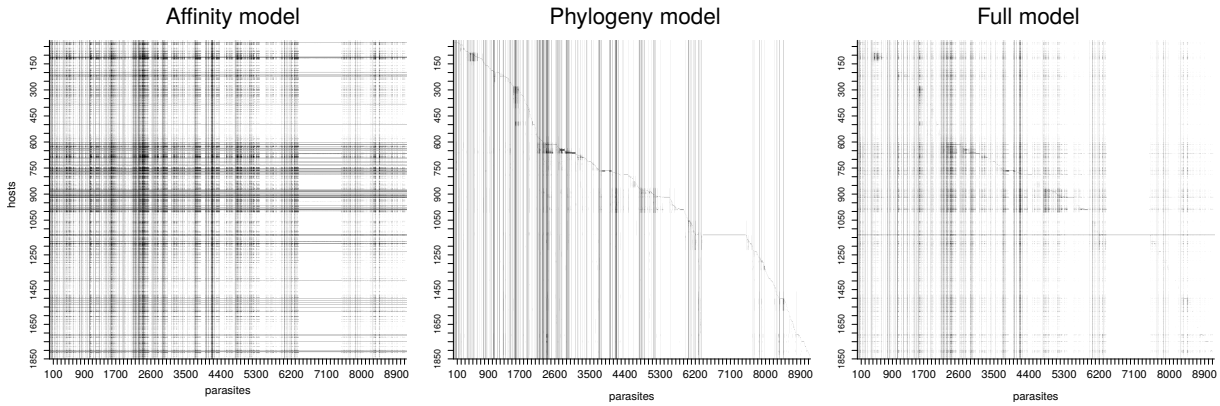

Figure 4: Posterior interaction matrices for the affinity, phylogeny, and full models run on the full dataset. Axis labels represent indexes for hosts (rows) and parasites, with hosts ordered to match Figure 1.

Predictions from the affinity model indicate Europe as a hotspot of undocumented links (Fig. 5), likely representing the volume of research on well-studied taxa in member countries. If common, we might expect that such links would already be well-documented. However, these predictions could be critical for large scale public health initiatives if they implicate widespread or abundant species as reservoirs of emerging infections. Further, the affinity model is more likely to identify parasite sharing among distantly related hosts, which have the potential to

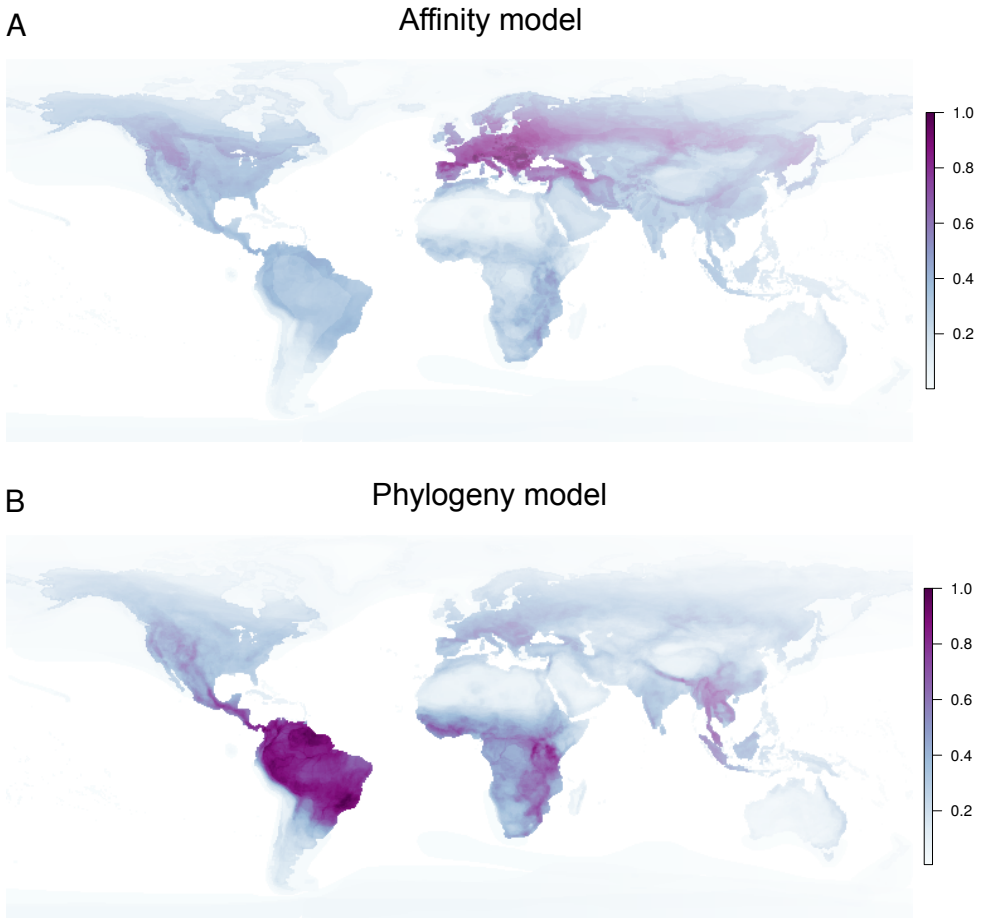

Figure 5: Global hotspot maps of relative risk of undocumented host-parasite interactions. Maps were generated by summing the probabilities of undocumented host-parasite interactions per host species, summing across host species per cell with range maps for terrestrial mammals from the IUCN, then standardized to a scale of 0 - 1. A) represents the relative risk based on the predictions from the affinity model, while B) represents the phylogeny model. The hotspot map generated by the full model was largely comparable to that produced by the affinity model, but with higher relative risk in eastern Africa (Fig. SM 2).

result in high host mortality following host shifts (*30*). Nonetheless, top predictions for the affinity and full models also included multiple links that are unlikely due to a mismatch in host ecology. For example, domestic cattle (*Bos taurus*) are predicted to be susceptible to infection by *Anisakis simplex*. *A. simplex* is a trophically transmitted nematode that uses aquatic mammals as final hosts, with marine invertebrates and fish as intermediate hosts (*42*), implying

that cattle may only be exposed to the parasite if fed a marine-based diet.

Hotspots of predicted links from the phylogeny model show a spatial distribution in stark contrast to the affinity model, with highest density in tropical and central America, followed by tropical Africa and Asia (Fig. 5). It is perhaps no suprise that these under-studied regions, with high host and parasite diversity, are identified as centres of undocumented host-parasite associations. We suggest these links might capture both existing interactions among rarely studied taxa, and risks for future host shifts. Undocumented links highlighted by the phylogeny models also included fewer ecologically unlikely interactions, with only one unlikely interaction among the top predicted links we investigated: *Echinococcus granulosus* is typically maintained by a domestic cycle of dogs eating raw livestock offal (*43*), and while wild canids such as *Lycalopex gymnocercus* are known hosts (*44*), the model predicts *Lycalopex vetulus* to be susceptible. However, this interaction is unlikely as *L. vetulus* has a largely insectivorous diet (*45*), unlike the other members of its genus.

In the phylogeny model, we use information on the evolutionary relationships among taxa to improve predictions, however, the flexibility of the method allows for any information, such as trait or geographic data, to be incorporated as long as they can be represented by a distance matrix (*37*). With this approach, future models may be expanded to help exclude ecological mismatches not readily captured by host phylogeny. Similarly, the model could be extended to incorporate weighted rather than binary associations, allowing for modelling links as a function of prevalence, intensity of infection, or to explicitly incorporate the amount of evidence supporting each link. In this way sampling intensity may be directly incorporated into link predictions, and help identify weakly supported interactions or sampling artefacts that may benefit from additional investigation.

Because we did not include geographic constraints in our model, we may predict interactions that are unlikely to exist due to lack of geographic overlap, such as *T. cruzi*, which is currently

restricted to the Americas (*46*), infecting endangered African species such as black rhinoceros (*Diceros bicornis*), lowland gorilla (*Gorilla gorilla*), and chimpanzee (*Pan troglodytes*). Although natural infections of chimpanzees by this parasite are unlikely, we found a report of a fatal infection of a captive individual in Texas (*47*). In addition, we identified multiple infections documented only in captive animals, and while we did not find evidence of these infections in nature, they demonstrate that the model is able to identify biologically plausible infection risks that are relevant for captive populations, and which may represent future risks in the face of host or parasite translocation.

The importance of infectious diseases in conservation is often hampered by our lack of knowledge about the diversity of pathogens in natural systems (*3*). Applying link prediction methods to global host-parasite networks can highlight both historical and contemporary disease threats to biodiversity, and identify parasites with the potential to drive endangered species towards extinction. For example, the phylogeny models identified links reported only in literature from over 30 years ago, such as *T. cruzi* in the critically endangered cotton-top tamarin (*Saguinus oedipus*) (*48*) and the vulnerable black-crowned Central American squirrel monkey (*Saimiri oerstedii*) (*49*). Our guided literature search also found evidence of severe infections in several endangered species such as rabies and sarcoptic mange (*Sarcoptes scabiei*) in Dhole (*Cuon alpinus*) (*50*) and *Toxoplasma gondii* in critically endangered African wild dogs (*Lycaon pictus*) which caused a fatal infection in a pup (*51*). Our model also predicts that rabies and sarcoptic mange are likely to infect the endangered Darwin's fox (*Lycalopex fulvipes*). Disease spread via contact with domestic dogs (notably *Canine distemper virus*) is currently one of the main threats to this species (*52*). Considering that both rabies and sarcoptic mange from domestic dogs are implicated in the declines of other wild canids, they may pose a serious risk for the conservation of Darwin's foxes.

Our approach identified a number of parasites that could have notable impacts on public

health of humans and domesticated animals. These include parasites currently considered a risk for zoonotic transmission such as *Alaria alata*, an intestinal parasite of wild canids – a concern as other *Alaria* species have been reported to cause fatal illness in humans (*53*), and *Bovine viral diarrhea virus 1*, which is not currently considered to be a human pathogen, but is highly mutable, has the ability to replicate in human cell lines, and has been isolated from humans on rare occasions (*54*). However, there is a large amount of effort that goes into studying infectious diseases of humans and domestic species, and it is likely that most contemporary associations among humans and described parasites have been recorded, even if not included in the aggregated databases because they occur rarely or are difficult to detect. For example, we predicted that humans should be infected by *Bartonella grahamii*, and found that the first recorded case was in an immunocompromised patient in 2013 (*55*). Similarly, humans are predicted to be susceptible to *Mycoplasma haemofelis*, which was again reported in someone who was immunocompromised (*56*), indicating that while these infections may pose little risk for a large portion of the human population, they are a serious concern for the health of immunocompromised individuals. These examples demonstrate our framework has the capacity to predict known human diseases, highlight parasites that are recognized zoonotic risks, and identify a number of parasites that are currently unrecognized as zoonotic risks.

Link prediction in global host-parasite networks is the first step in an iterative process of prediction and verification whereby likely links are identified, queried, and new links are added, allowing predictions to be updated (Fig. 6). With this approach we can efficiently build more comprehensive global host-parasite interaction databases. For example, we identified a number of interactions that were documented for the first time after the input databases were assembled (e.g. *T. cruzi* in horses (*57*), *Nematodirus spathiger* in *Gazella leptoceros* (*58*), *Toxoplasma gondii* in *Papio anubis* (*59*)). As we move down the list of most probable links, we will uncover links with infectious organisms that are less well studied, but which may emerge as public

health burdens in the future. These links represent key targets for disease surveillance. By disseminating predictions to veterinarians, conservation managers, and public health officials, we can spread awareness of potential threats and enhance surveillance by incorporating these taxa within existing disease and biodiversity monitoring programs. An important next step will be to move beyond the binary notion of infection used here and attempt to classify the nature of the association between host and parasite. In this way we may be able to predict not only the presence or absence of a particular host-parasite interaction, but the epidemiological role each host plays in parasite transmission, the impact of infection on host fitness (*30*), and better understand the ecologies of reservoir versus spillover hosts.

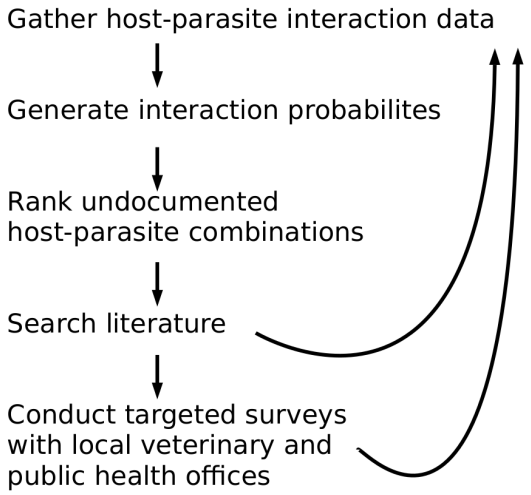

Figure 6: Conceptual outline of integrating link prediction and literature searching into proactive disease surveillance.

Conclusion

We have demonstrated that missing links in global databases of host-parasite associations can be predicted accurately using information on known associations and the evolutionary relationships among host species. Applying this method to host-parasite associations for mammal

host species we are able to make robust predictions even with sparse input data, indicating that this method may be applied to disparate host-parasite systems. We suggest that link prediction represents a cost-effective approach for supporting disease surveillance and for spreading awareness of future disease risks for humans, domestic animals, and wildlife. Global change in the form of shifting climates and alteration of natural habitats has the potential to bring into contact previously isolated host and parasite populations, increasing opportunities for disease spillover. Strengthening baseline knowledge of the diversity of host-parasite interactions is an essential step for building effective approaches to mitigate the impacts of multi-host diseases.

Materials & Methods

Data

We aimed to generate a list of highly probable, yet previously undocumented host-parasite interactions for humans, domesticated animals, and wildlife. First, we identified four recently published global host-parasite databases for mammals derived from primary literature, genetic sequence databases, and natural history collections, and which report host and parasite names as Latin binomials. The Global Mammal Parasite Database 2.0 (GMPD) (*15*) contains records of disease causing organisms (viruses, bacteria, protozoa, helminths, arthropods, and fungi) in wild ungulates (artiodactyls and perrisodactyls), carnivores, and primates drawn from over 2700 literature sources published through 2010 for ungulates and carnivores, and 2015 for primates. The static version of the Enhanced Infectious Disease Database (EID2) (*60*) contains 22,515 host-pathogen interactions from multiple kingdoms based on evidence published between 1950 - 2012 extracted from the NCBI Taxonomy database, NCBI Nucleotide database, and PubMed citation and index. The Host-Parasite Database of the Natural History Museum, London (*61*) contains over a quarter of a million host-parasite records for helminth parasites extracted from 28,000 references published after 1922, and is digitally accessible via the R package *helminthR*

(*62*). Finally, Olival et al. (*17*) compiled a database of 2,805 mammal-virus associations for every recognized virus found in mammals, representing 586 unique viral species and hosts from 15 mammalian orders.

