

Phylogenetic and biogeographical traits predict unrecognized hosts of zoonotic leishmaniasis

Caroline K. Glidden^{1*}, Aisling Roya Murrin¹, Rafaella Albuquerque Silva², Adrian A. Castellanos³, Barbara A. Han³, Erin A. Mordecai¹

¹Department of Biology, Stanford University, Stanford, California, United States of America 94305

² Secretaria de Vigilância em Saúde, Ministério da Saúde do Brasil, Distrito Federal, Brasília

³Cary Institute of Ecosystem Studies, Millbrook, New York, United States of America 12545

*cglidden@stanford.edu

Abstract

The spatio-temporal distribution of leishmaniasis, a parasitic vector-borne zoonotic disease, is significantly impacted by land-use change and climate warming in the Americas. However, predicting and containing outbreaks is challenging as the zoonotic *Leishmania* system is highly complex: leishmaniasis (visceral, cutaneous and mucocutaneous) in humans is caused by up to 14 different *Leishmania* species, and the parasite is transmitted by dozens of sandfly species and is known to infect almost twice as many wildlife species. Despite the already broad known host range, new hosts are discovered almost annually and *Leishmania* transmission to humans occurs in absence of a known host. As such, the full range of *Leishmania* hosts is undetermined, inhibiting the use of ecological interventions to limit pathogen spread and the ability to accurately predicting the impact of global change on disease risk. Here, we employed a machine learning approach to generate trait profiles of known zoonotic *Leishmania* wildlife hosts (mammals that are naturally exposed and susceptible to infection) and used trait-profiles of known hosts to identify potentially unrecognized hosts. We found that biogeography, phylogenetic distance, and study effort best predicted *Leishmania* host status. Traits associated with global change, such as agricultural land-cover, urban land-cover, and climate, were among the top predictors of host status. Most notably, our analysis suggested that zoonotic *Leishmania* hosts are significantly undersampled, as our model predicted just as many unrecognized hosts as unknown hosts. Overall, our analysis facilitates targeted surveillance strategies and improved understanding of the impact of environmental change on local transmission cycles.

Author summary

Leishmaniasis is a zoonotic, vector borne disease of poverty with a high burden throughout the Americas: within Latin America there are an estimated 58,500 new cases per year and 54,050 years of life lost due to disability. Although the World Health Organization has targeted leishmaniasis for elimination and control by 2030, the disease remains a persistent threat. Across the Americas, particularly in Central America, the southeastern United States, and perimeters of the Amazon Basin, risk of infection is increasing in geographic extent and elevation. While it is known that *Leishmania* parasites, the causative agent of leishmaniasis, are maintained in the environment via a mammalian host, the full suite of wildlife hosts has yet to be documented, which significantly hinders control efforts. Here, we use machine learning and ecological and evolutionary trait profiles of known hosts to identify unrecognized potential wildlife hosts of *Leishmania*. We identify 136 mammals in the Americas that are likely to be exposed to and infected by zoonotic *Leishmania* in the wild. The high number of unrecognized potential hosts emphasizes a need to better invest in studying the ecological epidemiology of leishmaniasis. The study provides information and tools to support targeted intervention and management of this important poverty-associated disease.

Introduction

Leishmaniasis, a debilitating and sometimes fatal parasitic disease, is one of the 20 neglected tropical diseases targeted for control and elimination by the United Nations by 2030 (1). However, leishmaniasis remains a persistent threat throughout the tropics and subtropics. In the Americas, leishmaniasis cases occur from the southern United States to Argentina, including the Caribbean, with an estimated 58,500 new cases each year (2). Further, incidence is rapidly shifting in geographic extent with land-use and climate change (3–7). For instance, hotspots of leishmaniasis have significantly expanded throughout Costa Rica in the last two decades, in concert with agricultural intensification, deforestation, and reforestation (3).

Treatment for leishmaniasis has many side effects and are extremely expensive (2,8), thus control and elimination hinges on limiting transmission. *Leishmania* parasites are transmitted by sandflies (Order: Diptera; Family: Psychodidae) and maintained in the environment via a multitude of sylvatic and domestic mammalian hosts. As such, controlling transmission requires ecological interventions targeting either component of the transmission cycles. For instance, clearing vector habitat proximate to households, supporting populations of natural enemies that regulate populations of vectors or small mammals (e.g., fish and mosquito larvae), and/or modifying distribution of wildlife food sources are sustainable and effective interventions in preventing transmission of flaviviruses, henipaviruses, *Plasmodium* spp. (malaria), and *Brucella abortus* (brucellosis) (9,10). Similar interventions could be employed against *Leishmania*, however, successful application requires species identification of vectors, hosts, and their corresponding interactions with their biotic and abiotic environment.

Leishmaniasis was first diagnosed by Western medicine in the Americas in 1909, at the time dogs and foxes were believed to be the zoonotic hosts (11). Continued research throughout the century hypothesized that rodents, and in some cases opossums and sloths, were the primary zoonotic hosts of these pathogens (11). However, especially with the advancement of molecular diagnostic techniques, the number and taxonomic diversity of hosts has since expeditiously increased. In the Americas, there are now 137 species of wildlife that are recognized as potential hosts (i.e., animals that are naturally exposed and become infected once exposed), of which around 60 may act as competent hosts (i.e., animals can become infectious once infected, and may maintain the pathogen in the environment as reservoir hosts) (12,13). Wild hosts are now known to include eight orders of mammals including Chiroptera (bats), Carnivora, Cingulata (armadillos), Didelphimorphia (opossums), Lagomorpha (rabbits), Pilosa (sloths and anteaters), Primates, and Rodentia. However, even with progress in host discovery, the full suite of *Leishmania* hosts is not yet described: *Leishmania* spillover occurs in absence of known hosts and new hosts are discovered almost annually (14–17) .

115

116 Limited knowledge on the transmission cycle makes it difficult to understand the
117 mechanisms driving *Leishmania* distribution and spillover. As previously stated, unknown
118 transmission cycles impede effective ecological interventions to reduce spillover risk to
119 people. Additionally, global change can alter host dynamics to amplify spillover risk. For
120 example, land-use change may increase host densities and/or push hosts into closer
121 proximity to anthropophilic vectors and humans (9). Climate warming may also shrink or
122 expand the range of hosts, further changing the rate of contact among hosts, vectors, and
123 humans (18). As such, the incomplete classification of possible *Leishmania* hosts could
124 lead to unexpected emergence and hotspots of human leishmaniasis.

125

126 Study designs optimal for detecting *Leishmania* hosts, such as systematic sampling of
127 large populations of a diversity of species through space and time, are limited by logistics
128 and cost of sampling. *Leishmania* host discovery could benefit from strong *a priori*
129 hypotheses about which mammals to target for sampling effort; taxa specific sampling
130 could help to efficiently use time and monetary resources, ultimately expediting the full
131 classification of the *Leishmania* transmission cycle. Through the last decade, machine
132 learning approaches have been used to generate predictions of hosts of a multitude of
133 different pathogens, including betacoronaviruses (19), flaviviruses (20), Rabies virus (21),
134 and *Borrelia burgdoferi* (causative agent of Lyme disease) (22). These methods have
135 facilitated more targeted laboratory and field work of novel pathogens, and they have great
136 potential to improve our understanding of diseases of poverty with broad host and vector
137 ranges.

138

139 Host status for any parasite or pathogen is driven by two processes: exposure and
140 physiological susceptibility (23,24). Exposure depends on the environmental conditions
141 needed to support pathogen transmission and, in the case of vector-borne disease, vector
142 reproduction, contact with competent hosts, and vector survival (25). Following exposure,
143 physiological susceptibility then depends on pathogen interaction with host cells to gain

entry and, in some cases, avoid immune attack and replicate (25). Host traits related to exposure and/or susceptibility likely interact in non-linear and higher-order combinations to delineate zoonotic *Leishmania* hosts from non-hosts. In brief, the effect of a trait on host status may vary in magnitude or direction depending on trait value and traits may become more or less important under different conditions. Hypothetically, for example, the average temperature of a species range may have a positive effect on host status between 25-30 °C, but a negative effect outside of this range. Depending on biting activity of vector species, animals that forage on the ground may be more likely to be hosts if they are also nocturnal, whereas animals that are arboreal may be more likely to be hosts if they are crepuscular but less likely to be hosts if they are nocturnal. Thus, when viewing the zoonotic *Leishmania* system in one or two dimensions (e.g., only considering phylogeny or habitat use), it is difficult to determine unifying traits useful for identifying hosts. To date, it has been challenging to combine these processes across scales from physiological to evolutionary to ecological to biogeographic to predict their overall impact on infection potential. Machine learning, particularly tree-based methods, offers a set of new tools that can be used to incorporate and dissect this complexity by increasing the dimensionality in which the *Leishmania* system can be studied, allowing for multiple predictors that may have nonlinear and interactive effects on host status. The flexibility of these models allows for pinpointing combinations of mammalian traits unique to the confluence of *Leishmania* exposure and susceptibility, by learning patterns of known hosts. With this information, we can identify likely host species as targets for *Leishmania* surveillance efforts and the key traits that distinguish host species from non-hosts.

Here we use tree-based machine learning (extreme gradient boosted regression, 'Xgboost') and mammalian traits, including but not limited to biogeographical, phylogenetic, and life-history features, to: (i) describe trait profiles of hosts for zoonotic *Leishmania* (*Viannia*) and *Leishmania* (*Leishmania*) parasites; and (ii) use the trait profiles to predict unrecognized wild hosts of zoonotic *Leishmania* within the subgenera *Leishmania* (*Viannia*) and *Leishmania* (*Leishmania*).