To amalgamate these databases, host names were standardized to those in Wilson & Reeder (*63*) and used in the Fritz et al. (*40*) mammal phylogeny. Virus names were standardized to the 2015 version of the International Convention on Viral Taxonomy (*64*). For non-viral parasites there exists no single accepted authority for nomenclature or species concepts, however all parasite names were thoroughly checked for typographical and formatting errors. When potentially synonymous names were identified (e.g. *Cylicocyclus insigne* and *Cylicocyclus insignis*), online searches of primary literature and taxonomic databases including the Integrated Taxonomic Information System (www.itis.gov), Catalogue of Life (www.catalogueoflife.org), World Register of Marine Species (www.marinespecies.org), Encyclopedia of Life (www.eol.org), NCBI Taxonomy Database (www.ncbi.nlm.nih.gov/taxonomy), UniProt (www.uniprot.org), and the Global Names Resolver (resolver.globalnames.org) were conducted to resolve taxonomic conflicts. Synonymous names were corrected to the name with the majority of references, or to the preferred name in recently published literature or taxonomic revision when this information was available. Host associations for parasite names that were later split into multiple species were removed from the dataset (e.g. *Bovine papillomavirus*). All hosts and parasites reported below the species level were assigned to their respective species (e.g. *Alcelaphus buselaphus jacksoni* was truncated to *Alcelaphus buselaphus*), and any species reported only to genus level or higher (e.g. *Trichostrongylus sp.*) was removed. Our requirement for source databases to report Latin binomials facilitated taxonomy harmonization, but resulted in the exclusion of databases such as the GIDEON, which is a useful resource for modelling outbreaks of human infections based on disease common names, but does not provide unambiguous parasite scientific names (*65*). Of the 215 human diseases reported in Smith et al. (2014), we identified 197 that could be

attributed to a predominant causal agent or group of organisms, all of which were already represented in our more comprehensive dataset. The remaining 18 human diseases represented those with no known causes (e.g. Brainerd diarrhea, Kawasaki disease), and those regularly caused by a diverse range of pathogens (e.g. viral conjuntivitis, invasive fungal infection, tropical phagedenic ulcer)."

The resulting network includes 29,112 documented associations among 1835 host and 9149 parasite species (Fig. 1). To our knowledge this constitutes the largest host-parasite interaction database assembled for mammals, and includes parasites from diverse groups including viruses, bacteria, protozoa, helminths, arthropods, and fungi, and wild, domestic, and human hosts. The resulting matrix is quite sparse, with $\sim$0.17% of the $\sim$16.8 million possible links having documented interactions. Humans are documented to associate with 2,064 parasites (47% of which associate with another mammal in the database), and comprise 7% of all interactions. Parasite species are largely represented by helminths (63.9%), followed by bacteria (13.1%) and viruses (7.89%). The number of documented interactions per species (degree distribution) varies considerably and is shown to be linear on the log scale for both hosts and parasites (Fig. SM 1). This database comprises presence-only data as we do not have consistent information about the strength or nature of each interaction, only that it has been documented at least once by any accepted method (direct observation, genetic sequencing, or serology). Therefore interactions are taken as binary (0/1 for a given host-parasite interaction). Absences should not be considered true absences and are likely to include some host-parasite associations that (1) have been observed but which are not recorded in the original database, (2) currently exist but have not yet been observed or are undocumented, and that (3) may be realized in the future given sufficient opportunity. Our link prediction framework attempts to identify these cases.

Statistical analyses

We apply the link prediction model of Elmasri et al. (*37*) to the amalgamated dataset of 29,112 documented and $\sim$16.8 million potential host-parasite interactions. The model has three variants: the "affinity" model which generates predictions based only on the number of observed interactions for each host and parasite, the "phylogeny" model which is informed only by host evolutionary relationships, and the "full" model which layers both components. The affinity model is fit by preferential attachment whereby hosts and parasites that have many interacting species in the network are assigned higher probabilities of forming novel interactions. The phylogeny model uses the similarity of host species based on evolutionary distances to assign higher probability to parasites interacting with hosts closely related to their documented host species, and lower probability of interacting with hosts that are distantly related. To account for uncertainty in the phylogeny, and allow the model to place more or less emphasis on recent versus deeper evolutionary relationships, we fit a tree scaling parameter (η) based on an accelerating-decelerating model of evolution (*66*). This transformation, which allows for changes in the relative evolutionary distances among hosts, was shown to have good statistical properties for link prediction in a subset of the GMPD (*37*). We apply all three model to the full dataset. The tree scaling parameter is applied across the whole phylogeny, but since the importance of recent versus deep evolutionary relationships among hosts is likely to vary across parasite types (*67*), we additionally run the models on the dataset subset by parasite taxonomy (arthropods, bacteria, fungi, helminths, protozoa, and viruses).

For each model we determined the number of iterations required for parameter convergence by visual inspection of parameter traceplots, auto-correlation plots, and effective sample size (see Elmasri et al. (*37*) for detailed discussion of convergence diagnostics). For parameter estimation and evaluation of model performance we ran each model using 10-fold cross validation holding out links for which there is a minimum of two observed interactions (the model would

not be able to recover interactions for parasites that infect a single host species). To assess predictive performance and prevent overfitting, each model was evaluated using the average of these 10-fold cross-validations. For each fold, we set a different set of the observed interactions (1s) to unknowns (0s), and attempted to predict them using a model fit to the remaining interactions. To assess predictive performance, we calculated the area under the receiver operating characteristic (ROC) curve, which is a popular measure of potential predictive ability for binary outcomes. For each model, an ROC curve is obtained by thresholding the predictive probabilities in each cross-validation, calculating the true positive and false positive rates per fold, and taking the average across the 10 folds. In this way the posterior predictive interaction matrix is generated at the threshold value that maximizes the area under the ROC curve (AUC). However, because our study is motivated by the belief that some 0s in our data are actually unobserved 1s, AUC may not be the best metric for comparing model performance as it increases when models correctly recover observed 0s. Therefore, in addition to AUC, we also assess models based on the percent of 1s accurately recovered.

Targeted literature searches

As additional validation, and to determine the utility of the model, we identify the top 10 most likely links in each model iteration (model type and data subset) which were not documented in the database and conducted online searches of primary and grey literature for evidence of these associations. Searches were conducted in Google Scholar by using both the host and parasite Latin binomials in quotes and separated by the AND boolean operator (e.g. "Gazella leptoceros" AND "Nematodirus spathiger"). If this returned no hits, the same search was conducted using the standard Google search engine in order to identify grey literature sources. For models run on the full dataset, we also investigate the top ten links for domesticated mammals (*Bison bison*, *Bos sp.*, *Bubalus bubalis*, *Camelus sp.*, *Capra hircus*, *Canis lupus*, *Cavia porcellus*, *Equus*

asinus, *Equus caballus*, *Felis catus*, *Felis silvestris*, *Lama glama*, *Mus musculus*, *Oryctolagus cuniculus*, *Ovis aries*, *Rangifer tarandus*, *Rattus norvegicus*, *Rattus rattus*, *Sus scrofa*, *Vicugna vicugna*), and wild host species separately. For domesticated animals and humans, if the Latin binomials returned no hits, the search strategy was repeated using host common names (e.g. "human", "pig", "horse").

Together we ran the three models on the full dataset and identified the global top ten links, plus the top ten links for domesticated hosts and wild hosts, returning 90 links. We also ran the three models for each of the six parasite subsets, returning 180 links. Confirmed evidence of infection was taken as physical, genetic, or serological identification of a parasite infecting a given host species, except in situations where the authors explicitly state that their parasite identification was uncertain due to known serological cross-reactivity, absence of clear genetic similarity to known reference sequences, or unconfirmed visual diagnosis made from afar. Missing links were deemed unlikely when they involved an ecological mismatch between host and parasite, such as the trophically transmitted parasites infecting the wrong trophic level or unlikely ingestion pathway, infections of captive hosts outside of their native ranges, and parasites that are typically considered to be non-pathogenic or commensal. However, lack of known current geographic overlap among wide ranging hosts and parasites was not considered as enough evidence to define a link as unlikely. In total we extracted 270 highly likely undocumented links to target; however, as there was some overlap in the top links among models and data subsets, this resulted in 177 unique links that we investigated for published evidence of infection.

Data & code

The amalgamated host-parasite matrix, results of all models, top predictions from each model, and R scripts to format predictions and reproduce the figures can be found at doi: 10.6084/m9.figshare.8969882. Code used to run the model is available at `github.com/melmasri/HP-prediction`.

Acknowledgements

This work would not exist without the McGill Statistics-Biology Exchange (S-BEX) organized by Zofia Taranu, Amanda Winegardner, and Russel Steele. Additional thanks are due to Will Pearse, Ignacio Morales-Castilla, and Klaus Schliep for support and advice. MJF was supported by a Vanier NSERC CGS, the CIHR Systems Biology Training Program, the Quebec Centre for Biodiversity Science, and the McGill Biology Department. ME was supported by the McGill Department of Mathematics and Statistics, FRQNT, and an NSERC PDF.

Author Contributions

MF and JD designed the study, ME, MF, JD, and DS designed the methods, MF compiled the data, MF and ME conducted the analyses, MF wrote the manuscript with input from JD. The authors declare no conflicts of interest.

References

1. S. Cleaveland, M. K. Laurenson, L. H. Taylor, *Philosophical Transactions of the Royal Society B: Biological Sciences* **356**, 991 (2001).

2. L. H. Taylor, S. M. Latham, M. E. Woolhouse, *Philosophical Transactions of the Royal Society B: Biological Sciences* **356**, 983 (2001).

3. K. F. Smith, K. Acevedo-Whitehouse, A. B. Pedersen, *Animal Conservation* **12**, 1 (2009).

4. D. Grace, J. Gilbert, T. Randolph, E. Kang'ethe, *Tropical Animal Health and Production* **44**, 67 (2012).

5. International Livestock Research Institute, *Zoonoses Report 4* pp. 1–119 (2012).

6. B. Fisher, T. Christopher, *Ecological Economics* **62**, 93 (2007).

7. J. Halliday, *et al.*, *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 2872 (2012).

8. T. Dallas, S. Huang, C. Nunn, A. W. Park, J. M. Drake, *Proceedings of the Royal Society B: Biological Sciences* **284** (2017).

9. M. Viana, *et al.*, *Trends in Ecology & Evolution* **29**, 270 (2014).

10. P. Roeder, J. Mariner, R. Kock, *Philosophical Transactions of the Royal Society B: Biological Sciences* **368** (2013).

11. R. S. Miller, M. L. Farnsworth, J. L. Malmberg, *Preventive Veterinary Medicine* **110**, 119 (2013).

12. P. Büscher, *et al.*, *Trends in Parasitology* **34**, 197 (2018).

13. D. Molyneux, D. P. Sankara, *PLoS Neglected Tropical Diseases* **11**, 4 (2017).

14. M. J. Farrell, L. Berrang-Ford, T. J. Davies, *Environmental Research Letters* **8**, 015036 (2013).

15. P. R. Stephens, *et al.*, *Ecology letters* **19**, 1159 (2016).

16. P. R. Stephens, *et al.*, *Ecology* pp. 42–49 (2017).

17. K. J. Olival, *et al.*, *Nature* **546**, 646 (2017).

18. J. M. Gomez, C. L. Nunn, M. Verdu, *Proceedings of the National Academy of Sciences* **110**, 7738 (2013).

19. S. Pilosof, S. Morand, B. R. Krasnov, C. L. Nunn, *PLoS ONE* **10** (2015).

20. I. Morales-Castilla, M. G. Matias, D. Gravel, M. B. Araújo, *Trends in Ecology & Evolution* **30**, 347 (2015).

21. C. L. Nunn, S. Altizer, K. E. Jones, W. Sechrest, *The American Naturalist* **162**, 597 (2003).

22. S. Huang, J. M. Drake, J. L. Gittleman, S. Altizer, *Evolution* **69**, 621 (2015).

23. V. O. Ezenwa, S. A. Price, S. Altizer, N. D. Vitone, K. C. Cook, *Oikos* **115**, 526 (2006).

24. T. Kamiya, K. O'Dwyer, S. Nakagawa, R. Poulin, *Biological Reviews of the Cambridge Philosophical Society* **89**, 123 (2014).

25. T. J. Davies, A. B. Pedersen, *Proceedings of the Royal Society B: Biological Sciences* **275**, 1695 (2008).

26. S. Huang, O. R. P. Bininda-Emonds, P. R. Stephens, J. L. Gittleman, S. Altizer, *Journal of Animal Ecology* **83**, 671 (2014).

27. M. P. Braga, E. Razzolini, W. A. Boeger, *Journal of Animal Ecology* **84**, 487 (2015).

28. A. D. Luis, *et al.*, *Ecology Letters* **18**, 1153 (2015).

29. E. K. Shwab, *et al.*, *Proceedings of the National Academy of Sciences* **115**, 201722202 (2018).

30. M. J. Farrell, T. J. Davies, *Proceedings of the National Academy of Sciences* **116**, 7911 (2019).

31. B. A. Han, J. P. Schmidt, S. E. Bowden, J. M. Drake, *Proceedings of the National Academy of Sciences* **112**, 7039 (2015).

32. D. Gravel, T. Poisot, C. Albouy, L. Velez, D. Mouillot, *Methods in Ecology and Evolution* **4**, 1083 (2013).

33. I. Bartomeus, *et al.*, *Functional Ecology* **30**, 1894 (2016).

34. F. Ricci, L. Rokach, B. Shapira, *Introduction to Recommender Systems Handbook* (Springer, 2011).

35. D. J. Watts, S. H. Strogatz, *Nature* **393**, 440 (1998).

36. P. Jordano, J. Bascompte, J. M. Olesen, *Ecology Letters* **6**, 69 (2003).

37. M. Elmasri, M. J. Farrell, T. J. Davies, D. A. Stephens, *Annals of Applied Statistics* **in press** (2020).

38. J. J. Wiens, *et al.*, *Ecology Letters* **13**, 1310 (2010).

39. R. D. Page, *International Journal for Parasitology* **23**, 499 (1993).

40. S. A. Fritz, O. R. P. Bininda-Emonds, A. Purvis, *Ecology letters* **12**, 538 (2009).

41. IUCN, The IUCN Red List of Threatened Species. (2019).

42. K. Buchmann, F. Mehrdana, *Food and Waterborne Parasitology* **4**, 13 (2016).

43. B. Otero-Abad, P. R. Torgerson, *PLoS Neglected Tropical Diseases* **7** (2013).

44. M. Lucherini, E. M. Luengos Vidal, *Mammalian Species* **820**, 1 (2008).

45. J. C. Dalponte, *Mammalian Species* **847**, 1 (2009).

46. A. J. Browne, *et al.*, *Scientific Data* **4**, 1 (2017).

47. Y. R. Bommineni, E. J. Dick, J. S. Estep, J. L. Van de Berg, G. B. Hubbard, *Journal of Medical Primatology* **38**, 247 (2009).

48. C. J. Marinkelle, *Annals of Tropical Medicine & Parasitology* **76**, 121 (1982). PMID: 6807227.

49. O. E. Sousa, *Revista de biologia tropical* **20**, 167 (1972).

50. L. Durbin, A. Venkataraman, S. Hedges, W. Duckworth, *Canids: Foxes, Wolves, Jackals and Dogs*, C. Sillero-Zubiri, M. Hoffmann, D. Macdonald, eds. (Island Press, 2005), chap. 8, pp. 210–219.

51. J. Van Heerden, M. G. Mills, M. J. Van Vuuren, P. J. Kelly, M. J. Dreyer, *Journal of the South African Veterinary Association* **66**, 18 (1995).

52. E. Silva-Rodríguez, *et al.*, Lycalopex fulvipes. The IUCN Red List of Threatened Species 2016 (2016).

53. T. M. Murphy, *et al.*, *Parasitology Research* **111**, 283 (2012).

54. P. H. Walz, *et al.*, *Journal of Veterinary Internal Medicine* **24**, 476 (2010).

55. J. Oksi, *et al.*, *Journal of Clinical Microbiology* **51**, 2781 (2013).

56. A. P. dos Santos, *et al.*, *Emerging Infectious Diseases* **14**, 1922 (2008).

57. L. K. Bryan, *et al.*, *Veterinary Parasitology* **216**, 13 (2016).

58. Y. Said, M. Gharbi, M. Mhadhbi, M. Dhibi, S. Lahmar, *Parasitology* **145**, 901 (2018).

59. D. Kamau, *et al.*, *The 11th Jkuat Scientific, Technological and Industrialization Conference and Exhibitions Conference Proceedings* (2016).

60. M. Wardeh, C. Risley, M. K. Mcintyre, C. Setzkorn, M. Baylis, *Scientific Data* **2**, 150049 (2015).

61. D. I. Gibson, R. A. Bray, E. A. C. Harris, Host-Parasite Database of the Natural History Museum, London (2005).

62. T. Dallas, *Ecography* **39**, 391 (2016).

63. D. E. Wilson, D. M. Reeder, *Mammal Species of the World: A Taxonomic and Geographic Reference* (Johns Hopkins University Press, Baltimore, Maryland, 2005), third edn.

64. E. J. Lefkowitz, *et al.*, *Nucleic Acids Research* **46**, D708 (2018).

65. K. F. Smith, *et al.*, *Journal of the Royal Society Interface* **11**, 1 (2014).

66. L. J. Harmon, *et al.*, *Evolution* **64**, 2385 (2010).

67. A. W. Park, *et al.*, *Proceedings of the Royal Society B: Biological Sciences* **285** (2018).

Predicting missing links in global host-parasite networks
Supplementary Materials

Maxwell J. Farrell[1,2*], Mohamad Elmasri[3], David Stephens[3] & T. Jonathan Davies[4,5]

[1]Ecology & Evolutionary Biology Department, University of Toronto
[2]Center for the Ecology of Infectious Diseases, University of Georgia
[3]Department of Mathematics & Statistics, McGill University
[4]Botany, Forest & Conservation Sciences, University of British Columbia
[5]African Centre for DNA Barcoding, University of Johannesburg
[*]To whom correspondence should be addressed: e-mail: maxwell.farrell@utoronto.ca

Supplementary Results

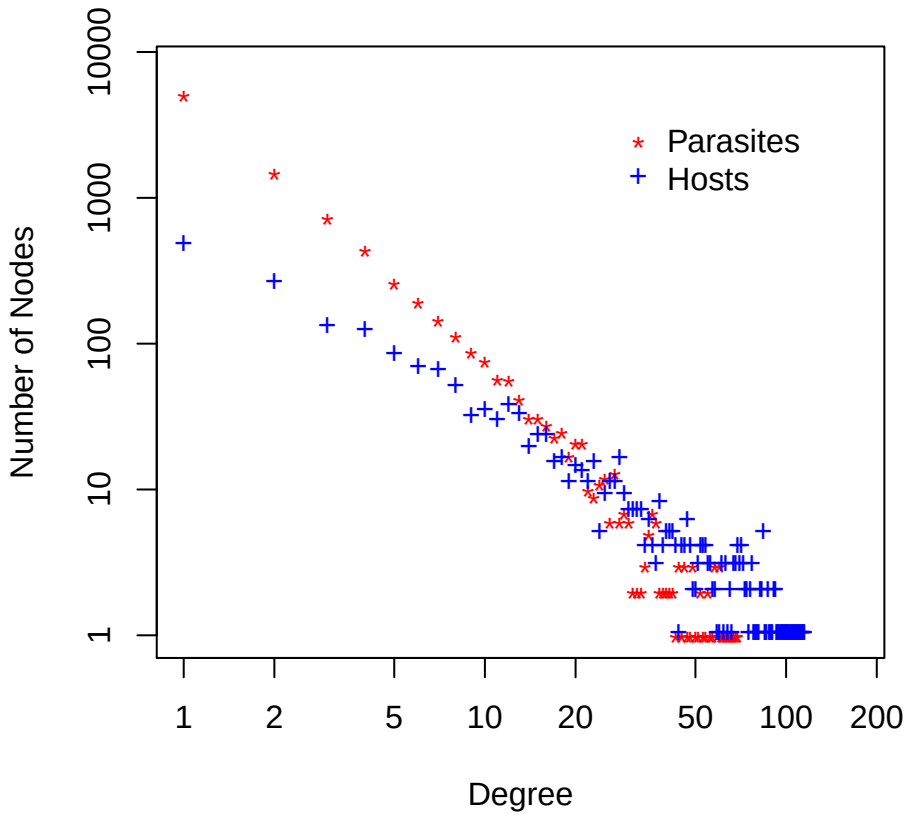

Figure SM 1: Degree distributions of the number of associations (degree) for hosts and parasites (nodes). The distributions are linear on the log scale, indicating a power-law in the number of observed associations per species across the network.

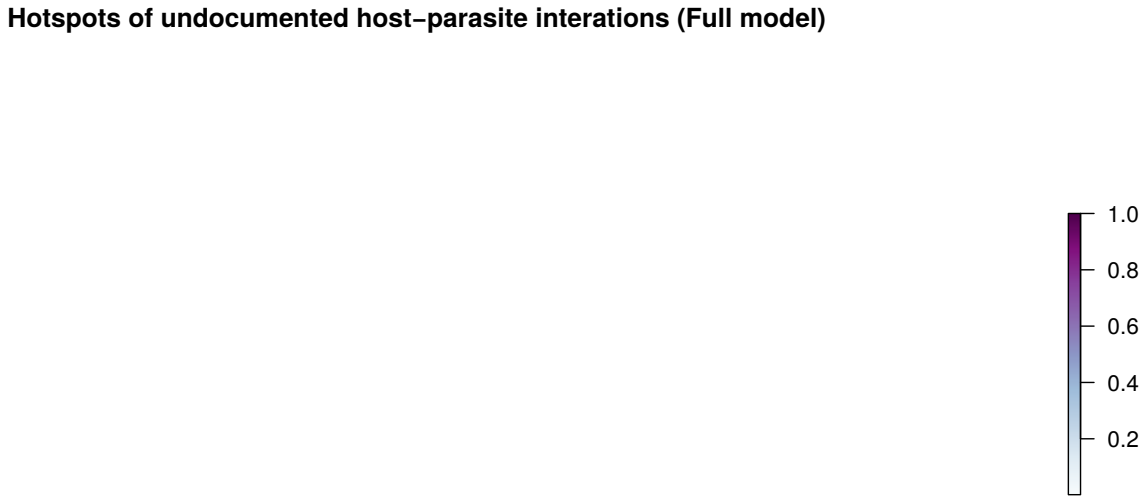

Figure SM 2: Global hotspot map of relative risk of undocumented host-parasite interactions based on predictions from the full model. The map was generated by summing the probabilities of undocumented host-parasite interactions per host species, summing across host species per cell with range maps for terrestrial mammals from the IUCN, then standardized to a scale of 0 - 1.

Model diagnostics

All models showed high predictive accuracy in cross-fold validation: area under the receiver operating characteristic curve (AUC) values ranging from 0.842 - 0.978, where a maximum AUC of 1 signifies perfect predictive accuracy, and with between 72.54% and 98.00% of the held-out documented interactions successfully recovered (Fig. 2A, Table SM 1; see Fig. 4 for posterior interaction matrices for the full dataset). While remaining high overall, AUC tended to decrease with the total size of the interaction matrix, with the phylogeny only model demonstrating the lowest AUC when applied to the largest interaction matrices – the full matrix, and the helminth subset (Fig. 2A). The percent documented interactions (1s) correctly recovered from the held-out portion also decreased the total size of the interaction matrix (Fig. 2C), with the full model outperforming the other models in all subsets except for the full dataset and the virus subset (Table SM 1). For each model, we conducted literature searches for the top ten most likely links without documentation in the current database. This resulted in 177 unique host-parasite links after removing duplicate predictions across models and subsets (see Materials & Methods). Of the undocumented links for which literature searches were conducted, we identified 72 links with evidence of infection (direct observation, genetic sequencing, or positive serology), and an additional 14 links with some evidence, but for which additional confirmatory data are required (e.g. antibodies but no confirmed cases for human infections, known cross-reactivity of the serological test used, an unconfirmed visual diagnosis, or the identification of a genetically similar but previously unknown parasite). Of the remaining links for which we could not find conclusive evidence, we highlight 39 that should be targeted for surveillance. These include links where there is known geographic overlap in the ranges of the host and parasite and host ecologies likely facilitate exposure. We also identify a number of links that are highly likely in the model, but are unlikely due to the mode of disease transmission, non-overlapping host and parasite geographies, or potential competitive interactions with closely related parasites. Overall the full and phylogeny only models tended to identify a greater number of links with published evidence, and fewer ecologically unlikely links (Fig. 2D).

Data subset	Model	AUC	% 1s recovered
Full dataset	Affinity only	0.938	92.60
	Phylogeny only	0.842	72.54
	Full model	0.908	83.66
Arthropods	Affinity only	0.939	87.89
	Phylogeny only	0.946	92.81
	Full model	0.952	93.76
Bacteria	Affinity only	0.974	94.09
	Phylogeny only	0.944	91.89
	Full model	0.963	95.44
Fungi	Affinity only	0.970	95.59
	Phylogeny only	0.963	93.09
	Full model	0.978	98.00
Helminths	Affinity only	0.937	85.00
	Phylogeny only	0.879	81.45
	Full model	0.919	86.70
Protozoa	Affinity only	0.963	93.90
	Phylogeny only	0.966	93.66
	Full model	0.956	94.33
Viruses	Affinity only	0.935	89.81
	Phylogeny only	0.920	89.16
	Full model	0.911	89.20

Table SM 1: Average model performances diagnostics after 10-fold cross validation: area under the receiver operating characteristic curve (AUC), and percent documented interactions (1s) correctly recovered from the held-out portion.

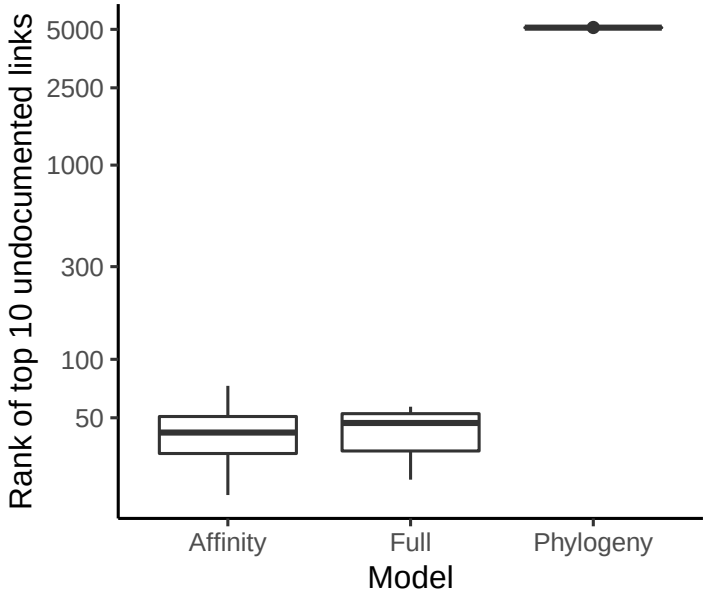

Figure SM 3: Boxplots of the ranks of the top 10 predicted links with no documentation in the original dataset, for each of the three model variants applied to the full dataset.

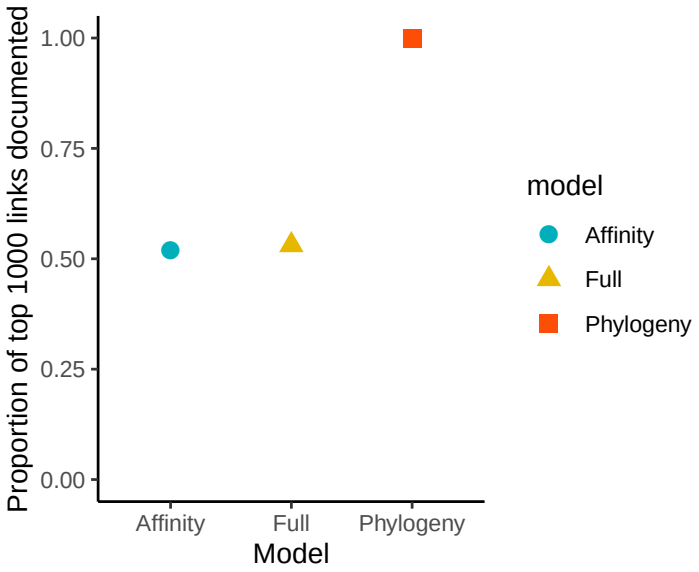

Figure SM 4: The proportion of the top 1000 predicted links which had already been documented in the original dataset. Results are shown for each of the model variants applied to the full dataset.

Phylogeny scaling

To account for uncertainty in the phylogenetic distances among hosts and improve prediction, the model estimates a tree scaling parameter (η) based on an early-burst model of evolution (*1*). Across models, η was estimated to be positive, corresponding to a model of accelerating evolution, and suggesting less phylogenetic conservatism in host link associations among closely related taxa than predicted under a pure-Brownian motion model (*1*). Not surprisingly, η varied when the data was subset by parasite type (Figs. SM 5 & SM 6). Interestingly, arthropods and fungi were estimated to have the smallest η parameters (both ~ 8.15), perhaps reflecting the tendency for fungi to include opportunistic pathogens such as *Pneumocystis carinii* and *Chrysosporium parvum*. In contrast, larger η was estimated for helminths and viruses (10.28 and 9.54 respectively), consistent with the observation of Park et al. (*2*) that mean phylogenetic specificity is similar in these two groups, though viruses are more variable and contain more extreme specialist and generalist parasites. Overall the full dataset was estimated with an η parameter most similar to the helminth subset (10.76), reflecting the representation of helminths in the full dataset (roughly 64% of the observed host-parasite interactions). The discrepancy in phylogeny scaling across the full model and the subsets by parasite type likely contributes to explaining why the performance of the phylogeny only model is lower in the full dataset. Future extensions may benefit from developing a more flexible model to allow for an interaction between phylogenetic scaling and parasite taxonomy, rather than using a single scaling parameter.