174 We aim for this analysis to identify animal species that are likely to be exposed and
175 infected in the wild; due to extreme data sparsity, we do not employ our methods to identify
176 competent hosts (i.e., animals that become infectious once infected and may act as
177 reservoirs given the density and distribution of their populations). Our findings can be used
178 to identify hosts that can then be surveyed at larger spatio-temporal scales, investigated
179 for competence, or used as sentinel species. As such, while our modeling approach does
180 not necessarily identify competent hosts that can transmit zoonotic *Leishmania* on to
181 sandflies and eventually humans, it does identify animals likely to be exposed to and
182 susceptible to zoonotic *Leishmania* infection, facilitating the first step of determining
183 epidemiological importance and fine-tuning surveillance efforts.

184

185 **Methods**

186 **Data collection**

187 Leishmaniasis is caused by protists within the *Leishmania* genus. There are 14 known species
188 of zoonotic *Leishmania* (i.e., they spill over from animals to humans) in the Americas. These
189 *Leishmania* species are divided between two subgenera: *Leishmania* (*Viannia*) and *Leishmania*
190 (*Leishmania*). In the Americas, *Leishmania* (*Viannia*) consists of ten species, nine of which are
191 zoonotic; *Leishmania* (*Leishmania*) consists of seven species, five of which are zoonotic (26).
192 *Leishmania* (*Viannia*) spp. cause cutaneous and mucocutaneous leishmaniasis, which
193 manifests as skin lesions that are susceptible to painful secondary bacterial infections, and/or
194 destruction of the mucus membranes of the nose, mouth, and throat (2). In addition to
195 cutaneous and mucocutaneous leishmaniasis, species within the *Leishmania* (*Leishmania*)
196 subgenera (*L. (L.) infantum* and, less frequently, *L. (L.) mexicana*) cause fatal visceral
197 leishmaniasis. Visceral leishmaniasis affects internal organs, typically causing enlargement of
198 the spleen and liver (2). The following analyses were conducted for each subgenus (*L. (Viannia)*
199 and *L. (Leishmania)*) so to better described the ecological associations of each taxonomic group
200 of parasites and their corresponding disease manifestations.

201
202 *Zoonotic Leishmania host status:* We gathered *Leishmania* host status from all wild,
203 endemic, and invasive terrestrial mammals with ranges in Mexico, Central America, and
204 South America using recent reviews(12,13) in addition to a Web of Science query on
205 February 12, 2021 and February 22, 2021 (see supplementary text 1 for the specific
206 search strings). Host status was defined as a binary trait: an animal has been naturally
207 infected by one of the 14 *Leishmania* species known to cause leishmaniasis in humans
208 (1), or there is no record of infection by a *Leishmania* species known to cause
209 leishmaniasis in humans (0). The latter category (“non-positives”) includes animals that
210 have been tested for *Leishmania* and have not been found to be infected as well as those
211 that have not been tested for *Leishmania* infection. Ideally, the analysis would include true-
212 positives (1) and true negatives (0). However, due to the intensive, systematic sampling
213 required to declare an animal as a true-negative (longitudinal sampling across large
214 geographic scales), there is not enough data to conduct the analysis with true-negatives.
215 Our analysis (using true-positives and non-positives) still enables us to identify species
216 that have not yet been tested but have a high probability of being a host. Infection was
217 determined via detection of *Leishmania* species specific antigens or genetic material.
218 Species identification was performed either using animal tissue samples directly or using
219 live parasites cultured from animal tissue. To ensure that all relevant mammals were
220 included in our study, a list of species names of endemic Latin American mammals was
221 retrieved from the International Union for Conservation of Nature (IUCN) Red List
222 Inventory (27) and a list of invasive mammals was retrieved from the Global Invasive
223 Species Database (28). We cross-referenced our host status data table with the Global
224 Infectious Disease and Epidemiology Network (29). We then limited our analysis to wild
225 mammals that had at least a 10% range overlap with reported human cases of
226 leishmaniasis for inclusion in our analysis. In total, 86.64% (1460/1685) of species were
227 retained for the *L. Viannia* analysis and 87.24% (1470/1685) of species were retained for
228 the *L. Leishmania* analysis. Zoonotic *Leishmania* ranges were constructed by outlining a

concave polygon around cases of disease occurrence in humans (30) (Fig 1a, 1c). We conducted all data filtering and downstream analyses separately for each subgenus.

Traits associated with Leishmania exposure and susceptibility: Our analysis leverages patterns of traits of known hosts to predict identities of unrecognized hosts. We used host traits related to *Leishmania* exposure (via sandfly bites) and susceptibility, including life-history, biogeographical, and phylogenetic traits (S1 Table).

For life-history and some habitat use traits, we collected traits from panTHERIA (31), mammalDIET (32), and EltonTraits (33). We additionally used IUCN range maps and satellite imagery to extract data associated with species biogeography, including variables related to climate (34,35), land-cover type (36), and global human modification index, an aggregate measure of anthropogenic pressure on a landscape (37), within each species range (see S1 Table for full description of traits and satellite imagery used). Ranges of endemic species were retrieved from the IUCN database and ranges of invasive species in Mexico, Central America, and South America were constructed by building a concave polygon around species occurrence points from (38). *Leishmania* and invasive mammal species ranges were built using the R (v 4.0.2) packages *speciesgeocodeR* (39) and *concaveman* (40). We additionally included IUCN-designated main habitat and habitat breadth (IUCN) as traits. To quantify phylogenetic traits, we downloaded pairwise divergence time (phylogenetic distance) from TimeTree.org (41) and reduced the dimensionality of this matrix by mapping species in ordination space using principal coordinate analysis in base R. TimeTree is a knowledge-base that has collected and synthesized species divergence times from > 3000 peer-reviewed studies. We included whether an animal is invasive or endemic in the given *Leishmania* range and zoonotic host status using GISD and GIDEON. Finally, we accounted for sampling effort by downloading the number of citations found on PubMed per species using the R package *easyPubmed* (42). We choose to use PubMed as this indicates the biomedical study effort, as opposed to a more general measure of study effort that would be estimated using the number of

citations on Web of Science. Biomedical study effort better represents the number of times a species was studied in the context of testing for pathogen assemblages.

Data analysis

We applied extreme gradient boosted regression (XGboost) in the R package *xgboost* (43) to use mammal trait data to predict the probability that a species is a *Leishmania* (*Leishmania*) or *Leishmania* (*Viannia*) host. Extreme gradient boosted regression is a machine learning algorithm that creates an ensemble of weak decision trees to form a stronger prediction model by iteratively learning from weak classifiers and combining them into a strong classifier (i.e., boosting) (44). Gradient boosted regression is flexible in that it allows for non-linearity, both among features (i.e., interactions) and between features and predictions, collinearity between features, and non-random patterns of missing data. Further, XGboost handles extremely unbalanced data well by weighting positive labels to increase class separability and allowing for a regularization parameter to prevent overfitting to a few positive labels – an advantage when analyzing our data set with relatively few known hosts compared to the number of mammal species examined. The ability to handle non-random patterns of missing data and extremely unbalanced data makes the XGBoost algorithm best suited for our study as compared to other machine learning methods, such as random forest and support machine vector, which severely overfit unbalanced data (random forest) and/or cannot handle missing data without use of imputation (random forest, support machine vector). Prior to analysis we removed traits with > 0.7 correlation to increase interpretability of variable importance. For the *L. Viannia* analysis 32/66 were retained in the final analysis; for the *L. Leishmania* analysis 33/66 were retained. To maximize the number of traits retained in the model, out of a pair of correlated traits, the one with the highest number of correlations to other traits and/or the lowest data coverage was dropped. We used a box-cox transformation of extremely skewed variables to reduce the influence of extreme values on model performance. Categorical traits were one-hot-encoded so that each category was represented as a binary trait. Finally, traits with < 10% coverage (i.e., less than 10% of the species had data

on the trait) were removed and not used in the analysis (both for training the model and prediction).

Model performance and hyperparameter tuning: The accuracy of our model was evaluated using nested cross-validation. Nested cross-validation, as opposed to K-fold cross-validation, produces the least biased evaluation of model performance when using a small sample size (45). Three-fold cross validation was used to generate model predictions and evaluate overall model performance, while 5-fold cross validation was performed to train hyperparameters within each 3-fold cross validation step. In other words, the full dataset was split into three outer folds. Hyperparameters were tuned using Bayesian optimization, optimizing over a parameter space aimed to reduce overfitting (low training rate, low ratio of samples used in trees, low ratio of features used in trees, low maximum tree depth, and high regularization). Using the hyperparameters that yielded the best performance (minimum logloss), a final model was then trained using the full data within these two folds and model predictions were calculated on the third hold-out fold. This process was repeated until out-of-sample predictions were made for each fold. As model output may be dependent on the distribution of the data across each fold, nested cross validation was repeated 25 times using 25 unique splits of the data. Our dataset was extremely unbalanced (i.e., many more 0s than 1s), which makes our model vulnerable to overfitting. To determine if our model was simply fitting spurious correlations in the data, we conducted a target shuffling analysis (46,47). We repeated the nested cross-validation analysis but randomized the response variables (*Leishmania* host status) for each iteration. We then calculated the mean AUC from the target shuffling experiment. AUC, the area under the receiver operator characteristic curve, is a classification metric that measures the probability that model output for a randomly chosen positive label (known host) will be higher than a randomly chosen negative label (animal with no record of zoonotic *Leishmania* infection). AUC ranges from 0 to 1, with an AUC of 1 indicating that the model perfectly classifies all samples, while an AUC < 0.5 indicates the model performs no better than a coin flip. We found that our model performed minutely better

than a coin flip in the target shuffling experiment (average *L. (Leishmania)* target-shuffled AUC = 0.53; average *L. (Viannia)* target-shuffled AUC = 0.54), thus we adjusted our final AUC values by $AUC_{\text{final model}} - (AUC_{\text{target shuffled model}} - 0.5)$.

In our final model, citation count was among the top 10 most important variables. To ensure that we were not simply predicting well-studied animals, we repeated the analysis above while replacing the response variable with study effort (the number of citations on PubMed). To reduce computational time, we ran ten iterations instead of 25. We concluded that our model was not simply predicting well-studied species by evaluating how well the traits in our model predict study effort, measured as pseudo- R^2 . Additionally, we checked that trait responses (i.e., partial dependence plots) that predicted publication number were not overly similar to the functional form of the trait responses that predicted host status.

Leishmania host trait profiles: To assess *Leishmania* host trait profiles, we first identified important features using SHAP (Shapley Additive Explanations) calculated via the R package *SHAPforxgboost* (48). Shapley scores represent the average marginal contribution of a feature to a prediction across all combinations of features; the contribution of the feature is interpreted as the change in prediction associated with that feature in relation to the average model prediction (44). SHAP calculates local feature contribution by evaluating the contribution of each feature to the prediction for each observation in a dataset. Here, local contribution refers to the contribution of each feature to the prediction for each individual species (local corresponds to the level of a single sample, which, for our analysis, is a species). Global feature contribution can be obtained by aggregating the Shapley scores for each feature across all observations; global feature contribution is the average contribution of each feature across all species (global corresponds to the level of the entire dataset). To generate a measure of uncertainty in feature contribution, we trained the model with 70% of the data, and calculated Shapley scores for the remaining 30% of observations. We repeated the above procedure 100 times, using different subsets

of data with each iteration. We then described trait profiles using SHAP partial dependence plots, which map the relationship between each value of a feature in the validation set and the contribution of the feature to model predictions.

Leishmania host predictions: Following (21), we identified animals as unrecognized zoonotic *Leishmania* spp. hosts using the Shapley value classification criterion. The sum of the Shapley scores across features for each species sums to the prediction for each species, relativized by the average model prediction across species. A summed Shapley value of 0 represents the average prediction in the dataset (49,50). To account for model uncertainty, we used a bootstrapping procedure to calculate Shapley scores for models fit with 70% of the data 100x. We classified an animal as a newly predicted host if > 95% of Shapley scores were > 0; in other words, we classified an animal as a newly predicted host if predicted probability of being a host was greater than the average probability for that model iteration for > 95% of model iterations (21).

The full record of species infection status and references, trait data-base and R code can be found at https://github.com/ckglidden/leish_hosts.

Results

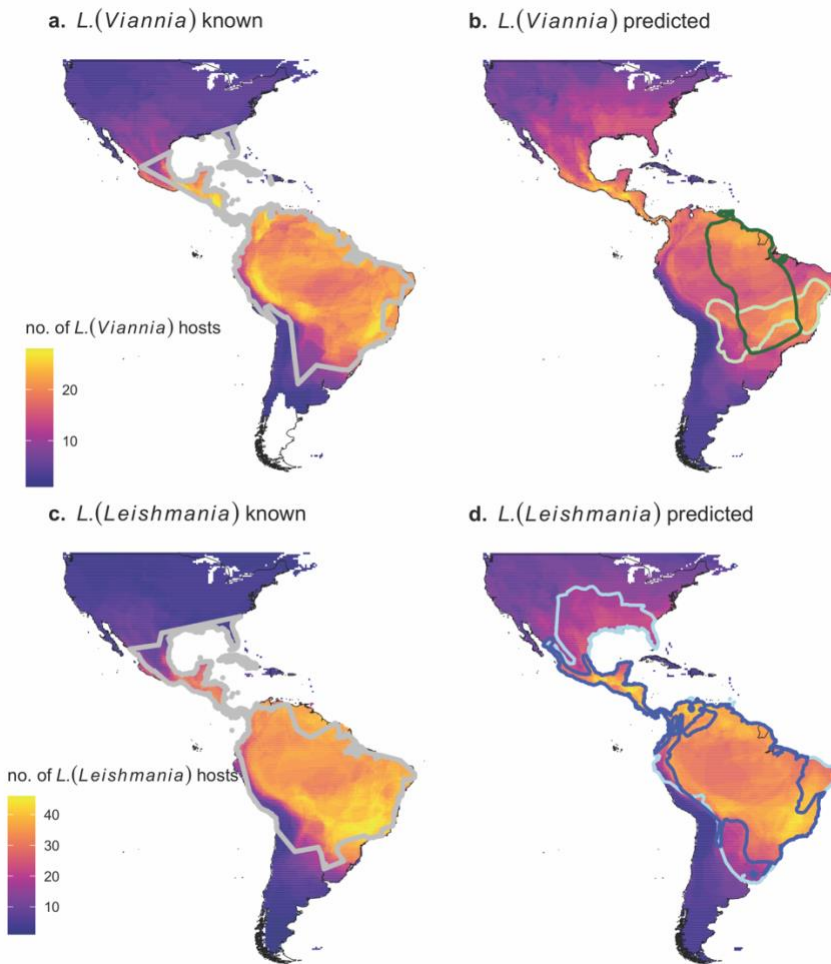


Fig 1. The ranges of zoonotic *Leishmania* and their known and predicted hosts extend throughout the Americas. The number of overlapping ranges of known *Leishmania* hosts are depicted in (a & c), with the range of human cases of *L. (Viannia)* and *L. (Leishmania)* outlined in grey. The number of overlapping ranges of newly predicted hosts are outlined in (b & d). *Hylaeamys megacephalus* and *Calomys callosus* are unrecognized hosts with the highest and second highest predictions for *L. (Viannia)* according to Shapley values. Their ranges are outlined in dark green and light green, respectively, in (b). *Dasypus novemcinctus* and *Leopardus wiedii* are currently hosts with the highest and second highest Shapley values for *L. (Leishmania)*. Their ranges are outlined in dark blue and light blue, respectively, in (d). An exhaustive list of newly predicted hosts is listed in S2 Table. Base maps were created in R using open source shapefiles from Natural Earth (admin-0 countries: <https://www.naturalearthdata.com/downloads/10m-cultural-vectors/>;

<https://www.natureearthdata.com/about/terms-of-use/>) (51); species ranges were created in R using open source shape files from IUCN (<https://www.iucnredlist.org/resources/spatial-data-download>; <https://www.iucnredlist.org/terms/terms-of-use>) (27).

Leishmania (Viannia): Through our literature search, we found 96 known *L. (Viannia)* hosts out of 1460 mammals that occur within the geographic range of *L. (Viannia)* parasites (Fig 1a; Fig 2a). After accounting for target shuffling, our model performed moderately well with an average out of sample AUC of 0.81 (95% CI: 0.80 - 0.82) and in sample AUC of 0.94.

L. (Viannia) trait profiles: After removing highly correlated traits, 30 features were used to predict *L. (Viannia)* hosts. A mix of biogeography, life-history, phylogeny, and study effort covariates were among the top important model features when estimating global feature contribution (Fig 3). Phylogenetic distances, reduced into six axes of a PCoA ordination, were, on average, the most important trait for *Leishmania (Viannia)* hosts. Specifically, animals were more likely to be hosts if they were rodents and opossums, and less likely to be hosts if they were non-human primates, carnivores, and even toed ungulates (S1, S2 Figs). Similar to other zoonotic pathogen systems, *L. (Viannia)* hosts live at high population densities and demonstrate signatures of fast-paced life-histories, indicated by short gestation periods (Fig 4a). Further, *L. (Viannia)* host ranges are large, have high proportions of crop and urban land cover, and, on average, peak at ~28°C during the warmest quarter of the year (Fig 4a). Geographic coordinates of the species' native ranges were also an important predictors: animals were more likely to be hosts if the minimum longitude of their native range was around western Central America. Notably, although study effort (number of PubMed citations, normalized with a Box-Cox transformation) was among the most important traits, our supplementary analyses indicated that the host models were not simply predicting well-studied hosts. Specifically, the supplementary model predicting study effort (number of PubMed citations) performed poorly (average out-of-sample $R^2 = 0.19$). Further, traits that predict citation count, and the response

curves of those traits, greatly differed from traits that predict *L. (Viannia)* host status (S3-S4 Figs). As such, while *L. (Viannia)* hosts are likely understudied, we are not simply reporting the trait profiles of well-studied mammals.

L. (Viannia) hosts: Eighty-three animals were identified as likely hosts (Fig 2, S2 Table). Out of these newly predicted hosts, the majority were rodents, bats, and opossums. Further, 14 of these hosts are known hosts of *L. (Leishmania)* (S2 Table). The newly predicted hosts with the highest summed Shapley scores were the large headed rice rat (*Hylaeamys megacephalus*) and the large vesper mouse, *Calomys callosus* (Fig 1). In contrast to known hosts, the range of newly predicted hosts extends to the southern tip of South America and there is a high density of species ranges that overlap in the southern United States (Fig 1a, 1b).