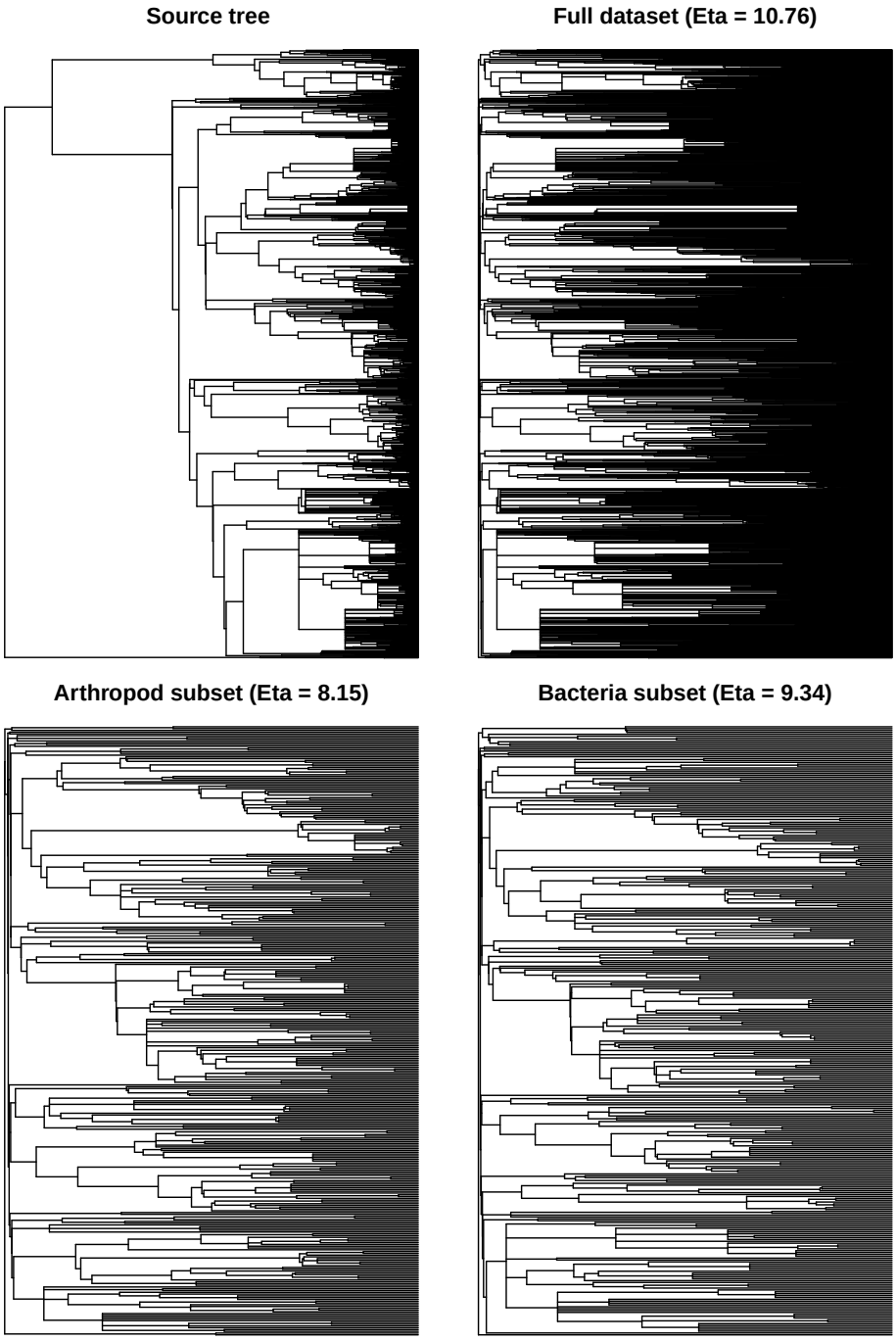

Figure SM 5: The source phylogeny pruned to include only hosts in the amalgamated dataset, and the phylogenies scaled by mean estimated Eta for the phylogeny only models applied to the full dataset, and arthropod and bacteria subsets.

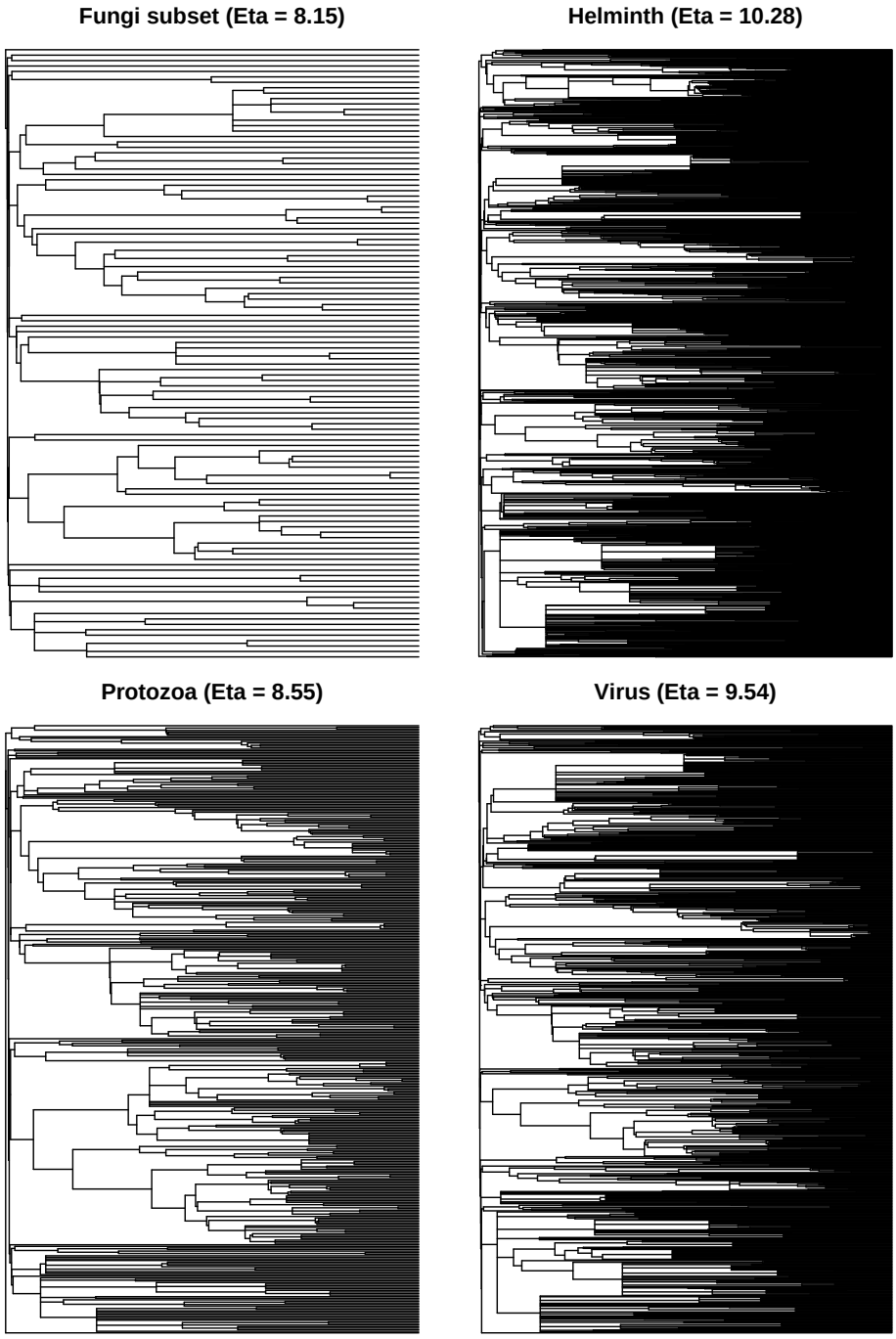

Figure SM 6: Phylogenies scaled by mean estimated Eta for the phylogeny only models applied to the fungi, helminth, protozoa, and virus subsets.

Top predicted links and literature search results

Dataset	Model	Unique Hosts	Unique Parasites	Links with evidence	Unlikely links
Full	Affinity only	1	10	1	1
Full-Domestics	Affinity only	3	9	3	2
Full-Wild	Affinity only	10	1	8	0
Full	Phylogeny only	10	2	4	0
Full-Domestics	Phylogeny only	9	2	7	0
Full-Wild	Phylogeny only	10	2	4	0
Full	Full model	9	3	7	0
Full-Domestics	Full model	5	6	5	2
Full-Wild	Full model	10	1	8	0
Arthropod	Affinity only	6	5	2	0
Arthropod	Phylogeny only	10	3	4	0
Arthropod	Full model	9	2	3	0
Bacteria	Affinity only	1	10	2	0
Bacteria	Phylogeny only	9	3	4	0
Bacteria	Full model	6	7	3	1
Fungi	Affinity only	3	9	2	5
Fungi	Phylogeny only	10	1	1	0
Fungi	Full model	9	3	5	1
Helminths	Affinity only	3	7	2	3
Helminths	Phylogeny only	8	8	2	1
Helminths	Full model	5	6	0	5
Protozoa	Affinity only	7	4	4	3
Protozoa	Phylogeny only	10	2	7	0
Protozoa	Full model	8	3	6	2
Viruses	Affinity only	4	8	4	0
Viruses	Phylogeny only	10	1	6	0
Viruses	Full model	10	2	8	0

Table SM 2: Numbers of unique hosts, unique parasites, links for which evidence was identified in the literature, and unlikely links based ecological mistmatch for each of the top 10 predicted links per model.

Affinity only – Full Dataset

Host	Parasite	Status
Homo sapiens	Taenia mustelae	
Homo sapiens	Bluetongue virus	
Homo sapiens	Bovine viral diarrhea virus 1	
Homo sapiens	Neospora caninum	
Homo sapiens	Mastophorus muris	
Homo sapiens	Plagiorchis vespertilionis	Confirmed
Homo sapiens	Carnivore protoparvovirus 1	
Homo sapiens	Alaria alata	
Homo sapiens	Taenia pisiformis	
Homo sapiens	Physocephalus sexalatus	Unlikely

Table SM 3: Top 10 undocumented links with highest probablity of interaction: Affinity only model - full data set.

- A recent molecular phylogeny of the *Taenia* genus supported the creation of a new genus and renaming of *Taenia mustelae* to *Versteria mustelae* (*3*). Since then there has been a report of fatal infection of a previously unknown *Versteria* species in a captive orangutan *Pongo pygmaeus* cloesly related to species found in wild mustelids, suggesting the need for increased vigilance of *Versteria* infections in humans (*4*).

- While bluetongue virus is known to infect a wide range of ruminants, it is not currently considered to infect humans (*5*).

- Bovine viral diarrhea viruses are not considered to be human pathogens, but there is some concern about zoonotic potential as they are highly mutable, have the ability to replicate in human cell lines, and have been isolated from humans on rare occasions (*6*).

- Although antibodies to *Neospora caninum* have been reported in humans, the parasite has not been identified in human tissues and the zoonotic potential is not known (*7*).

- *Mastophorus muris* is a rodent-specific nematode that requires arthropods as intermediate hosts and while this makes it unlikely to infect humans, it was recently documented in an urban population of rats in the UK (*8*), indicating the potential for human exposure.

- While natural infections of *Plagiorchis* species in humans are rare, the first case of human infection by the bat parasite *Plagiorchis vespertilionis* was reported in 2007 (*9*). The source of infection is uncertain, it has been suggested that freshwater fish and snails may be undocumented intermediate hosts and infection was due to ingestion of raw freshwater fish.

- Carnivore protoparvovirus can infect a number of hosts in the order *Carnivora* (*10*), though there seems to be no evidence of human infection.

- *Alaria alata*, an intestinal parasite of wild canids, has not been identified in humans, but is considered a potential zoonotic risk as other *Alaria* species have been reported to cause fatal illness in humans (*11*).

- Due to the characteristics of the biological cycle of *Tenia pisiformis* and the observation that it is innocuous in humans, this parasite has been used as a model for the study of other important zoonosis relevant to human health including *T. solium* (*12*).

- The definitive hosts of *Physocephalus sexalatus* are commonly wild and domestic pigs, but it is sometimes found in other mammals and some reptiles (*13*). However the parasite uses beetles as an intermediate host, which makes human infection unlikely.

Affinity only – Full Dataset – Domestic Hosts

Host	Parasite	Status
Bos taurus	Trypanosoma cruzi	
Rattus norvegicus	Rabies lyssavirus	Confirmed
Bos taurus	Hymenolepis diminuta	
Bos taurus	Mesocestoides lineatus	Unlikely
Bos taurus	Capillaria hepatica	Confirmed
Bos taurus	St. Louis encephalitis virus	
Bos taurus	Canine distemper virus	
Bos taurus	Anisakis simplex	Unlikely
Ovis aries	Trypanosoma cruzi	Confirmed
Bos taurus	Taenia mustelae	

Table SM 4: Top 10 undocumented links with highest probablity of interaction for domesticated species: Affinity only model - full data set.

- Currently the role of cattle in the epidemiology of Chagas disease (caused by *Trypanosoma cruzi*) is unknown, though the majority of cattle in Latin America may be exposed (280 million heads; 1/4 of the world population) (*14*). (*15*) report that 177 species have been documented as susceptible to infection by *T. cruzi*, with domestic hosts in some cases being responsible for the maintenance of local parasite populations over long periods of time. While cattle have tested positive in serological studies, cows and other domestic species are also infected by *Trypanosoma* and *Phytomonas* species which can cause cross-reactions in diagnostic tests (*16*).

- Rabies has an extremely large host range and surprisingly rats are rarely reported as suffering from rabies, though there have been a few reported cases rabid *Rattus norvegicus* in the United States (*17*).

- The natural hosts of *Hymenolepis diminuta* are rats and cattle have not been found to be susceptible to infection, however *H. diminuta* eggs have been found in the feces of dairy cattle, likely the result of ingesting forage contaminated with rodent feces (*18*).

- *Mesocestoides lineatus* has a three-stage lifecycle with two intermediate hosts and a large range of carnivorous mammals as definitive hosts (*19*). Human infections of the tapeworm *Mesocestoides lineatus* are rare but can occur through the consumption of chickens, snails, snakes, or frogs (*20*) and therefore it is unlikely that cows will ingest the intermediate life stages of this parasite.

- *Capillaria hepatica* (syn. *Calodium hepaticum*), is a globally distributed zoonotic parasite which uses rodents as main hosts, but is known to cause infection in over 180 mammalian species, including cattle (*21*).

- Birds are the primary vertebrate hosts for St. Louis encephalitis virus, though amplification by certain mammals has been suggested (*22*). There is some serological evidence of infection in

domestic mammals, including cattle (*23*). The common vector *Culex nigripalpus* feeds primarily on birds, but shows a seasonal shift from avian hosts in the spring to mammalian hosts in the summer, indicating it may be able to act as a bridging vector among different host species (*22*).

- Canine distemper virus infects a wide range of hosts within the order Carnivora, but has also been found to cause fatal infection in some non-human primates and peccaries (*24*).
- *Anisakis simplex* uses cetaceans as final hosts, with marine invertebrates and fish as intermediate hosts (*25*). Whales are infected through ingestion, indicating that while cattle may be susceptible, though they would need sufficient exposure to marine based feed.
- *Ovis aries* is reported to be infected by *Trypanosoma cruzi* (*15*).
- We found no evidence of *Bos taurus* infection by *Taenia mustelae*.

Affinity only – Full Dataset – Wild Hosts

Host	Parasite	
Cervus elaphus	Rabies lyssavirus	Confirmed
Ondatra zibethicus	Rabies lyssavirus	Confirmed
Capreolus capreolus	Rabies lyssavirus	Confirmed
Sorex araneus	Rabies lyssavirus	
Apodemus sylvaticus	Rabies lyssavirus	Confirmed
Odocoileus virginianus	Rabies lyssavirus	Confirmed
Syncerus caffer	Rabies lyssavirus	
Dama dama	Rabies lyssavirus	Confirmed
Pan troglodytes	Rabies lyssavirus	Confirmed
Microtus arvalis	Rabies lyssavirus	Confirmed

Table SM 5: Top 10 undocumented links with highest probablity of interaction for wild host species: Affinity only model - full data set.

- For the affinity only model run on the full dataset, all of the top ten interactions involve rabies, which is unsurprising as rabies is commonly said to be capable of infecting all mammal species (*26*) and is the parasite with the largest number of documented hosts in the database (176). Of these ten most likely hosts, we found evidence of rabies infection in eight (*Cervus elaphus* and *Odocoileus virginianus* (*27*), *Ondatra zibethicus* (*28*), *Capreolus capreolus* (*29*), *Pan troglodytes* (*30*), *Apodemus sylvaticus* (*31*), *Dama dama* (*32*), and successful experimental infection of *Microtus arvalis* (*33*)).