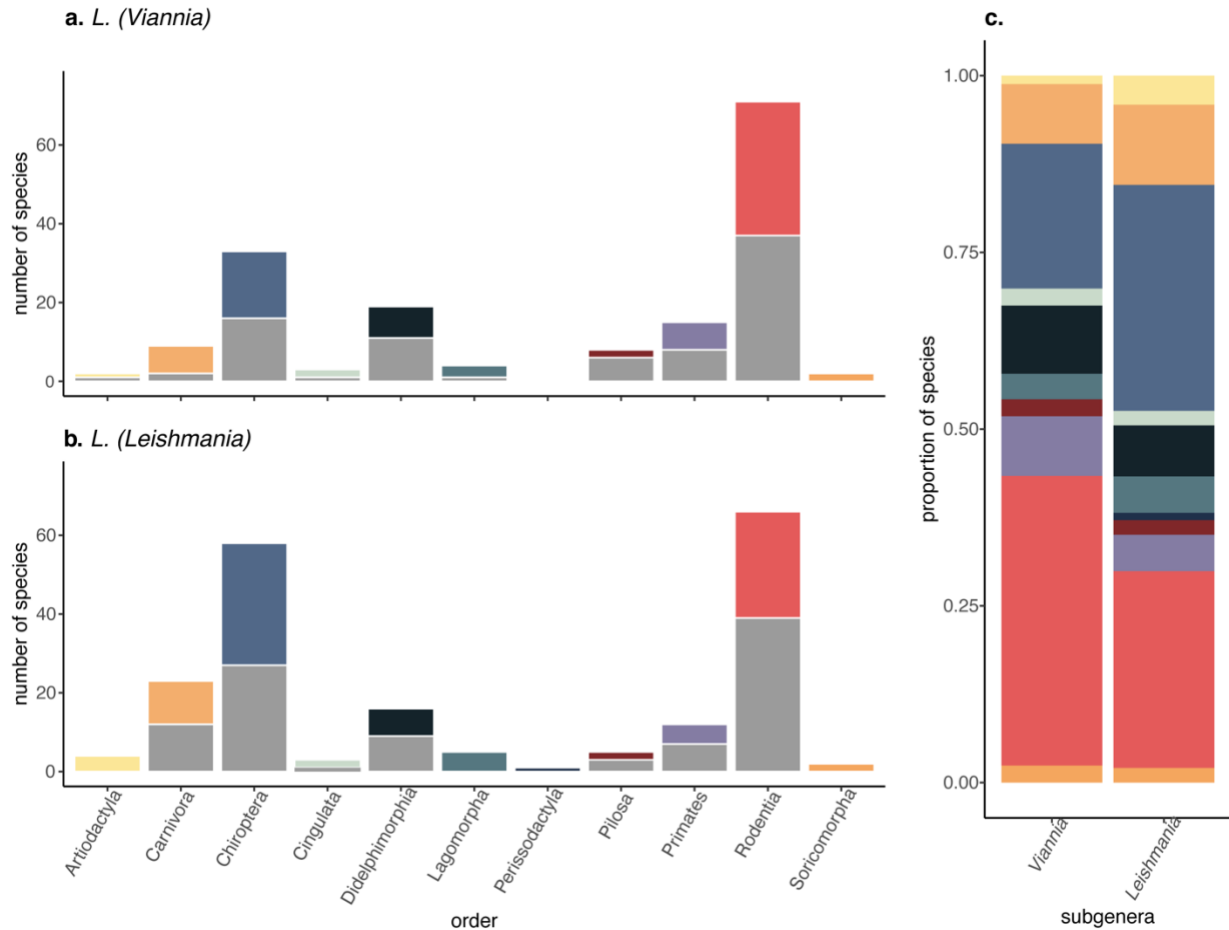


Fig 2. The number of known and predicted hosts per order for *L. (Viannia)* (a) and *L. (Leishmania)* (b). Grey bars represent the number of known hosts within each order, colored bars represent the number of newly predicted hosts per order. Animals were classified as newly predicted if > 95% of mean SHAP values were greater than 0 (a-b). Bars in (c) represent the proportion of newly classified hosts within each order; colors match (a-b). Animal outlines were downloaded from PhyloPic (<http://phylopic.org/>; <http://phylopic.org/about/>) (52).

Leishmania (Leishmania): Through our literature search, we found 110 known hosts out of 1470 mammals that have overlapping ranges with the zoonotic *L. (Leishmania)* range (Fig 1c). Of these, 40 animals were hosts for both *Leishmania* subgenera. After accounting for target shuffling, our *L. (Leishmania)* hosts model performed moderately well with an average out of sample AUC of 0.84 (95% CI: 0.83 - 0.85) and in sample AUC of 0.94.

L. (Leishmania) trait profiles: After removing highly correlated traits, 32 features were used to predict *L. (Leishmania)* hosts. On average, study effort was the most important trait (i.e., had the highest global feature contribution), followed by the land cover/land-use composition of the species range, average temperature in the warmest quarter of the species range, range area, phylogenetic distance, litter size, gestation length, and longitudinal extent of the species range (Fig 3). Similarly, to *L. (Viannia)* hosts, *L. (Leishmania)* hosts show signatures of fast-paced life histories, indicated by short gestation lengths and large litter sizes, and have ranges with high degrees of crop and urban cover (Fig 3b). Further, the range of *L. (Leishmania)* hosts are similar to *L. (Viannia)* hosts in that the mean temperature in the warmest quarter peaks around 28°C (Fig 4b). On average, geographic coordinates of the species' native range were also an important predictor: Animals were more likely to be hosts if the minimum longitude of the range was around western Central America, while the maximum longitude of the range was western Europe, which indicates a role of invasive species in the *L. (Leishmania)* transmission cycle. While study effort was among the most important traits, host traits explained little

variation in number of PubMed citations (average out-of-sample $R^2 = 0.17$) and traits that predict citation count, and the response curves of those traits, greatly differed from traits that predict *L. (Leishmania)* host status suggesting that traits important for discriminating host status were not just those that predicted study effort (S5-S6 Figs).

L. (Leishmania) hosts: Using the Shapley classification criterion, 98 animal species that are currently unrecognized hosts were labeled as newly predicted hosts (Fig 2, S2 Table). Out of these newly predicted hosts, the majority were rodents, bats, and carnivores. Nine newly predicted hosts are known hosts of *L. (Viannia)*. Forty-five newly predicted hosts were newly predicted hosts for both the *L. (Leishmania)* and *L. (Viannia)* subgeneras (S2 Table). The newly predicted hosts for the *L. (Leishmania)* with the top summed Shapley scores included the nine-banded armadillo (*Dasypus novemcinctus*) and margay (*Leopardus wiedii*) (Fig 1). The nine-banded armadillo is a known host of *L. (Viannia)* parasites, with a range that extends from northern Argentina to the central United States (Fig 1d).

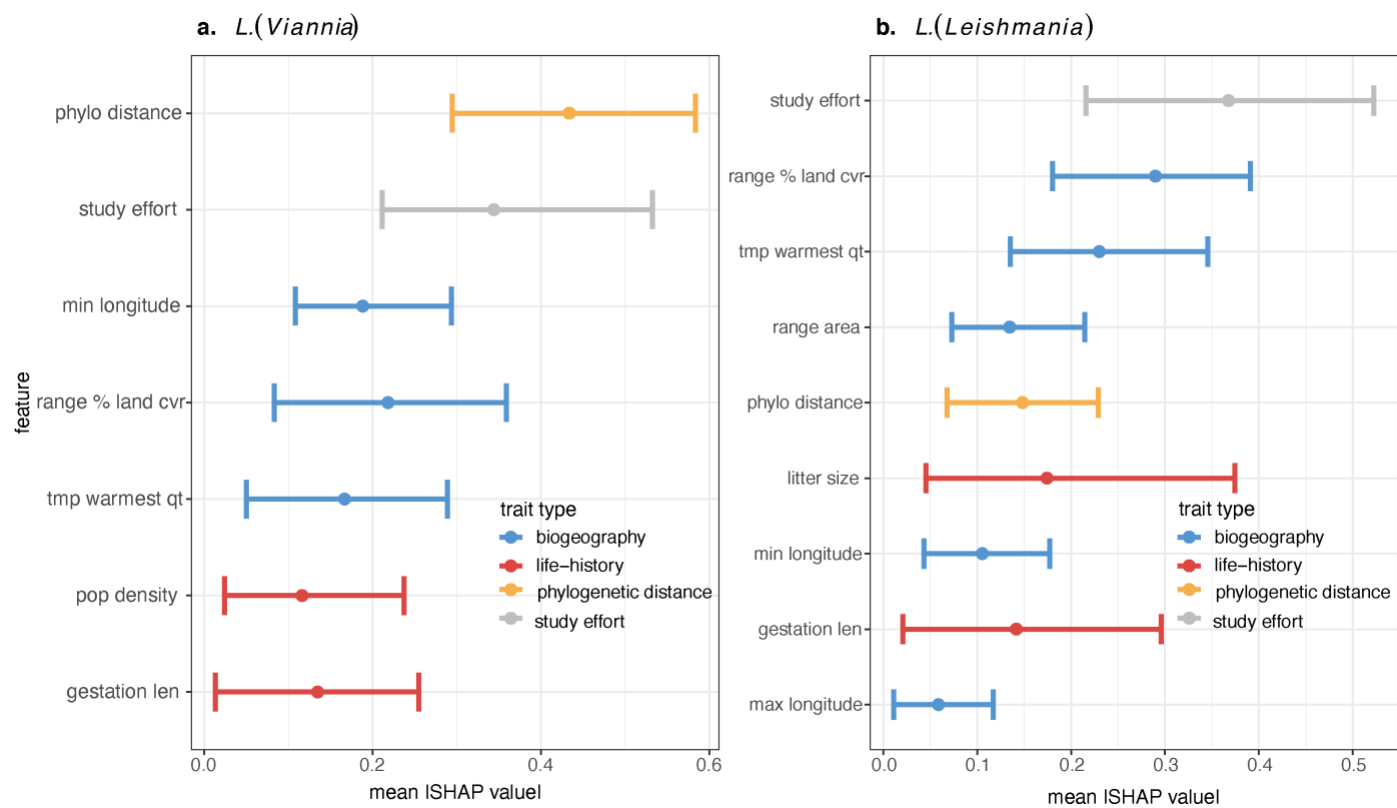
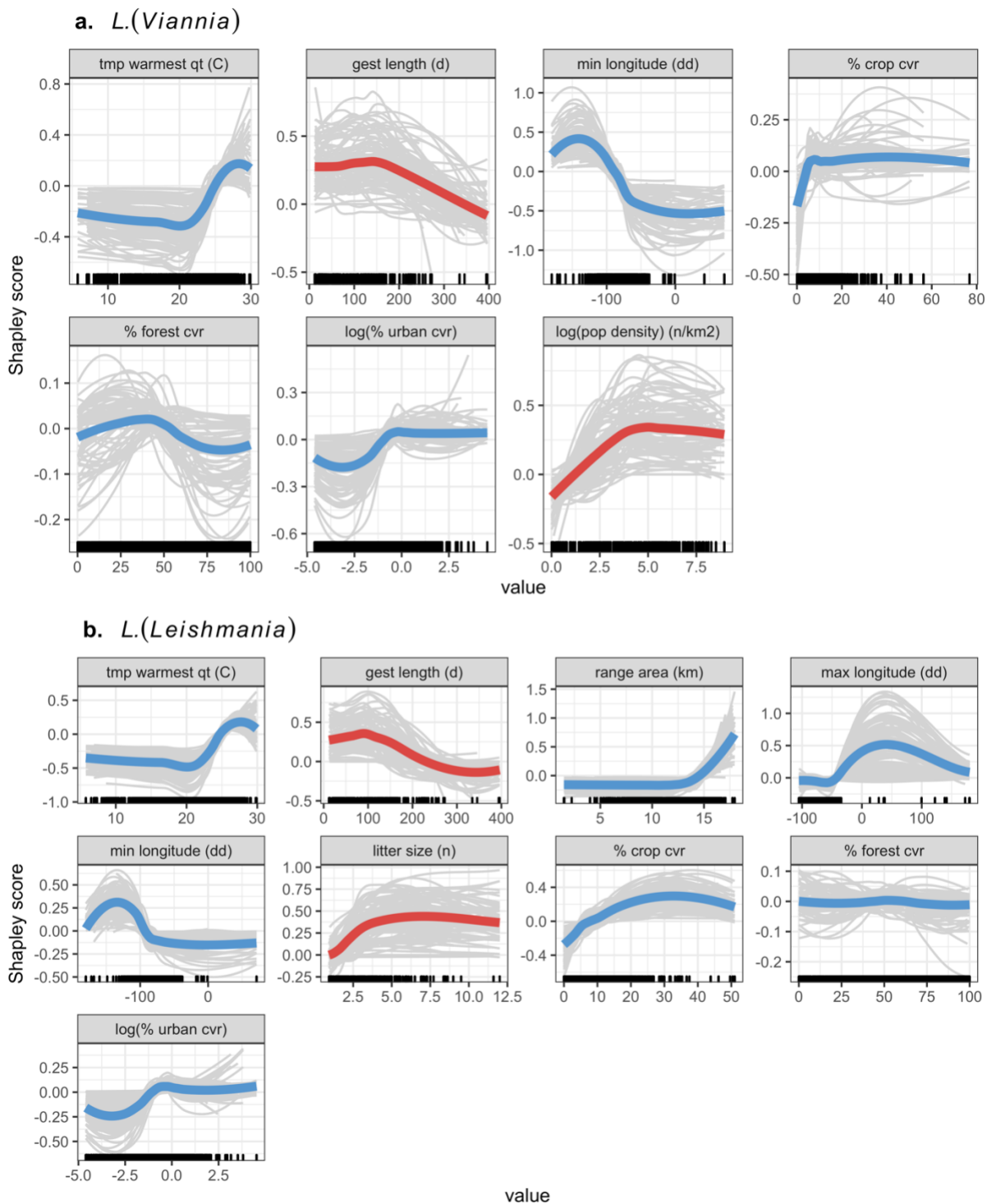


Fig 3. Biogeography, life-history, and phylogenetic traits all significantly contributed to model predictions for host status. Biogeographical traits (minimum longitude of the range, maximum longitude of the range, % cover of land-use/land cover in the species range, average temperature in the warmest quarter in the species range, range area) are colored in blue, life-history traits (population density, gestation length, litter size) are colored in red, phylogenetic distance (location along PCoA ordination axes) is colored in orange, and study effort is colored in grey. Points are the absolute value of the mean Shapley importance (mean |SHAP value|) for the trait across all mammals (i.e., global feature contribution), bars represent the absolute values of the 0.05-0.95 percentiles. Only features with 0.05 percentiles > 0 are shown.



468 **Fig 4. Hosts have fast-paced life-histories and live in proximity to humans.** Shapley
 469 partial dependence plots showing the effect of each feature on *L. (Viannia)* (a) and *L.*
 470 (*Leishmania*) (b) host status after accounting for the average effect of the other features
 471 in the model. Colored lines represent the average effect across model iterations, while
 472 grey lines show each individual model iteration (model fit with 80% of data) (blue =
 473 biogeographical traits; red = life-history). Features with global mean feature contribution

scores > 0 for > 95% of model iterations are shown. Rug plots on the x-axis indicate the distribution of the data.

Discussion

Our literature search identified 137 wild mammals in the Americas as known zoonotic *Leishmania* hosts, and our model predicted there to be an additional 136 likely hosts (animals that are likely to be exposed and infected in the wild, but not necessarily infectious upon infection). The majority of the predicted hosts for both subgenera fell within the Rodentia order (*L. (Leishmania)* N = 27; *L. (Viannia)* N: = 34), with a similar number of predicted *Leishmania (Leishmania)* hosts within the Chiroptera order (N = 32). While their contribution to model predictions varied between subgenera, key traits included phylogenetic distance, human associated land-use, climate, population density, and study effort.

For both *Leishmania* subgenera, biogeographical features encompassing descriptors of species ranges were among the top five contributors to host predictive accuracy. Notably, zoonotic *L. (Leishmania)* hosts are more likely to be animals with larger ranges than non-hosts. Geographic range area was also reported to be a shared predictive feature among zoonotic hosts and reservoirs in other systems (21,53). Further, ranges of *L. (Leishmania)* and *L. (Viannia)* hosts are made up of higher proportions of crop cover and urban cover than ranges of mammals that are not *Leishmania* hosts. These results are in accordance with previous findings that, in general, zoonotic host richness and relative abundance increases with anthropogenic pressure (54). Further, our results align with recent literature showing that risk of leishmaniasis is associated with agricultural intensification and urbanization (4,55). *L. (Viannia)* hosts are marginally more likely to have ranges with moderate forest cover, than non-hosts whereas there is little effect of forest cover on *L. (Leishmania)* host status. Vectors of cutaneous leishmaniasis, which is caused by all species of zoonotic *Leishmania* except *L. infantum*, are more likely to live in forests with

high integrity and low human modification (56), while hosts are more common in disturbed landscapes (57). Our analysis adds support to the latter finding, highlighting that transmission of parasites associated with cutaneous leishmaniasis likely occurs at forest-human settlement interfaces. The incomplete overlap between vector ranges and host ranges may explain the highly localized dynamics of cutaneous leishmaniasis spillover (i.e., transmission from wildlife to humans via the sandfly vector) (6,58). Spillover occurs when reservoir hosts, vectors, and humans overlap in space and time so that forward transmission occurs from animal to sandfly to human. If vector and host distributions only overlap at habitat interfaces, area of reservoir host, sandfly, and human co-occurrence may be low and variable through time. Finally, average temperature of the warmest quarter was an important predictor of host status, with the feature contribution to model predictions peaking around 28°C. In the case of zoonotic leishmaniasis, temperature is likely a predictor of *Leishmania* exposure as this trait profile reflects temperatures associated with vector survival and parasite development (59,60).

The profiles of important life history traits for *Leishmania* hosts resembled those observed in other comparative analyses of zoonotic diseases. Specifically, for both *Leishmania* subgenera, the probability of host status peaked at low to median lengths of gestation, a pattern similar to that of rodent reservoirs of zoonotic pathogens and competent hosts of zoonotic tick-borne pathogens (53,61). Further, the probability of zoonotic *L. (Leishmania)* host status increased with litter size, a trend that parallels that of both rodents reservoirs of zoonotic pathogens and bat and carnivore reservoirs of rabies virus (21,53). Altogether, our findings suggest that, like other zoonotic reservoirs and hosts, zoonotic *Leishmania* hosts may have fast-paced life-histories. Animals with fast-paced life-histories are believed to maximize fitness by expeditiously investing in short-term reproductive effort -- such as through younger ages at sexual maturity, short gestation lengths, greater litter sizes -- at the cost of long-term survival (62). As a result, animals with fast-paced life histories are thought to divert energetic resources away from effective adaptive immunity (63), or utilize responses conducive to tolerance but not resistance (64), allowing them to

allocate maximal resources to short term reproductive effort. As such, animals with fast-paced life-histories may be more susceptible to infections.