Phylogeny only – Full Dataset

Host	Parasite	Status
Vulpes rueppellii	Rabies lyssavirus	
Vulpes macrotis	Rabies lyssavirus	Confirmed
Vulpes ferrilata	Rabies lyssavirus	
Lycalopex culpaeus	Rabies lyssavirus	
Lycalopex fulvipes	Rabies lyssavirus	
Lycalopex griseus	Rabies lyssavirus	Confirmed
Lycalopex gymnocercus	Rabies lyssavirus	
Cuon alpinus	Rabies lyssavirus	Confirmed
Speothos venaticus	Rabies lyssavirus	Confirmed
Holochilus chacarius	Schistosoma mansoni	

Table SM 6: Top 10 undocumented links with highest probablity of interaction (Phylogeny only model - full data set)

- We did not find any record of rabies infecting *Vulpes rueppellii*, but this should be investigated as this disease is known to cause severe declines in wild canids (*34*).
- Rabies has been documented to infect the endangered San Joaquin kit fox (*Vulpes macrotis mutica*) and is suggested to have caused a catastrophic decline of the species in the 1990s (*35*).
- Domestic dogs are considered a major predation threat to the Tibetan fox (*Vulpes ferrilata*) (*36*), and rabies is confirmed to circulate in wild and domestic animals in Tibet (*37*).
- Domestic dogs alter the ecology of Andean foxes (*Lycalopex culpaeus*), have been observed hunting them, and are a potential source of infection (*38*).
- Diseases from domestic dogs (largely canine distemper) is considered a major threat to the endangered *Lycalopex fulvipes* (*39*), indicating that rabies may also pose a risk.
- There is one report of serological evidence of rabies in *Lycalopex griseus* in Chile in 1989 (*40*) (reported as *Pseudalopex griseus*, a formerly accepted name (*41*)).
- We did not find evidence of rabies infection in *Lycalopex gymnocercus*, however its distribution overlaps with species known to be important in the transmission of rabies in Brazil (*42*).
- The endangered Dhole (*Cuon alpinus*) is known to suffer from rabies and was a source of fatal human infections during an outbreak in the 1940s (*43*).
- Rabies has been reported as potentially infecting *Speothos venaticus* (*44*) and there is a report of an individual with positive serology (*45*).
- We could not find evidence of *Schistosoma mansoni* infection in *Holochilus chacarius*. Congener *Holochilus braziliensis* was experimentally shown to be a viable host, although infection resulted in host death (*46*).

Phylogeny only – Full Dataset – Domestic Hosts

Host	Parasite	Status
Bison bison	Rabies lyssavirus	Confirmed
Bos grunniens	Rabies lyssavirus	Confirmed
Bos frontalis	Rabies lyssavirus	
Bos javanicus	Rabies lyssavirus	
Vicugna vicugna	Rabies lyssavirus	
Rattus rattus	Rabies lyssavirus	Confirmed
Rattus norvegicus	Rabies lyssavirus	Confirmed
Cavia porcellus	Rabies lyssavirus	Confirmed
Oryctolagus cuniculus	Rabies lyssavirus	Confirmed
Bos grunniens	Toxoplasma gondii	Confirmed

Table SM 7: Top 10 undocumented links with highest probablity of interaction for domesticated species: Phylogeny only model - full data set.

- Rabies in *Bison bison* is considered rare, but there are multiple cases reported (*47*).
- Rabies has been reported to infect yak (*Bos grunniens*) in Nepal (*48*).
- We could not find evidence of rabies infection in *Bos frontalis*, but as this is a semi-wild and endangered species (*49*) and other *Bos* species are susceptible, the disease may pose a conservation risk. This may also be the case for the endangered *Bos javanicus*.
- We did not find a specific report of rabies infection in *Vicugna vicugna*, although all South American camelids are noted to be susceptible and display clinical signs of infection (*50*).
- There are documented cases of rabies infecting *Rattus rattus* (ex. (*51*)), though it appears to be rare.
- Rabies in *Rattus norvegicus* was predicted by the affinity only model with the full dataset for domestic hosts.
- Pet guinea pigs *Cavia porcellus* have been infected with rabies after being bitten by a raccoon (*52*).
- There are reported cases of rabid *Oryctolagus cuniculus* in the United States (*17*).
- *Toxoplasma gondii* is known to infect *Bos grunniens* and cause severe economic losses (*53*).

Phylogeny only – Full Dataset – Wild Hosts

Host	Parasite	Status
Vulpes rueppellii	Rabies lyssavirus	
Vulpes macrotis	Rabies lyssavirus	Confirmed
Vulpes ferrilata	Rabies lyssavirus	
Lycalopex culpaeus	Rabies lyssavirus	
Lycalopex fulvipes	Rabies lyssavirus	
Lycalopex griseus	Rabies lyssavirus	Confirmed
Lycalopex gymnocercus	Rabies lyssavirus	
Cuon alpinus	Rabies lyssavirus	Confirmed
Speothos venaticus	Rabies lyssavirus	Confirmed
Holochilus chacarius	Schistosoma mansoni	

Table SM 8: Top 10 undocumented links with highest probablity of interaction for wild host species: Phylogeny only model - full data set.

- These top predictions are the same as for the phylogeny only model on the full dataset (all wild host species in the top 10 links).

Full model – Full Dataset

Host	Parasite	Status
Rattus norvegicus	Rabies lyssavirus	Confirmed
Cervus elaphus	Rabies lyssavirus	Confirmed
Homo sapiens	Taenia mustelae	
Homo sapiens	Bluetongue virus	
Ondatra zibethicus	Rabies lyssavirus	Confirmed
Capreolus capreolus	Rabies lyssavirus	Confirmed
Rattus rattus	Rabies lyssavirus	Confirmed
Sorex araneus	Rabies lyssavirus	
Oryctolagus cuniculus	Rabies lyssavirus	Confirmed
Apodemus sylvaticus	Rabies lyssavirus	Confirmed

Table SM 9: Top 10 undocumented links with highest probablity of interaction (Full model - full data set)

- All of the top links were previously discussed except for Rabies infection of *Sorex araneus*, for which we could not find any published evidence.

Full model – Full Dataset – Domestic Hosts

Host	Parasite	Status
Rattus norvegicus	Rabies lyssavirus	Confirmed
Rattus rattus	Rabies lyssavirus	Confirmed
Oryctolagus cuniculus	Rabies lyssavirus	Confirmed
Bos taurus	Trypanosoma cruzi	
Bos taurus	Hymenolepis diminuta	
Ovis aries	Trypanosoma cruzi	Confirmed
Bos taurus	Mesocestoides lineatus	Unlikely
Bos taurus	Anisakis simplex	Unlikely
Bos taurus	Capillaria hepatica	Confirmed
Ovis aries	Hymenolepis diminuta	

Table SM 10: Top 10 undocumented links with highest probablity of interaction for domesticated species: Full model - full data set.

- All of the top links were previously discussed except for *Hymenolepis diminuta* infection of sheep, for which we could not find any published evidence.

Full model – Full Dataset – Wild Hosts

Host	Parasite	Status
Cervus elaphus	Rabies lyssavirus	Confirmed
Ondatra zibethicus	Rabies lyssavirus	Confirmed
Capreolus capreolus	Rabies lyssavirus	Confirmed
Sorex araneus	Rabies lyssavirus	
Apodemus sylvaticus	Rabies lyssavirus	Confirmed
Odocoileus virginianus	Rabies lyssavirus	Confirmed
Syncerus caffer	Rabies lyssavirus	
Dama dama	Rabies lyssavirus	Confirmed
Aepyceros melampus	Rabies lyssavirus	Confirmed
Rupicapra rupicapra	Rabies lyssavirus	Confirmed

Table SM 11: Top 10 undocumented links with highest probablity of interaction for wild host species: Full model - full data set.

- Most of these links were predicted by models discussed above except for rabies infection in *Aepyceros melampus*, which has been identified as suffering from spillover infections (*54*), and *Rupicapra rupicapra*, which has been documented in Europe (*31*).

Arthropods – Affinity only

Host	Parasite	Status
Vulpes vulpes	Rhipicephalus evertsi	
Vulpes vulpes	Rhipicephalus appendiculatus	
Vulpes vulpes	Amblyomma hebraeum	
Cervus elaphus	Rhipicephalus evertsi	
Aepyceros melampus	Sarcoptes scabiei	Confirmed
Vulpes vulpes	Hyalomma truncatum	
Cervus elaphus	Rhipicephalus appendiculatus	
Sus scrofa	Rhipicephalus evertsi	
Bos taurus	Sarcoptes scabiei	Confirmed
Odocoileus virginianus	Sarcoptes scabiei	

Table SM 12: Top 10 undocumented links with highest probablity of interaction: Affinity only - arthropod subset.

- *Rhipicephalus evertsi* and *R. appendiculatus* are common ticks in East and Southern Africa (*55*) and are unlikely to interact with non-African hosts such as *Cervus elaphus* and *Vulpes vulpes* (though there are some populations of *Vulpes vulpes* in North Africa). Future iterations of our implemented link prediciton framework may benefit from the inclusion of information on geographic range overlap among host species, however this may reduce the ability to identify future host-parasite associations that may occur given range expansions or species translocations.

- *Amblyomma hebraeum*, the main vector of *Ehrlichia ruminantium* in southern Africa, prefer large hosts such as cattle and wild ruminants, although immature stages are found to feed on a wide range of hosts including scrub hares, guineafowl, and tortoises (*56*). Considering the species is restricted to southern Africa, it is unlikely to interact with Vulpes vulpes, although it's wide host range during immaturity indicates that it may be suscpetible given the opportunity.

- (*57*) compiled a list of documented host species of sarcoptic mange (caused by *Sarcoptes scabiei*) which includes *Aepyceros melampus* and cattle (*Bos taurus*).

- *Hyalomma truncatum* is found across sub-Saharan Africa, where it commonly infests domestic and wild herbivores, and domestic dogs (*56*). Considering the species is restricted to Africa, it is unlikely to interact with Vulpes vulpes, although it's tendency to infest domestic dogs indicates that it may be suscpetible given the opportunity.

- *Rhipicephalus appendiculatus* was found to be the most prevalent tick species on domestic pigs in the Busia District of Kenya (*58*), indicating that increased monitoring may also identify *R. evertsi* on domestic pigs.

- While multiple cervids have been reported with sarcoptic mange (*57*), we cannot find any report of infection in white-tailed deer *Odocoileus virginianus*, although there are numerous reports of

infection with mange caused by *Demodex sp.*, including the host specific *Demodex odocoilei* (*59*), potentially indicating competition among *Sarcoptes* and *Demodex* species.

Arthropods – Phylogeny only

Host	Parasite	
Canis adustus	Sarcoptes scabiei	
Lycalopex fulvipes	Sarcoptes scabiei	
Lycalopex vetulus	Sarcoptes scabiei	
Vulpes lagopus	Pulex irritans	
Cerdocyon thous	Sarcoptes scabiei	Confirmed
Chrysocyon brachyurus	Sarcoptes scabiei	
Cuon alpinus	Sarcoptes scabiei	Confirmed
Speothos venaticus	Sarcoptes scabiei	Confirmed
Vulpes macrotis	Dermacentor variabilis	Confirmed
Ovis ammon	Sarcoptes scabiei	

Table SM 13: Top 10 undocumented links with highest probablity of interaction: Phylogeny only model - Arthropod subset.

- There does not appear to be a published record of sarcoptic mange in *Canis adustus*, however in areas with sympatric jackal species *C. adustus* usually display ecological segregation through preferring denser vegetation (*60*). This may indicate that while *C. adustus* may be susceptible to sarcoptic mange, differences in the ecologies of this species relative to other canids may limit transmission making overt infections difficult to document.

- *Lycalopex fulvipes* is endangered (*39*), meaning that its small population sizes and restricted geographic range may reduce exposure to *S. scabiei*, however as sarcoptic mange is implicated in the declines of other wild canids, it should be targeted in disease monitoring programs for this species.

- While *Lycalopex vetulus* is not endangered, it displays some adaptability to anthropogenic disturbance (*61*), which may expose it to sarcoptic mange through contact with domestic dogs. In addition, *Lycalopex vetulus* is sympatric with the crab eating fox (*Cerdocyon thous*) – a documented host of *S. scabiei* (*57*). The IUCN reports a gap in conservation actions for *L. vetulus* regarding the role of disease in population regulation, and their status as reservoirs of scabies, canine distemper, leishmaniasis, and rabies (*61*).

- While the Arctic fox (*Vulpes lagopus*) is considered the most important terrestrial game species in the Arctic (*62*), we were unable to find documented infection by the "human flea" (*Pulex irritans*). *P. irritans* is thought to be unable to persist in Arctic envrionments due to the temperature thresholds necessary for breeding (*63*), though this may change in the future with continued Arctic warming.

- (*57*) compiled a list of documented host species of sarcoptic mange (caused by *Sarcoptes scabiei*) which includes *Cerdocyon thous*.

- *Chrysocyon branchyurus* has one report of clinical signs suggestive of sarcoptic mange-like infestation (*64*).
- The endangered Dhole (*Cuon alpinus*) has been documented as suffering from mange as early as 1937 (*43*) and appear to be especially susceptible to disease outbreaks due to their large group sizes and amicable behaviour within packs.
- *S. scabei* was identified in *Speothos venaticus* (*65*), and identified as potentially contributing to the loss of individuals from a group in Mato Grosso, Brazil (*66*).
- *Vulpes macrotis* was documented as infested with *Dermacentor variabilis* in Southwest Texas in 1950 (*67*).
- We cannot find a record of mange in *Ovis ammon*, though outbreaks of sarcoptic mange have been documented in ibex and blue sheep in the Taxkorgan Reserve, China, in which *O. ammon* are also present, although this population has received little study (*68*).

Arthropods – Full model

Host	Parasite	Status
Cerdocyon thous	Sarcoptes scabiei	Confirmed
Cervus elaphus	Rhipicephalus evertsi	
Vulpes velox	Sarcoptes scabiei	
Aepyceros melampus	Sarcoptes scabiei	Confirmed
Vulpes macrotis	Sarcoptes scabiei	Confirmed
Chrysocyon brachyurus	Sarcoptes scabiei	
Odocoileus virginianus	Rhipicephalus evertsi	
Odocoileus virginianus	Sarcoptes scabiei	
Lycalopex vetulus	Sarcoptes scabiei	
Bos taurus	Sarcoptes scabiei	Confirmed

Table SM 14: Top 10 undocumented links with highest probablity of interaction: Full model - arthropod subset.

- We were unable to find any published evidence of *Sarcoptes scabiei* infesting *Vulpes velox*. However, sarcoptic mange is known to infest several species of canids, is prevalent in coyotes (*Canis latrans*) within the range of *Vulpes velox* (*69*). Criffield2009 surveyed for *S. scabiei* on *V. velox*, but no clinical signes were observed, and suggest that as the grey fox (*Urocyon cineroargenteus*) is somewhat resistant to mange in laboratory tests, *V. velox* may be similarly resistant, though there have been no experimental infestations. Finally, *V. velox* is debated to be conspecific with *Vulpes macrotis* (*70*), which has been documented with sarcoptic mange (*71*).