However, *L. (Viannia)* hosts only had one trait (gestation length) clearly associated with a fast-paced life history, suggesting weaker evidence for the “fast-paced” host hypothesis for this subgenus. Phylogenetic distance was the most important trait for *L. (Viannia)* hosts. We hypothesize that phylogenetic relationships may better account for host physiological defenses against *Leishmania* than life-history traits. If this is the case, inclusion of specific genetic markers associated with *Leishmania*-host cell interactions may improve model predictions.

Notably, study effort significantly contributed to host predictions for both subgenera. While our follow up analyses do not indicate that model results are biased by uneven study effort, this finding may suggest that zoonotic *Leishmania* hosts are generally under sampled. Though leishmaniasis is a neglected tropical disease primarily impacting resource-poor communities, state-of-the-art zoonotic *Leishmania* diagnostic and surveillance techniques are available (16,57,65–69). As such, under sampling *Leishmania* hosts primarily stems from both the lack of funding to employ these tools during large scale surveillance efforts and the large number of animal species that are potentially hosts. The main leishmaniasis research funding mechanism within the USA, the National Institute of Health (NIH), primarily focuses on the molecular biology of leishmaniasis work and does not prioritize research elucidating the ecology of the sandfly-wildlife transmission cycle. In fact, a simple search for “leishmania” on the NIH and National Science Foundation grants database yields a combined total of three awards aiming to elucidate transmission cycles of *Leishmania* in the Americas (searched October 10, 2022). Researchers across Latin America are faced with similar, if not more severe, funding constraints. For example, in Brazil, a major hotspot of leishmaniasis, the federal administration from 2019-202 cut 90% of federal science funding (70). While our model results facilitate more targeted field sampling for particular species, optimized surveillance should consist of an iterative

feedback between local knowledge and model predictions to guide field work, and updating the model with new results from field sampling to guide ongoing surveillance and field studies (16,19). Importantly, to reduce study effort bias going forward, care should be taken to estimate and survey under studied areas and species in addition to newly predicted species, particularly in areas with high rates of human spillover with poorly known hosts. A lack of investment in understanding the basic ecology of *Leishmania* slows research progress by obscuring the major drivers of disease dynamics and spillover transmission to humans. Leishmaniasis surveillance and management would benefit from larger funding mechanisms prioritizing understanding the ecology of this group of pathogens.

In some cases, our host predictions are supported by multiple independent but complementary lines of evidence (summarized below and in S3 Table). Non-specific *Leishmania* antigens, antibodies, or DNA has been detected in the crab-eating racoon (*Procyon cancrivorus*), neotropical river otter (*Lontra longicaudis*), six-banded armadillo (*Euphractus sexcinctus*), coyote (*Canis latrans*), and the grey short-tailed opossum (*Monodelphis domestica*) (12,68,71,72). As tests were not parasite species specific, it is unconfirmed if they were infected with zoonotic *Leishmania* species, but our model suggests they may be. Further, *Leishmania*-like flagellates were discovered in Derby's wooly opossum (*Caluromys derbianus*) (71); however, to our knowledge this species has not been tested for zoonotic *Leishmania* species.

A number of animals predicted to be hosts for *L. (Leishmania)* were also found to co-occur with vectors of *L. (L.) mexicana* in Mexico: the hairy big-eyed bat (*Chiroderma villosum*), Brazilian brown bat (*Eptesicus brasiliensis*), great sac-winged bat (*Saccopteryx bilineata*), tent-making bat (*Uroderma bilobatum*), Mexican woodrat (*Neotoma mexicana*), black footed pygmy rice rat (*Oligoryzomys nigripes*), Coue's rice rat (*Oryzomys couesi*), fulvous harvest mouse (*Reithrodontomys fulvescens*), nine-banded armadillo (*Dasypus novemcinctus*), lowland paca (*Cuniculus paca*), Virginia opossum (*Didelphis virginiana*),

kinkajou (*Potos flavus*), racoon (*Procyon lotor*), white-nose coati (*Nasua narica*), and the margay (*Leopardus wiedii*) (16,73). Notably, the nine-banded armadillo and margay had the two highest summed Shapley scores as potential *L. (Leishmania)* hosts. The nine-banded armadillo is a known host of *L. (Viannia)* as well as other zoonotic pathogens including the causative agent of leprosy (74).

Host competence (i.e., the ability to become infected and infectious) has been demonstrated in the lab for four of the predicted species: the Virginia opossum (*Didlephis virginiana*), the large vesper mouse (*Calomys callosus*), spix's yellow toothed cavy (*Galea spixii*), and the Mexican free-tailed bat (*Tadarida brasiliensis*) (71,75–77). Although host competence for these animals has been tested in the laboratory, infection/exposure has not been tested in the wild or animals have not tested positive in the wild (possibly due to minimal sampling of the species). Our model suggests that these animals are also likely to be naturally infected in the wild and thus warrant further investigation. The large vesper mouse is one of the top predicted species for *L. (Viannia)*. Surprisingly, we found no field studies testing wild large vesper mice for zoonotic *Leishmania* infection. Finally, our model suggests that a handful of species naturally infected with *L. infantum* in Europe are also likely to be exposed to and susceptible to zoonotic *Leishmania* in the Americas: the red fox (*Vulpes vulpes*), the European hare (*Lepus europaeus*), and the European rabbit (*Oryctolagus cuniculus*) (78–80).

Importantly, while there are 14 species of zoonotic *Leishmania* parasites in the Americas, our analysis divides these species into two subgenera but does not identify *Leishmania* species specific hosts. Future work could leverage the information presented by the current analysis to identify *Leishmania* species specific hosts. Additionally, our analysis identifies animals that are likely to be infected in the wild but does not narrow the host pool down to competent hosts (animals that are likely to be both infected *and* infectious in the wild). Next steps of model development could identify trait profiles of competent hosts that delineate them from non-competent hosts. Notably, (56) used machine learning to predict

unrecognized vectors of leishmaniasis. Further, metabarcoding of arthropods has started to identify sandfly feeding networks (66,81) that may be useful in linking model predictions of vectors and hosts to local transmission cycles.

Overall, models for both subgenera identified a number of new hosts, including but not limited to the nine-banded armadillo, the margay, the large vesper mouse, and the large headed rice rat, that can be targeted for future surveillance and research to determine their ecological significance in leishmaniasis transmission and spillover to humans. Next steps should include ground truthing the role of newly identified potential hosts in the zoonotic *Leishmania* transmission cycle, through field sampling and laboratory experiments, through engagement with local communities, and iteratively updating the current analysis as new information is collected, taking careful consideration to not perpetuate the effects of study bias. Although a cautious first step in host discovery work, our study can immediately inform surveillance efforts and study design. For instance, host predictions may provide a starting point for implication of hosts during local outbreaks of human leishmaniasis by providing a targeted list of likely suspects to sample. While the study of leishmaniasis has historically focused on rodents and small marsupials, our study suggests that capture methods should more commonly be supplemented by bat mist-netting and/or techniques to capture small to medium sized carnivores. Finally, many newly predicted hosts have ranges that extend throughout the southeastern USA. Ecological niche models suggest that climate warming may increase the range of *Leishmania* vectors as far north as Canada, and leishmaniasis endemicity within the US was recently confirmed (82–84). Similarly, the range of predicted hosts for *L. (Viannia)* extends farther south through Argentina than currently known hosts. Our analysis provides tools for monitoring leishmaniasis risk under global change as these hosts could be used as sentinel species for tracking leishmaniasis range expansion. Testing these hosts for leishmaniasis infection beyond the current putative boundaries of zoonotic *Leishmania* ranges would help to evaluate if range expansion has already occurred as well as provide an updated baseline for future monitoring.

648

649 Identifying local transmission cycles is key for reducing leishmaniasis spillover and
650 disease burden in people, especially in the face of land-use and climate change, which
651 likely influence the geographic distribution and extent of vector—host—human contact and
652 were strongly associated with novel host predictions in our models. We estimate that there
653 are likely many unknown hosts of leishmaniasis throughout the Americas, highlighting the
654 complexity of the transmission cycle and a need to increase study effort on the ecology
655 and epidemiology of this system.