- *Sarcoptes scabiei* has been documented in the endangered San Jaoaquin kit fox *Vulpes macrotis mutica*, and is considered a significant threat to it's conservation (*71*).

- *Rhipicephalus evertsi* is a common tick in East and Southern Africa (*55*) and is unlikely to interact with non-African hosts such as *Odocoileus virginianus*. Future iterations of our link prediction method may benefit from the inclusion of information on geographic range overlap among host species, however this may reduce the ability to identify future host-parasite associations that may occur given range expansions or species translocations.

- The remaining links are discussed above.

Bacteria – Affinity only

Host	Parasite	Status
Homo sapiens	Bartonella grahamii	Confirmed
Homo sapiens	Anaplasma marginale	
Homo sapiens	Anaplasma bovis	
Homo sapiens	Mycoplasma mycoides	
Homo sapiens	Mycoplasma haemofelis	
Homo sapiens	Chlamydophila pecorum	
Homo sapiens	Lawsonia intracellularis	
Homo sapiens	Mycoplasma conjunctivae	
Homo sapiens	Chlamydia psittaci	Confirmed
Homo sapiens	Histophilus somni	

Table SM 15: Top 10 undocumented links with highest probablity of interaction: Affinity only model - bacteria subset.

- *Bartonella grahamii* is a pathogen of rodents worldwide, but was first identified as causing an infection in an immunocompromised human in 2013 (*72*).
- *Anaplasma bovis*, causal agent of bovine anaplasmosis, is not currently considered zoonotic (*73*), but *Anaplasma phagocytophilum* the causative agent of human anaplasmosis is placed is as sister taxa to *A. bovis* in a recent phylogeny (*74*). Similarly, *A. marginale*, the causative agent of anaplasmosis in cattle, is also not considered zoonotic, but it reaches high prevalence in cattle and humans are likely exposed to the tick vector (*73*).
- *Mycoplasma mycoides* is not typically thought to infect humans, but there is one report of disease and positive serology in a farm worker exposed to multiple calves infected with *M. mycoides subsp. mycoides LC* (*75*).
- There has been one documented case of infection in an immunocompromised human with a *Mycoplasma haemofelis*-like bacteria (*76*) and the authors note that disease-causing latent mycoplasma infections in immunocompromised and non-immunocompromised patients are an emerging issue.
- The zoonotic potential of *Chlamydophila pecorum* is not known, although it is associated with abortions in small ruminants (*77*).
- *Lawsonia intracellularis* was recently recognised as the cause of an emerging intestinal disease in horses (Equine proliferative enteropathy), but is currently not considered to be zoonotic (*78*).
- *Mycoplasma conjunctivae* causes a highly contagious ocular infection of sheep, goats, and wild Caprinae, and is possibly zoonotic as it has been associated with eye inflammation in young children (*79*).

- *Chlamydophila psittaci* (formerly known as *Chlamydia psittaci*) is a known zoonotic disease from birds (*77*).
- *Histophilus somni* is a pathogen of bovine and ovine hosts, but has not been documented to infect humans (*80*).

Bacteria – Phylogeny only model

Host	Parasite	
Canis aureus	Leptospira interrogans	
Canis mesomelas	Leptospira interrogans	
Canis mesomelas	Anaplasma phagocytophilum	
Saguinus geoffroyi	Escherichia coli	Confirmed
Equus burchellii	Escherichia coli	Confirmed
Equus zebra	Escherichia coli	Confirmed
Lycaon pictus	Leptospira interrogans	
Vulpes lagopus	Leptospira interrogans	
Vulpes velox	Leptospira interrogans	
Canis latrans	Escherichia coli	Confirmed

Table SM 16: Top 10 undocumented links with highest probablity of interaction: Phylogeny only - bacteria subset.

- *Leptospira sp.* are commonly regarded as infecting a wide range of mammals (*81*). (*82*) identified *Leptospira interrogans* serovar *canicola* in the urine of jackals in Israel though did not identify the particular species. However, *Canis aureus* is the only jackal species present in the country (*83*) providing some support for this host-parasite association, though the findings of (*82*) should be verified.

- We did not find any documentation of leptospirosis in *Canis mesomelas* or *Lycaon pictus* but considering it is found in multiple species in Africa including domestic dogs (*84*), wild canids are likely to be exposed.

- (*85*) sequenced DNA from blood samples of *Canis mesomelas* in South Africa and identified 16S rDNA sequences very similar to *Anaplasma phagocytophilum*. This study also identified other *Anaplasma* species indicating the potential for 16S rDNA sequencing to gather evidence of predicted host-parasite and discover previously unknown pathogens.

- *E. coli* is a ubiquitous commensal microbe of vertebrates (*86*) and pathogenicitiy is linked to particular strains, indicating that our approach may be expanded by identifying the host ranges of particular subspecies or virulent strains of common commensal bacteria. *Canis latrans* has been identified as harbouring atypical enteropathogenic *E. coli* and may serve as a reservoir in agricultural areas near the United States-Mexico border (*87*). Wild Equus burchellii have been found to harbour antibiotic resistant *E. coli* in South Africa (*88*) and Tanzania (*89*), and captive *Equus zebra hartmannae* have been found with antibiotic resistant *E. coli* (*90*). Similarly, antibiotic resistant *E. coli* have been found in captive *Saguinus geoffroyi* (*91*).

- We did not find any evidence of *Leptospira* infections in *Vulpes lagopus*. We did find one report of positive serology for *Leptospira interrogans* in *Vulpes velox macrotis* (*92*), though there is debate as to whether this subspecies is actually its own species *Vulpes macrotis* (*41*).

Bacteria – Full model

Host	Parasite	Status
Ovis canadensis	Escherichia coli	
Bison bison	Escherichia coli	Confirmed
Ovis aries	Mycobacterium bovis	Confirmed
Homo sapiens	Bartonella grahamii	Confirmed
Ovis canadensis	Leptospira interrogans	
Zalophus californianus	Anaplasma phagocytophilum	Unlikely
Ovis canadensis	Anaplasma phagocytophilum	
Phoca vitulina	Leptospira interrogans	
Homo sapiens	Anaplasma marginale	
Bison bison	Anaplasma phagocytophilum	

Table SM 17: Top 10 undocumented links with highest probablity of interaction: Full model - bacteria subset.

- *E. coli* is ubiquitous commensal microbe of vertebrates (*86*) and *Ovis canadensis* has been surveyed for pathogenic *E. coli* in Washington, USA, though no individuals tested positive (*93*). *Bison bison* have been highlighted as a potentially important reservoir of pathogenic *E. coli O157:H7* for human infection (*94*).
- While there is some debate whether sheep are relatively immune or highly susceptible to infection by *Mycobacterium bovis*, spillover infections have been documented to occur when animals are exposed to contaminated pasture (*95*).
- *Ovis canadensis* has been identified as exposed to *Leptospira interrogans* through serological surveys (*96*).
- While tick-borne *Anaplasma phagocytophilum* is found to persist in a large range of terrestrial mammalian hosts (*97*), we find no evidence of infection in *Zalophus californianus* or any other marine mammals.
- We identified one survey of *Anaplasma sp.* in Ovis canadensis in Montana which conducted testing for *Anaplasma phagocytophilum*, though no individuals were positive (*98*). (*98*) speculate that this lack of infection despite the potential for exposure may be due to the exclusion of Anaplasma genotypes such as A. ovis, which is found to infect *Ovis canadensis*.
- Regular exposure of *Phoca vitulina* to *Leptospira interrogans* has been reported (*99*), though low antibody titers were interpreted as exposure rather than infection.
- We did not find evidence of *Anaplasma phagocytophilum* infection in American bison (*Bison bison*), though *A. phagocytophilum* is known to infect European bison (*Bison bonasus*) in Poland (*97*).
- The remaining links are discussed above.

Fungi – Affinity only model

Host	Parasite	Status
Homo sapiens	Geomyces destructans	Unlikely
Homo sapiens	Chrysosporium parvum	Confirmed
Homo sapiens	Neocallimastix frontalis	Unlikely
Homo sapiens	Pilobolus kleinii	Unlikely
Bos taurus	Pneumocystis carinii	Confirmed
Homo sapiens	Trichophyton terrestre	
Phascolarctos cinereus	Pneumocystis carinii	
Homo sapiens	Pilobolus crystallinus	Unlikely
Homo sapiens	Pilobolus heterosporus	Unlikely
Homo sapiens	Chaetomidium arxii	

Table SM 18: Top 10 undocumented links with highest probablity of interaction: Affinity only - fungi subset.

- *Geomyces destructans* is the cause of white nose syndrome in multiple bat species (*100*), but is not documented to be zoonotic.
- *Chrysosporium parvum* and related species are soil fungi that cause pulmonary infections in rodents, fossorial mammals, their predators, and occasionally humans, though the taxonomy of these pathogens is muddied in the literature (*101*).
- *Neocallimastix frontalis* appears to be a commensal fungi of bovid rumens (*102*) and we cannot find documentation of zoonotic infection.
- *Pilobolus sp.* play a role in the decomposition of herbivore dung and although they are non-pathogenic to herbivores, they can facilitate the spread of attached parasitic lungworms because of their projectile dispersal system (*103*). We could find no evidence of human infections.
- *Pneumocystis carinii* belongs to a genus that normally reside in the pulmonary parenchyma of a wide range of mammals (*104*). It is capable of causing life threatening pneumonia in immunocompromised hosts and is documented as causing infections in cattle (*105*), though we found no evidence of infection in *Phascolarctos cinereus*.
- *Trichophpyton terrestre* is part of a large species complex with some variants documented to cause human infection (*106*).
- We cannot find any documentation of *Chaetomidium arxii* infection in humans, although this genus is well known for its opportunistic animal and human pathogens (*107*).

Fungi – Phylogeny only

Host	Parasite	Status
Macaca sylvanus	Pneumocystis carinii	
Felis silvestris	Pneumocystis carinii	
Mustela lutreola	Pneumocystis carinii	
Ateles paniscus	Pneumocystis carinii	
Pan troglodytes	Pneumocystis carinii	
Gorilla beringei	Pneumocystis carinii	
Mustela erminea	Pneumocystis carinii	
Ovis canadensis	Pneumocystis carinii	
Nyctereutes procyonoides	Pneumocystis carinii	
Vulpes vulpes	Pneumocystis carinii	Confirmed

Table SM 19: Top 10 undocumented links with highest probablity of interaction: Phylogeny only - fungi subset.

- *Pneumocystis carinii* belongs to a genus that normally reside in the pulmonary parenchyma of a wide range of mammals (*104*). It is capable of causing life threatening pneumonia in immunocompromised hosts is documented as causing infections cattle and pigs (*108*), *Vulpes vulpes* (*109*), and species in the genera *Macaca* and *Mustela* (*109*).

Fungi – Full model

Host	Parasite	Subset
Bos taurus	Pneumocystis carinii	Confirmed
Phascolarctos cinereus	Pneumocystis carinii	
Sus scrofa	Pneumocystis carinii	Confirmed
Homo sapiens	Geomyces destructans	Unlikely
Capra hircus	Pneumocystis carinii	Confirmed
Oryctolagus cuniculus	Pneumocystis carinii	Confirmed
Lagenorhynchus obliquidens	Pneumocystis carinii	
Homo sapiens	Chrysosporium parvum	Confirmed
Macaca sylvanus	Pneumocystis carinii	
Cervus elaphus	Pneumocystis carinii	

Table SM 20: Top 10 undocumented links with highest probablity of interaction: Full model - fungi subset.

- *Pneumocystis carinii* has been documented to infect pigs (*108, 104*), goats (*104*), and *Oryctolagus cuniculus* (*109*). While it has been documented in other cervids (*109*), we find no evidence of infection in *Cervus elaphus*.
- The remaning top links are discussed above.

Helminths – Affinity only model

Host	Parasite	Status
Homo sapiens	Taenia mustelae	
Bos taurus	Hymenolepis diminuta	
Ovis aries	Hymenolepis diminuta	
Homo sapiens	Mastophorus muris	
Ovis aries	Echinococcus multilocularis	Unlikely
Bos taurus	Mesocestoides lineatus	Unlikely
Ovis aries	Mesocestoides lineatus	Unlikely
Homo sapiens	Plagiorchis vespertilionis	Confirmed
Bos taurus	Capillaria hepatica	Confirmed
Ovis aries	Capillaria hepatica	

Table SM 21: Top 10 undocumented links with highest probablity of interaction: Affinity only - helminth subset.

- Although the distribution, ecology, and epidemiology of *Echinococcus multilocularis* in North America is still largely unknown, it does not appear to infect sheep or any other ungulates as it is maintained in a carnivore-rodent prey cycle (*110*).
- *Mesocestoides lineatus* has a three-stage lifecycle with two intermediate hosts and a large range of carnivorous mammals as definitive hosts (*19*). Human infections of the tapeworm *Mesocestoides lineatus* are rare but can occur through the consumption of chickens, snails, snakes, or frogs (*20*) and therefore it is unlikely that sheep will consume the intermediate life stages of this parasite.
- *Capillaria hepatica* (syn. *Calodium hepaticum*), is a globally distributed zoonotic parasite which uses rodents as main hosts, and while it is known to cause infection in over 180 mammalian species, a recent review of hosts did not include domesticated sheep (*21*). However, a recent study of *Capillaria* in Brazil identified two cases which were possibly caused by *C. hepatica* (*111*).
- The remaning links are discussed above.

Helminths – Phylogeny only

Host	Parasite	Status
Holochilus chacarius	Schistosoma mansoni	
Onychogalea unguifera	Echinococcus granulosus	
Canis adustus	Echinococcus granulosus	
Gazella leptoceros	Nematodirus spathiger	Confirmed
Kobus vardonii	Cotylophoron cotylophorum	
Onychogalea unguifera	Rugopharynx australis	
Lycalopex vetulus	Echinococcus granulosus	Unlikely
Canis mesomelas	Trichinella spiralis	Confirmed
Gazella leptoceros	Trichostrongylus vitrinus	
Onychogalea fraenata	Progamotaenia festiva	

Table SM 22: Top 10 undocumented links with highest probablity of interaction: Phylogeny only - helminth subset.

- *Schistosoma mansoni* infection in *Holochilus chacarius* was predicted by the phylogeny only model in the full dataset (discussed above).
- *Echinococcus granulosus* has not been reported to infect northern nail-tail wallabies (*Onychogalea unguifera*), other wallaby species including endangered bridled nail-tailed wallaby (*Onychogaela fraenata*) are involved in the transmission the parasite in Australia (*112*). *Onychogaela unguifera* may also be involved in *Echinococcosus* transmission, but its parasites may not be as well studied compared to the bridled nail-tail wallaby due to its stable conservation status.
- Similarly, *Rugopharynx australis* is known to infect multiple wallaby species, however the diversity of *Rugopharynx* and their susceptible hosts is still being discovered (*113*), suggesting that *Onychogalea ungifera* may be a promising target for future study.
- While there does not appear to be evidence of *Echinococcosus granulosus* infection in *Canis adustus*, other *Canis* species in Africa are known hosts (*114*), indicating that this should be a target for future surveillance.
- A recent molecular survey of gastrointestinal parasites of wild ruminants in Tunisia identified *Nematodirus spathiger* in engandered *Gazella leptoceros* that were genetically identical to those found in other domestic and wild ruminants (*115*). This is an example of a successful exploratory study aimed at describing the diversity of parasites in threatened species.
- We did not find much information on the parasites of the near threatened *Kobus vardonii*, however it is known to inhabit floodplains and grasslands near permanent water in south-central Africa (*116*) where is likely to be exposed to *Cotylophoron cotylophoron*, a “rumen fluke” which emerge from snail intermediate hosts and encyst on vegetation, later being ingested by ruminant definitive hosts in East Africa (*117*).

- *Echinococcus granulosus* is usually maintained by a domestic cycle of dogs eating raw livestock offal (*114*), and while its vertebrate-eating congener *Lycalopex gymnocercus* has been documented to host the parasite (*118*), *Lycalopex vetulus* is unlikely to become infected with *E. granulosus* as it has a largely insectivorous diet (*119*).

- *Canis mesomelas* has been reported with infection of *Trichinella spiralis* in the Kruger National Park, South Africa (*120*).

- *Gazella leptoceros* is also predicted to be susceptible to *Trichostrongylus vitrinus*. Although (*115*) did not identify this parasite in their study, *T. vitrinus* has been documented in lambs in Tunisia (*121*), indicating potential range overlap with *G. leptoceros*.

- *Progamotaenia festiva* is known to infect multiple *Onychogalea* species (*122*), but we could not find evidence of infection in the endangered *Onychogalea fraenata*.

- The remaining links are discussed above.

Helminths – Full model

Host	Parasite	Subset
Homo sapiens	Taenia mustelae	
Bos taurus	Hymenolepis diminuta	
Ovis aries	Hymenolepis diminuta	
Ovis aries	Echinococcus multilocularis	Unlikely
Capra hircus	Hymenolepis diminuta	
Bos taurus	Mesocestoides lineatus	Unlikely
Ovis aries	Mesocestoides lineatus	Unlikely
Bos taurus	Anisakis simplex	Unlikely
Ovis aries	Anisakis simplex	Unlikely
Sorex araneus	Echinococcus granulosus	

Table SM 23: Top 10 undocumented links with highest probablity of interaction: Full model - helminth subset.

- The natural hosts of *Hymenolepis diminuta* are rats (*18*) and we find no evidence of infection in goats.
- *Anisakis simplex* uses cetaceans as final hosts, with marine invertebrates and fish as intermediate hosts (*25*). Whales are infected through ingestion, indicating that while sheep may be susceptible, though they would need sufficient exposure to marine based feed.
- We could not find evidence of *Echonococcus granulosus* infection in Sorex araneus, however this species and other *Sorex sp.* are known hosts of *Echinococcus multilocularis* (*123*).
- The other top links are discussed above.

Protozoa – Affinity only model

Host	Parasite	Status
Bos taurus	Trypanosoma cruzi	
Ovis aries	Trypanosoma cruzi	Confirmed
Homo sapiens	Neospora caninum	
Diceros bicornis	Trypanosoma cruzi	Unlikely
Diceros bicornis	Giardia intestinalis	
Pan troglodytes	Toxoplasma gondii	Confirmed
Gorilla gorilla	Toxoplasma gondii	Confirmed
Equus caballus	Trypanosoma cruzi	Confirmed
Pan troglodytes	Trypanosoma cruzi	Unlikely
Gorilla gorilla	Trypanosoma cruzi	Unlikely

Table SM 24: Top 10 undocumented links with highest probablity of interaction: Affinity only - protozoa subset.

- As *T. cruzi* is currently restricted to the Americas (*15*), it is unlikely to infect black rhinos (*Diceros bicornis*) or gorillas (*Gorilla gorilla*) in natural conditions, unless exported through human movement.
- *Giardia* has been identified in a captive bred *Diceros bicornis* calf (*124*) in San Diego, indicating the potential for grey literature from zoo and captive breeding facilities to inform potential host-parasite interactions.
- *Toxoplasma gondii* has been documented to infect chimpanzees (*Pan troglodytes*), and interestingly appears to mirror the infection-induced behaviour in rodents and humans, with infected chimpanzees attracted to the urine of leopards, their only natural predator (*125*).
- Recent finding of a *Gorilla gorilla* individual seropositive for *T. gondii* at a primate center in Gabon (*126*).
- *T. cruzi* was recently identified in *Equus caballus*, marking the first evidence of infection in equids (*127*).
- Although *T. cruzi* naturally occurs in the Americas, and thus natural infection of chimpanzees (*Pan troglodytes*) is unlikely, a fatal infection was documented in a captive individual in Texas (*128*).

Protozoa – Phylogeny only

Host	Parasite	Subset
Canis aureus	Toxoplasma gondii	
Saguinus oedipus	Trypanosoma cruzi	Confirmed
Cuon alpinus	Toxoplasma gondii	
Lycaon pictus	Toxoplasma gondii	Confirmed
Saimiri oerstedii	Trypanosoma cruzi	Confirmed
Mazama gouazoubira	Toxoplasma gondii	Confirmed
Nyctereutes procyonoides	Toxoplasma gondii	Confirmed
Saguinus niger	Trypanosoma cruzi	Confirmed
Capricornis swinhoei	Toxoplasma gondii	
Panthera tigris	Toxoplasma gondii	Confirmed

Table SM 25: Top 10 undocumented links with highest probablity of interaction: Phylogeny only - protozoa subset.

- *Canis aureus* with antibodies against *T. gondii* have been identified in captive animals in the United Arab Emirates (*129*).
- Two recent reviews of parasites in non-human primates find no documented infection of the critically endangered Cotton-top tamarin (*Saguinus oedipus*) by *Trypanosoma cruzi*, although multiple *Saguinus sp.* have been documented with infections (*130, 131*). However, a 1982 study of Colombian monkeys and marmosets identified *S. oedipus* as a host for *T. cruzi* for the first time (*132*). This highlights the potential conservation importance of this parasite for *S. oedipus* and the need for periodic disease surveys of critically endangered species.
- Similarly, *Saimiri oerstedii* is not listed by these reviews as a host of *T. cruzi*, although a 1972 study identifies *S. oerstedii* as a reservoir for the parasite in Panama (*133*). While this report should be followed up with contemporary diagnostic methods, this reiterates the difficulty of exhaustively searching the literature for interaction data and the utility of link prediction methods to for directing these efforts.
- *Toxoplasmoa gondii* infection in *Cuon alpinus* has rarely been investigated, except for one captive individual which tested negative in serological testing (*134*).
- High prevalence of antibodies against *Toxoplasma gondii* was found in wild dogs (*Lycaon pictus*) in the Kruger National Park, South Africa, and was documented as causing a fatal infection in one pup (*135*), indicating that this parasite has the potential to influence the population dynamics of this endangered canid.
- *T. gondii* has been identified in *Mazama gouanzoubira* from French Guiana (*136*) and *Nyctereutes procyonoides* (*137*) in China via genetic sequencing.
- *Trypanosoma cruzi* is documented to infect *Saguinus niger* (*131*).

- We did not find any reports of *T. gondii* infection in Taiwan serow *Capricornis swinhoei*, although direct evidence of infection has been found in Japanese serow (*138*) and *T. gondii* has been found to infect multiple animals in Taiwan (*139*).
- The Siberian tiger *Panthera tigris altaica* acts as a definitive host for *T. gondii* and is observed to naturally shed oocysts (*140*).

Protozoa – Full model

Host	Parasite	Status
Bos taurus	Trypanosoma cruzi	
Ovis aries	Trypanosoma cruzi	Confirmed
Pan troglodytes	Toxoplasma gondii	Confirmed
Gorilla gorilla	Toxoplasma gondii	Confirmed
Diceros bicornis	Trypanosoma cruzi	Unlikely
Diceros bicornis	Giardia intestinalis	
Equus caballus	Trypanosoma cruzi	Confirmed
Pan troglodytes	Trypanosoma cruzi	Unlikely
Bubalus bubalis	Toxoplasma gondii	Confirmed
Papio anubis	Toxoplasma gondii	Confirmed

Table SM 26: Top 10 undocumented links with highest probablity of interaction: Full model - protozoa subset.

- The first eight links were predicted by models discussed above.
- *Bubalus bubalis* is a well established host of *Toxoplasma gondii*, though they are considered resistant to clinical toxoplasmosis (*141*).
- Natural infections of *Toxoplasma gondii* in *Papio anubis* were recently documented via genetic sequencing (*142*).

Viruses – Affinity only

Host	Parasite	Subset
Homo sapiens	Bluetongue virus	
Homo sapiens	Bovine viral diarrhea virus 1	
Homo sapiens	Carnivore protoparvovirus 1	
Homo sapiens	Simian immunodeficiency virus	Confirmed
Pan troglodytes	Rabies lyssavirus	Confirmed
Macaca mulatta	Rabies lyssavirus	Confirmed
Homo sapiens	Bovine alphaherpesvirus 1	
Homo sapiens	Canine mastadenovirus a	
Cervus elaphus	Rabies lyssavirus	Confirmed
Homo sapiens	Ovine herpesvirus 2	

Table SM 27: Top 10 undocumented links with highest probablity of interaction: Affinity only - virus subset.

- Simian immunodificiency virus strains from wild primates have previously shifted to infect humans and are responsible for the AIDS pandemic (from HIV-1) (*143*).
- Rabies positive *Macaca mulatta* have recently been reported in India (*144*).
- Bovine alphaherpesvirus 1, the main casual agent of infectious bovine rhinotracheitis, is largely restricted to cattle and not currently considered to infect humans (*145*).
- *Canine mastadenovirus a*, formerly *Canine adenovirus 1* is known to infect dogs and circulate in wild carnivores (*10*), though we could not find evidence of human infection.
- *Ovine herpesvirus 2*, the casual agent of sheep associated malignant catarrhal fever, is asymptomatic in its natural hosts and cause severe disease in susceptible animals that are dead end hosts, however to date there is no evidence of infection in humans (*146*).
- The remaining links are discussed above.

Viruses – Phylogeny only

Host	Parasite	Status
Vulpes macrotis	Rabies lyssavirus	Confirmed
Lycalopex culpaeus	Rabies lyssavirus	
Lycalopex griseus	Rabies lyssavirus	Confirmed
Lycalopex gymnocercus	Rabies lyssavirus	
Myotis nattereri	Rabies lyssavirus	Confirmed
Myotis blythii	Rabies lyssavirus	Confirmed
Myotis myotis	Rabies lyssavirus	Confirmed
Myotis macrodactylus	Rabies lyssavirus	
Myotis mystacinus	Rabies lyssavirus	
Myotis dasycneme	Rabies lyssavirus	Confirmed

Table SM 28: Top 10 undocumented links with highest probablity of interaction: PHylogeny only - virus subset.

- The first four links were predicted by previous models and discussed above.
- Rabies has been isolated from a single *Myotis nattereri* individual in France (*147*).
- In 2017, an individual *Myotis blythii* from Croatia tested positive for antibodies against rabies (*148*).
- Rabies has been detected in *Myotis myotis* in a few European countries (*149*).
- We did not find evidence of rabies infection in *Myotis macrodactylus* or *Myotis mystacinus*.
- Identification of rabies positive *Myotis dasycneme* in the Netherlands (*150*).

Viruses – Full model

Host	Parasite	Subset
Pan troglodytes	Rabies lyssavirus	Confirmed
Macaca mulatta	Rabies lyssavirus	Confirmed
Cervus elaphus	Rabies lyssavirus	Confirmed
Homo sapiens	Bluetongue virus	
Macaca fascicularis	Rabies lyssavirus	Confirmed
Capreolus capreolus	Rabies lyssavirus	Confirmed
Rattus norvegicus	Rabies lyssavirus	Confirmed
Rattus rattus	Rabies lyssavirus	Confirmed
Chlorocebus aethiops	Rabies lyssavirus	
Sigmodon hispidus	Rabies lyssavirus	Confirmed

Table SM 29: Top 10 undocumented links with highest probablity of interaction: Full model - virus subset

- *Macaca fascicularis* is known to be susceptible to infection by *Rabies lyssavirus* and regularly used for development of rabies vaccines (*151*).
- We did not find evidence of rabies infection in *Chlorcebus aethiops*, however several cercopithecine monkeys are known to be susceptible to infection (*30*).
- *Sigmodon hispidus* has successfully been experimentally infected with rabies (*152*).
- The remaining links are discussed above.

References

1. L. J. Harmon, *et al.*, *Evolution* **64**, 2385 (2010).
2. A. W. Park, *et al.*, *Proceedings of the Royal Society B: Biological Sciences* **285** (2018).
3. M. Nakao, *et al.*, *International Journal for Parasitology* **43**, 427 (2013).
4. L. M. Lee, *et al.*, *Emerging Infectious Diseases* **22**, 707 (2016).
5. A. Spickler, Bluetongue, *Tech. rep.*, Institute for International Cooperation in Animal Biologies (2015).
6. P. H. Walz, *et al.*, *Journal of Veterinary Internal Medicine* **24**, 476 (2010).
7. J. P. Dubey, G. Schares, L. M. Ortega-Mora, *Clinical Microbiology Reviews* **20**, 323 (2007).
8. J. W. Mcgarry, A. Higgins, N. G. White, K. C. Pounder, U. Hetzel, *Zoonoses and Public Health* **62**, 44 (2015).
9. S.-M. Guk, J.-L. Kim, J.-H. Park, J.-Y. Chai, *Journal of Parasitology* **93**, 1225 (2007).
10. A. Balboni, *et al.*, *BMC Veterinary Research* **14**, 1 (2018).
11. T. M. Murphy, *et al.*, *Parasitology Research* **111**, 283 (2012).
12. M. A. Betancourt-Alonso, A. Orihuela, V. Aguirre, R. Vázquez, I. Flores-Pérez, *World Rabbit Science* **19**, 21 (2011).
13. C. T. McAllister, C. R. Bursey, J. F. Roberts, *Journal of Herpetological Medicine and Surgery* **14**, 10 (2004).
14. M. Giangaspero, *Clinical Microbiology: Open Access* **06**, 20 (2017).
15. A. J. Browne, *et al.*, *Scientific Data* **4**, 1 (2017).
16. R. E. Gürtler, M. V. Cardinal, *Acta Tropica* **151**, 32 (2015).
17. J. L. Fitzpatrick, J. L. Dyer, J. D. Blanton, I. V. Kuzmin, C. E. Rupprecht, *Journal of the American Veterinary Medical Association* **245**, 333 (2014).
18. S. Huang, O. R. P. Bininda-Emonds, P. R. Stephens, J. L. Gittleman, S. Altizer, *Journal of Animal Ecology* **83**, 671 (2014).
19. S. H. Cho, T. S. Kim, Y. Kong, B. K. Na, W. M. Sohn, *Korean Journal of Parasitology* **51**, 531 (2013).