656

657 **Acknowledgements**

658 We thank the Mordecai Lab for providing feedback on manuscript drafts.

659

660

661 **References**

- 662 1. WHO. The road map targets for 2030 [Internet]. The road map targets for 2030. 2022 [cited
663 2022 Jun 12]. Available from: [https://www.who.int/teams/control-of-neglected-tropical-](https://www.who.int/teams/control-of-neglected-tropical-diseases/ending-ntds-together-towards-2030/targets)
664 [diseases/ending-ntds-together-towards-2030/targets](https://www.who.int/teams/control-of-neglected-tropical-diseases/ending-ntds-together-towards-2030/targets)
- 665 2. PAHO. Leishmaniasis - PAHO/WHO | Pan American Health Organization [Internet]. 2022
666 [cited 2022 Jun 12]. Available from: <https://www.paho.org/en/topics/leishmaniasis>
- 667 3. Bayles BR, Rusk AE, Pineda MA, Chen B, Dagy K, Hummel T, et al. Spatiotemporal trends
668 of cutaneous leishmaniasis in Costa Rica. *Trans R Soc Trop Med Hyg*. 2021 May
669 8;115(5):569–71.
- 670 4. Bayles BR, Rusk A, Christofferson R, Agar G, Pineda MA, Chen B, et al. Spatiotemporal
671 dynamics of vector-borne disease risk across human land-use gradients: examining the role
672 of agriculture, indigenous territories, and protected areas in Costa Rica. *Lancet Glob Health*.
673 2020 Apr 1;8:S32.
- 674 5. Dutari LC, Loaiza JR. American Cutaneous Leishmaniasis in Panama: a historical review of
675 entomological studies on anthropophilic *Lutzomyia* sand fly species. *Parasit Vectors*. 2014
676 May 11;7(1):218.
- 677 6. Gonçalves NV, Miranda C do SC, Costa RJF da, Guedes JA, Matsumura ES de S, Costa
678 SBN da, et al. Cutaneous leishmaniasis: Spatial distribution and environmental risk factors
679 in the state of Pará, Brazilian Eastern Amazon. *J Infect Dev Ctries*. 2019 Oct
680 31;13(10):939–44.
- 681 7. Portella TP, Kraenkel RA. Spatial-temporal pattern of cutaneous leishmaniasis in Brazil.
682 *Infect Dis Poverty*. 2021 Jun;10(03):47–57.
- 683 8. Chakravarty J, Sundar S. Current and emerging medications for the treatment of
684 leishmaniasis. *Expert Opin Pharmacother*. 2019 Jul 3;20(10):1251–65.

- 685 9. Glidden CK, Nova N, Kain MP, Lagerstrom KM, Skinner EB, Mandle L, et al. Human-
686 mediated impacts on biodiversity and the consequences for zoonotic disease spillover. *Curr*
687 *Biol*. 2021 Oct 11;31(19):R1342–61.
- 688 10. Sokolow SH, Nova N, Pepin KM, Peel AJ, Pulliam JRC, Manlove K, et al. Ecological
689 interventions to prevent and manage zoonotic pathogen spillover. *Philos Trans R Soc Lond*
690 *B Biol Sci*. 2019 Sep 30;374(1782):20180342.
- 691 11. Lainson R, Shaw JJ. Epidemiology and ecology of leishmaniasis in Latin-America. *Nature*.
692 1978 Jun;273(5664):595–600.
- 693 12. Azami-Conesa I, Gómez-Muñoz MT, Martínez-Díaz RA. A Systematic Review (1990–2021)
694 of Wild Animals Infected with Zoonotic *Leishmania*. *Microorganisms*. 2021 May
695 20;9(5):1101.
- 696 13. Roque ALR, Jansen AM. Wild and synanthropic reservoirs of *Leishmania* species in the
697 Americas. *Int J Parasitol Parasites Wildl*. 2014 Aug 29;3(3):251–62.
- 698 14. López M, Erazo D, Hoyos J, León C, Fuya P, Lugo L, et al. Measuring spatial co-
699 occurrences of species potentially involved in *Leishmania* transmission cycles through a
700 predictive and fieldwork approach. *Sci Rep*. 2021 Mar 24;11(1):6789.
- 701 15. Medkour H, Davoust B, Levasseur A, Mediannikov O. Molecular Evidence of *Leishmania*
702 *infantum* and *Leishmania guyanensis* in Red Howler Monkey (*Alouatta seniculus*) from
703 French Guiana. *Vector-Borne Zoonotic Dis*. 2019 Dec;19(12):896–900.
- 704 16. Stephens CR, González-Salazar C, Sánchez-Cordero V, Becker I, Rebollar-Tellez E,
705 Rodríguez-Moreno Á, et al. Can You Judge a Disease Host by the Company It Keeps?
706 Predicting Disease Hosts and Their Relative Importance: A Case Study for Leishmaniasis.
707 *PLoS Negl Trop Dis*. 2016 Oct 7;10(10):e0005004.
- 708 17. Vieira TM, Silva S de O, Lima L, Sabino-Santos G, Duarte ER, Lima SM, et al. *Leishmania*
709 diversity in bats from an endemic area for visceral and cutaneous leishmaniasis in
710 Southeastern Brazil. *Acta Trop*. 2022 Apr 1;228:106327.
- 711 18. Carlson CJ, Albery GF, Merow C, Trisos CH, Zipfel CM, Eskew EA, et al. Climate change
712 increases cross-species viral transmission risk. *Nature*. 2022 Jul;607(7919):555–62.
- 713 19. Becker DJ, Albery GF, Sjödin AR, Poisot T, Bergner LM, Chen B, et al. Optimising
714 predictive models to prioritise viral discovery in zoonotic reservoirs. *Lancet Microbe*
715 [Internet]. 2022 Jan 10 [cited 2022 Jun 12]; Available from:
716 <https://www.sciencedirect.com/science/article/pii/S2666524721002457>
- 717 20. Han BA, Majumdar S, Calmon FP, Glicksberg BS, Horesh R, Kumar A, et al. Confronting
718 data sparsity to identify potential sources of Zika virus spillover infection among primates.
719 *Epidemics*. 2019 Jun 1;27:59–65.
- 720 21. Worsley-Tonks KEL, Escobar LE, Biek R, Castaneda-Guzman M, Craft ME, Streicker DG,
721 et al. Using host traits to predict reservoir host species of rabies virus. *PLoS Negl Trop Dis*.
722 2020 Dec 8;14(12):e0008940.
- 723 22. Becker DJ, Han BA. The macroecology and evolution of avian competence for *Borrelia*
724 *burgdorferi*. *Glob Ecol Biogeogr*. 2021;30(3):710–24.
- 725 23. Combes C. The Art of Being a Parasite [Internet]. Chicago, IL: University of Chicago Press;
726 2005 [cited 2023 Jan 5]. 280 p. Available from:
727 <https://press.uchicago.edu/ucp/books/book/chicago/A/bo3534652.html>

- 728 24. Sweeny AR, Albery GF. Exposure and susceptibility: The Twin Pillars of infection. *Funct*
729 *Ecol.* 2022 May 18;1365-2435.14065.
- 730 25. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, et al. Pathways to
731 zoonotic spillover. *Nat Rev Microbiol.* 2017 Aug;15(8):502–10.
- 732 26. Akhoundi M, Kuhls K, Cannet A, Votýpka J, Marty P, Delaunay P, et al. A Historical
733 Overview of the Classification, Evolution, and Dispersion of *Leishmania* Parasites and
734 Sandflies. *PLoS Negl Trop Dis.* 2016 Mar 3;10(3):e0004349.
- 735 27. The IUCN Red List of Threatened Species [Internet]. IUCN Red List of Threatened Species.
736 [cited 2022 Aug 31]. Available from: <https://www.iucnredlist.org/en>
- 737 28. GISD [Internet]. [cited 2023 Jan 5]. Available from: <http://www.iucngisd.org/gisd/>
- 738 29. Global Infectious Diseases and Epidemiology Network [Internet]. GIDEON. [cited 2023 Jan
739 5]. Available from: <https://www.gideononline.com/>
- 740 30. Herrera G, Barragán N, Luna N, Martínez D, De Martino F, Medina J, et al. An interactive
741 database of *Leishmania* species distribution in the Americas. *Sci Data.* 2020 Apr 3;7(1):110.
- 742 31. Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, et al. PanTHERIA: a species-
743 level database of life history, ecology, and geography of extant and recently extinct
744 mammals. *Ecology.* 2009;90(9):2648–2648.
- 745 32. Kissling WD, Dalby L, Fløjgaard C, Lenoir J, Sandel B, Sandom C, et al. Establishing
746 macroecological trait datasets: digitalization, extrapolation, and validation of diet
747 preferences in terrestrial mammals worldwide. *Ecol Evol.* 2014;4(14):2913–30.
- 748 33. Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. EltonTraits 1.0:
749 Species-level foraging attributes of the world's birds and mammals. *Ecology.*
750 2014;95(7):2027–2027.
- 751 34. Brun P, Zimmermann NE, Hari C, Pellissier L, Karger DN. Global climate-related predictors
752 at kilometre resolution for the past and future. *Earth Syst Sci Data Discuss.* 2022 Jun 27;1–
753 44.
- 754 35. Karger DN, Schmatz DR, Dettling G, Zimmermann NE. High-resolution monthly precipitation
755 and temperature time series from 2006 to 2100. *Sci Data.* 2020 Jul 23;7(1):248.
- 756 36. Buchhorn M, Lesiv M, Tsendbazar NE, Herold M, Bertels L, Smets B. Copernicus Global
757 Land Cover Layers—Collection 2. *Remote Sens.* 2020 Jan;12(6):1044.
- 758 37. Kennedy CM, Oakleaf JR, Theobald DM, Baruch-Mordo S, Kiesecker J. Managing the
759 middle: A shift in conservation priorities based on the global human modification gradient.
760 *Glob Change Biol.* 2019;25(3):811–26.
- 761 38. Rosa CA da, Ribeiro BR, Bejarano V, Puertas FH, Bocchiglieri A, Barbosa AL dos S, et al.
762 NEOTROPICAL ALIEN MAMMALS: a data set of occurrence and abundance of alien
763 mammals in the Neotropics. *Ecology.* 2020;101(11):e03115.
- 764 39. Töpel M, Zizka A, Calió MF, Scharn R, Silvestro D, Antonelli A. SpeciesGeoCoder: Fast
765 Categorization of Species Occurrences for Analyses of Biodiversity, Biogeography, Ecology,
766 and Evolution. *Syst Biol.* 2017 Mar 1;66(2):145–51.
- 767 40. Gombin J, Vaidyanathan R, Agafonkin V. concaveman. R. 2020;