20. A. Ito, C. M. Budke, *Travel Medicine and Infectious Disease* **12**, 582 (2014).

21. H. P. Fuehrer, *Parasitology Research* **113**, 641 (2014).

22. A. Kopp, *et al.*, *mBio* **4**, 1 (2013).

23. L. A. Diaz, *et al.*, *Emerging Infectious Diseases* **12**, 1752 (2006).

24. A. Beineke, W. Baumgärtner, P. Wohlsein, *One Health* **1**, 49 (2015).

25. K. Buchmann, F. Mehrdana, *Food and Waterborne Parasitology* **4**, 13 (2016).

26. N. Mollentze, R. Biek, D. G. Streicker, *Current Opinion in Virology* **8**, 68 (2014).

27. M. G. Birhane, *et al.*, *Journal of the American Veterinary Medical Association* **250**, 1117 (2017).

28. World Health Organization, *Rabies Bulletin Europe* **7**, 1 (1983).

29. J. Sempere, V. E. Sokolov, A. A. Danilkin, *Mammalian Species* pp. 1–9 (1996).

30. P. Gautret, *et al.*, *PLoS Neglected Tropical Diseases* **8**, 1 (2014).

31. F. Steck, A. Wandeler, *Epidemiologic reviews* **2**, 71 (1980).

32. H. Zhu, *et al.*, *Infection, Genetics and Evolution* **31**, 127 (2015).

33. E. Schindler, *Zeitschrift fur Tropenmedizin und Parasitologie* **8**, 233 (1957).

34. P. J. Fleming, *et al.*, *Food Webs* **12**, 14 (2017).

35. P. J. White, M. T. Hanson, W. H. Berry, J. J. Eliason, *Southwestern Naturalist* **45**, 204 (2000).

36. Z. Wang, X. Wang, Q. Lu, *Acta Theriologica* **52**, 215 (2007).

37. X. Y. Tao, *et al.*, *PLoS Neglected Tropical Diseases* **9**, 1 (2015).

38. G. Zapata-Ríos, L. C. Branch, *Biological Conservation* **193**, 9 (2016).

39. E. Silva-Rodríguez, *et al.*, Lycalopex fulvipes. The IUCN Red List of Threatened Species 2016 (2016).

40. M. F. C. Juan, C. Durn. R., *Avances en Ciencias Veterinarias* **4** (1989).

41. D. E. Wilson, D. M. Reeder, *Mammal Species of the World: A Taxonomic and Geographic Reference* (Johns Hopkins University Press, Baltimore, Maryland, 2005), third edn.

42. P. Carnieli, *et al.*, *Virus Research* **131**, 33 (2008).

43. L. Durbin, A. Venkataraman, S. Hedges, W. Duckworth, *Canids: Foxes, Wolves, Jackals and Dogs*, C. Sillero-Zubiri, M. Hoffmann, D. Macdonald, eds. (Island Press, 2005), chap. 8, pp. 210–219.

44. K. E. DeMatteo, *Canid News* **11**, 1 (2008).

45. R. S. P. Jorge, *et al.*, *Journal of Wildlife Diseases* **46**, 1310 (2010).

46. C. E. Borda, M. J. Rea, *Memorias do Instituto Oswaldo Cruz* **101**, 233 (2006).

47. C. L. Stoltenow, K. Solemsass, M. Niezgoda, P. Yager, C. E. Rupprecht, *Journal of Wildlife Diseases* **36**, 169 (2000).

48. D. Joshi, *H.M. Government Press,* **XVII**, 145 (1982).

49. C. Mei, *et al.*, *Scientific Reports* **6**, 1 (2016).

50. M. Fowler, *Revue scientifique et technique (OIE)* **15**, 155 (1996).

51. A. B. Carey, R. G. Mclean, *Journal of Applied Ecology* **20**, 777 (1983).

52. M. Eidson, *et al.*, *Journal of the American Veterinary Medical Association* **227**, 932 (2005).

53. K. Li, *et al.*, *Veterinary Parasitology* **205**, 354 (2014).

54. M. Grover, Spatio-temporal analysis of dog ecology and rabies epidemiology at a wildlife interface in the Lowveld Region of South Africa (Dissertation (MSc)–University of Pretoria, 2016), Ph.D. thesis (2016).

55. F. Jongejan, G. Uilenberg, *Revue scientifique et technique (OIE)* **13**, 1201 (1994).

56. A. Walker, *et al.*, *Ticks of Domestic Animals in Africa: a Guide to Identification of Species* (Bioscience Reports, Edinburgh, 2003).

57. S. Bornstein, T. Mörner, W. M. Samuel, *Parasitic Diseases of Wild Mammals*, W. M. Samuel, Margo J. Pybus, A. A. Kocan, eds. (Iowa Sate University Press, Ames, Iowa, 2002), pp. 107–119, second edn.

58. J. M. Kagira, *et al.*, *ISRN Veterinary Science* p. 650890 (2013).

59. N. M. Nemeth, *et al.*, *Veterinary Pathology* **51**, 633 (2014).

60. A. J. Loveridge, D. W. Macdonald, *Journal of Zoology* **259**, 143 (2003).

61. J. Dalponte, O. Courtenay, Lycalopex vetulus. The IUCN Red List of Threatened Species 2008 (2008).

62. A. Angerbjörn, M. Tannerfeldt, Vulpes lagopus. The IUCN Red List of Threatened Species 2014 . (2014).

63. P. C. Buckland, J. P. Sadler, *Journal of Biogeography* **16**, 115 (1989).

64. J. a. D. Luque, H. Müller, L. González, I. Berkunsky, *Mastozoologia Neotropical* **21**, 135 (2014).

65. R. Jorge, E. Lima, L. Lucarts, *XXXIII Congresso Anual da Sociedade de Zoológicos do Brasil, At Sorocaba, SP, Brasil* (2008).

66. E. de Souza Lima, *et al.*, *Wildlife Research* **39**, 512 (2012).

67. J. V. Irons, R. B. Eads, C. W. Johnson, O. L. Walker, M. A. Norris, *The Journal of Parasitology* **38**, 1 (1952).

68. G. B. Schaller, A. Kang, *Oryx* **42**, 100 (2008).

69. M. A. Criffield, M. V. Reichard, E. C. Hellgren, D. M. Leslie, K. Freel, *The Southwestern Naturalist* **54**, 492 (2009).

70. J. W. Dragoo, *et al.*, *American Society of Mammalogists* **71**, 318 (1990).

71. B. L. Cypher, *et al.*, *Journal of Wildlife Diseases* **53**, 46 (2017).

72. J. Oksi, *et al.*, *Journal of Clinical Microbiology* **51**, 2781 (2013).

73. V. Rar, I. Golovljova, *Infection, Genetics and Evolution* **11**, 1842 (2011).

74. J. Yang, *et al.*, *Parasites and Vectors* **10** (2017).

75. R. Gonçalves, Does Mycoplasma mycoides subsp. mycoides LC cause human infection?, *Tech. rep.*, Instituto Nacional de Investigação Agrária e Veterinária (INIAV) (2007).

76. A. P. dos Santos, *et al.*, *Emerging Infectious Diseases* **14**, 1922 (2008).

77. S. Barati, N. Moori-Bakhtiari, M. G. Najafabadi, H. Momtaz, L. Shokuhizadeh, *Iranian Journal of Microbiology* **9**, 288 (2017).

78. N. Pusterla, C. Gebhart, *Equine Veterinary Education* **21**, 415 (2009).

79. I. Lysnyansky, *et al.*, *Israel Journal of Veterinary Medicine* **62**, 79 (2007).

80. I. Sandal, T. J. Inzana, *Trends in Microbiology* **18**, 90 (2010).

81. J. L. Siembieda, R. A. Kock, T. A. McCracken, S. H. Newman, *Animal health research reviews / Conference of Research Workers in Animal Diseases* **12**, 95 (2011).

82. J. van der Hoeden, *Journal of Comparative Pathology and Therapeutics* **65**, 278 (1955).

83. T. Dayan, D. Simberloff, E. Tchernov, Y. Yom-Tov, *Biological Journal of the Linnean Society* **45**, 315 (1992).

84. K. J. Allan, *et al.*, *PLoS Neglected Tropical Diseases* **9**, 1 (2015).

85. B. L. Penzhorn, *et al.*, *Parasites and Vectors* **11**, 1 (2018).

86. O. Tenaillon, D. Skurnik, B. Picard, E. Denamur, *Nature Reviews Microbiology* **8**, 207 (2010).

87. M. T. Jay-Russell, A. F. Hake, Y. Bengson, A. Thiptara, T. Nguyen, *PLoS ONE* **9** (2014).

88. T. King, Wild and domestic animals as reservoirs of antibiotic resistant Escherichia coli in South Africa, Ph.D. thesis (2017).

89. A. A. Katakweba, *et al.*, *Journal of Applied Microbiology* **118**, 966 (2015).

90. M.-A. Frank, Antibiotic resistance in enteric Escherichia coli and Enterococcus spp. isolated from ungulates at Marwell Zoo England, Ph.D. thesis, University of Pretoria (2016).

91. J. B. Clayton, J. L. Danzeisen, A. M. Trent, T. Murphy, T. J. Johnson, *Frontiers in Veterinary Science* **1**, 1 (2014).

92. W. Standley, P. McCue, *Department of Defense Techincal Report* .

93. J. A. Langholz, T. J. R. Michele, *Human-Wildlife Interactions* **7**, 140 (2013).

94. S. Reinstein, J. T. Fox, X. Shi, M. J. Alam, T. G. Nagaraja, *Journal of Food Protection* **70**, 2555 (2016).

95. D. V. Cousins, *Revue scientifique et technique (OIE)* **20**, 71 (2001).

96. R. K. Clark, W. M. Boyce, D. A. Jessup, L. F. Elliott, *Journal of Zoo and Wildlife Medicine* **24**, 48 (1993).

97. S. Stuen, E. G. Granquist, C. Silaghi, *Frontiers in Cellular and Infection Microbiology* **3**, 1 (2013).

98. J. de la Fuente, *et al.*, *Journal of Wildlife Diseases* **42**, 381 (2006).

99. D. J. Greig, *et al.*, *Diseases of Aquatic Organisms* **111**, 93 (2014).

100. L. Warnecke, *et al.*, *Proceedings of the National Academy of Sciences* **109**, 6999 (2012).

101. G. M. Anstead, D. A. Sutton, J. R. Graybill, *Journal of Clinical Microbiology* **50**, 1346 (2012).

102. F. H. Gleason, A. V. Marano, *Hydrobiologia* **659**, 81 (2011).

103. A. M. Aluoch, *et al.*, *South African Journal of Botany* **111**, 182 (2017).

104. P. Danesi, *et al.*, *Fungal Biology* **120**, 191 (2016).

105. R. J. Palmer, O. P. Settnes, J. Lodal, A. E. Wakefield, *Applied and Environmental Microbiology* **66**, 4954 (2000).

106. C. Campbell, A. Borman, C. Linton, P. Bridge, E. Johnson, *Medical Mycology* **44**, 451 (2006).

107. M. Ma, *et al.*, *MicrobiologyOpen* **e597**, 1 (2018).

108. O. Settnes, S. A. Henriksen, *Acta veterinaria Scandinavica* **30**, 437 (1989).

109. J. Laakkonen, *International Journal for Parasitology* **28**, 241 (1998).

110. A. Massolo, S. Liccioli, C. Budke, C. Klein, *Parasite* **21**, 73 (2014).

111. P. E. F. Teixeira, *et al.*, *Revista Brasileira de Parasitologia Veterinaria* **27**, 226 (2018).

112. D. J. Jenkins, C. N. Macpherson, *Parasitology* **127**, S63 (2003).

113. N. B. Chilton, F. Huby-Chilton, A. V. Koehler, R. B. Gasser, I. Beveridge, *International Journal for Parasitology: Parasites and Wildlife* **5**, 124 (2016).

114. B. Otero-Abad, P. R. Torgerson, *PLoS Neglected Tropical Diseases* **7** (2013).

115. Y. Said, M. Gharbi, M. Mhadhbi, M. Dhibi, S. Lahmar, *Parasitology* **145**, 901 (2018).

116. IUCN SSC Antelope Specialist Group, Kobus vardonii, The IUCN Red List of Threatened Species 2016 (2016).

117. M. R. Laidemitt, *et al.*, *Parasitology* **144**, 131 (2017).

118. M. Lucherini, E. M. Luengos Vidal, *Mammalian Species* **820**, 1 (2008).

119. J. C. Dalponte, *Mammalian Species* **847**, 1 (2009).

120. E. Young, S. P. Kruger, *Journal of the South African Veterinary Association* **38**, 441 (1967).

121. H. Akkari, M. Gharbi, M. a. Darghouth, *Parasite* **19**, 407 (2012).

122. I. Beveridge, S. Shamsi, *Zootaxa* **1990**, 1 (2009).

123. M. Šarknas, *et al.*, *Veterinary Parasitology* **213**, 121 (2015).

124. D. C. Wagner, M. S. Edwards, *San Diego Animal Park* pp. 1–10 (1984).

125. C. Poirotte, *et al.*, *Current Biology* **26**, R98 (2016).

126. J. P. Akue, *et al.*, *Journal of Parasitology and Vector Biology* **10**, 1 (2018).

127. L. K. Bryan, *et al.*, *Veterinary Parasitology* **216**, 13 (2016).

128. Y. R. Bommineni, E. J. Dick, J. S. Estep, J. L. Van de Berg, G. B. Hubbard, *Journal of Medical Primatology* **38**, 247 (2009).

129. J. P. Dubey, *et al.*, *Veterinary Parasitology* **172**, 195 (2010).

130. K. Strait, J. G. Else, M. L. Eberhard, *Parasitic Diseases of Nonhuman Primates* (Elsevier, 2012), second edn.

131. B. Solórzano-García, G. Pérez-Ponce de León, *International Journal of Primatology* **39**, 155 (2018).

132. C. J. Marinkelle, *Annals of Tropical Medicine & Parasitology* **76**, 121 (1982). PMID: 6807227.

133. O. E. Sousa, *Revista de biologia tropical* **20**, 167 (1972).

134. S. Y. Zhang, M. X. Wei, Z. Y. Zhou, J. Y. Yu, X. Q. Shi, *Parasitology International* **49**, 171 (2000).

135. J. Van Heerden, M. G. Mills, M. J. Van Vuuren, P. J. Kelly, M. J. Dreyer, *Journal of the South African Veterinary Association* **66**, 18 (1995).

136. A. Mercier, *et al.*, *Infection, Genetics and Evolution* **11**, 1378 (2011).

137. D. H. Zhou, *et al.*, *Acta Tropica* **172**, 143 (2017).

138. C. Sakae, T. Ishida, *Journal of Parasitology* **98**, 224 (2012).

139. J.-C. Chen, Y.-J. Tsai, Y.-L. Wu, *Research in Veterinary Science* **102**, 184 (2015).

140. S. A. Elmore, *et al.*, *Trends in Parasitology* **26**, 190 (2010).

141. C. Alvarado-Esquivel, *et al.*, *BMC Veterinary Research* **10**, 1 (2014).

142. D. Kamau, *et al.*, *The 11th Jkuat Scientific, Technological and Industrialization Conference and Exhibitions Conference Proceedings* (2016).

143. F. Van Heuverswyn, *et al.*, *Nature* **444**, 164 (2006).

144. O. K. Bharti, *Journal of Travel Medicine* **23**, 1 (2016).

145. S. More, *et al.*, *EFSA Journal* **15**, 1 (2017).

146. B. Sharma, S. Parul, G. Basak, R. Mishra, *Journal of Entomology and Zoology Studies* **7**, 26 (2019).

147. E. Picard-Meyer, *et al.*, *PLoS ONE* **9** (2014).

148. I. Šimić, N. Krešić, I. Lojkić, T. Bedeković, *Rabies Serology Meeting* (2017).

149. J. Schatz, *et al.*, *Zoonoses and Public Health* **60**, 22 (2013).

150. J. Nieuwenhuijs, J. Haagsma, P. Lina, *Revue scientifique et technique (OIE)* **11**, 1155 (1992).

151. P. T. Ullas, A. Desai, S. N. Madhusudana, *World Journal of Vaccines* **02**, 36 (2012).

152. W. G. Winkler, N. J. Schneider, W. L. Jennings, *Journal of Wildlife Diseases* **8**, 99 (1972).