- 768 41. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines,
769 Timetrees, and Divergence Times. *Mol Biol Evol.* 2017 Jul 1;34(7):1812–9.
- 770 42. Fantini D. easyPubMed – Data Pulse [Internet]. 2019 [cited 2022 Aug 31]. Available from:
771 https://www.data-pulse.com/dev_site/easypubmed/
- 772 43. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the
773 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
774 [Internet]. San Francisco California USA: ACM; 2016 [cited 2022 Aug 31]. p. 785–94.
775 Available from: <https://dl.acm.org/doi/10.1145/2939672.2939785>
- 776 44. Molnar C. Interpretable Machine Learning [Internet]. 2nd ed. 2022 [cited 2022 Aug 31].
777 Available from: <https://christophm.github.io/interpretable-ml-book/>
- 778 45. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a
779 limited sample size. *PLOS ONE.* 2019 Nov 7;14(11):e0224365.
- 780 46. Fischhoff IR, Castellanos AA, Rodrigues JPGLM, Varsani A, Han BA. Predicting the
781 zoonotic capacity of mammals to transmit SARS-CoV-2. *Proc Biol Sci.* 2021 Nov
782 24;288(1963):20211651.
- 783 47. Martin JT, Fischhoff IR, Castellanos AA, Han BA. Ecological Predictors of Zoonotic Vector
784 Status Among *Dermacentor* Ticks (Acari: Ixodidae): A Trait-Based Approach. Gaff H, editor.
785 *J Med Entomol.* 2022 Sep 5;tjac125.
- 786 48. Liu Y, Just A. SHAPforxgboost: SHAP Plots for “XGBoost”. R Package Version 010. 2020;
- 787 49. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions [Internet]. arXiv;
788 2017 [cited 2022 Sep 21]. Available from: <http://arxiv.org/abs/1705.07874>
- 789 50. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. Explainable AI for
790 Trees: From Local Explanations to Global Understanding [Internet]. arXiv; 2019 [cited 2022
791 Sep 21]. Available from: <http://arxiv.org/abs/1905.04610>
- 792 51. Natural Earth - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales
793 [Internet]. [cited 2023 Jan 30]. Available from: <https://www.naturalearthdata.com/>
- 794 52. PhyloPic — Free Silhouette Images of Life Forms [Internet]. [cited 2023 Jan 30]. Available
795 from: <http://phylopic.org/>
- 796 53. Han BA, Schmidt JP, Bowden SE, Drake JM. Rodent reservoirs of future zoonotic diseases.
797 *Proc Natl Acad Sci.* 2015 Jun 2;112(22):7039–44.
- 798 54. Gibb R, Redding DW, Chin KQ, Donnelly CA, Blackburn TM, Newbold T, et al. Zoonotic
799 host diversity increases in human-dominated ecosystems. *Nature.* 2020
800 Aug;584(7821):398–402.
- 801 55. Lima ÁLM, de Lima ID, Coutinho JFV, de Sousa ÚPST, Rodrigues MAG, Wilson ME, et al.
802 Changing epidemiology of visceral leishmaniasis in northeastern Brazil: a 25-year follow-up
803 of an urban outbreak. *Trans R Soc Trop Med Hyg.* 2017 Oct 1;111(10):440–7.
- 804 56. Vadmal GM, Glidden CK, Han BA, Carvalho B, Castellanos AA, Mordecai EA. Data-Driven
805 Predictions of Potential *Leishmania* Vectors in Latin America [Internet]. bioRxiv; 2022 [cited
806 2022 Oct 10]. p. 2022.08.22.504897. Available from:
807 <https://www.biorxiv.org/content/10.1101/2022.08.22.504897v1>

- 808 57. Kocher A, Cornuault J, Gantier JC, Manzi S, Chavy A, Girod R, et al. Biodiversity and
809 vector-borne diseases: Host dilution and vector amplification occur simultaneously for
810 Amazonian leishmaniasis. *Mol Ecol*. 2022 Jan 9;
- 811 58. Silva J, Queiroz A, Moura I, Sousa RS, Guimarães LH, Machado PRL, et al. Dynamics of
812 American tegumentary leishmaniasis in a highly endemic region for *Leishmania (Viannia)*
813 *braziliensis* infection in northeast Brazil. *PLoS Negl Trop Dis*. 2017 Nov 2;11(11):e0006015.
- 814 59. Guzmán H, Tesh RB. Efectos de la temperatura y la dieta sobre el crecimiento y la
815 longevidad de flebotómíneos (Diptera: Psychodidae). *Biomédica*. 2000 Sep 1;20(3):190.
- 816 60. Hlavacova J, Votypka J, Volf P. The effect of temperature on *Leishmania* (Kinetoplastida:
817 Trypanosomatidae) development in sand flies. *J Med Entomol*. 2013 Sep;50(5):955–8.
- 818 61. Ostfeld RS, Levi T, Jolles AE, Martin LB, Hosseini PR, Keesing F. Life history and
819 demographic drivers of reservoir competence for three tick-borne zoonotic pathogens. *PloS*
820 *One*. 2014;9(9):e107387.
- 821 62. Demas N, Nelson R. *Ecoimmunology*. Oxford, New York: Oxford University Press; 2011.
822 656 p.
- 823 63. Martin LB, Weil ZM, Nelson RJ. Immune defense and reproductive pace of life in
824 *Peromyscus* mice. *Ecology*. 2007 Oct;88(10):2516–28.
- 825 64. Schwanz LE. Persistent Effects of Maternal Parasitic Infection on Offspring Fitness:
826 Implications for Adaptive Reproductive Strategies When Parasitized. *Funct Ecol*.
827 2008;22(4):691–8.
- 828 65. Caldart ET, Pinto-Ferreira F, Matos AMRN de, Pascoal ATP, Bertão-Santos A, Mitsuka-
829 Breganó R, et al. Evaluation of an active and early surveillance methodology for visceral
830 leishmaniasis by molecular detection in road-killed wild fauna. *Rev Bras Parasitol*
831 *Veterinária*. 2021;30(2):e027920.
- 832 66. Massey AL, Bronzoni RV de M, da Silva DJF, Allen JM, de Lázari PR, dos Santos-Filho M,
833 et al. Invertebrates for vertebrate biodiversity monitoring: Comparisons using three insect
834 taxa as iDNA samplers. *Mol Ecol Resour*. 2022;22(3):962–77.
- 835 67. Patiño LH, Castillo-Castañeda AC, Muñoz M, Jaimes JE, Luna-Niño N, Hernández C, et al.
836 Development of an Amplicon-Based Next-Generation Sequencing Protocol to Identify
837 *Leishmania* Species and Other Trypanosomatids in Leishmaniasis Endemic Areas.
838 *Microbiol Spectr*. 2021 Oct 31;9(2):e0065221.
- 839 68. Richini-Pereira VB, Marson PM, Hayasaka EY, Victoria C, da Silva RC, Langoni H.
840 Molecular detection of *Leishmania spp.* in road-killed wild mammals in the Central Western
841 area of the State of São Paulo, Brazil. *J Venom Anim Toxins Trop Dis*. 2014 Jun 16;20:27.
- 842 69. Shender LA, De Los Santos M, Montgomery JM, Conrad PA, Gherzi BM, Razuri H, et al.
843 Native Rodent Species Are Unlikely Sources of Infection for *Leishmania (Viannia)*
844 *braziliensis* along the Transoceanic Highway in Madre de Dios, Peru. *PLoS ONE*. 2014 Jul
845 25;9(7):e103358.
- 846 70. Kowaltowski AJ. Brazil's scientists face 90% budget cut. *Nature*. 2021 Oct
847 25;598(7882):566–566.
- 848 71. Araujo Carreira JC, Magalhães M de AFM, Brazil RP, da Silva AVM. *Leishmania* in
849 Marsupials—An Overview of Infection Records in the Americas and Australia. *Open J Anim*
850 *Sci*. 2017;07(03):315–43.

72. Rosypal AC, Alexander A, Byrd D, Weaver M, Stewart R, Gerhold R, et al. Survey Of Antibodies To *Leishmania Spp.* In Wild Canids From Pennsylvania And Tennessee. J Zoo Wildl Med. 2013;44(4):1131–3.
73. Stephens CR, Heau JG, González C, Ibarra-Cerdeña CN, Sánchez-Cordero V, González-Salazar C. Using biotic interaction networks for prediction in biodiversity and emerging diseases. PloS One. 2009 May 28;4(5):e5725.
74. Silva MB da, Portela JM, Li W, Jackson M, Gonzalez-Juarrero M, Hidalgo AS, et al. Evidence of zoonotic leprosy in Pará, Brazilian Amazon, and risks associated with human contact or consumption of armadillos. PLoS Negl Trop Dis. 2018 Jun 28;12(6):e0006532.
75. Barbosa PBBM, de Queiroz PVS, Jerônimo SMB, Ximenes M de FF de M. Experimental infection parameters in *Galea spixii* (Rodentia: Caviidae) with *Leishmania infantum chagasi*. Mem Inst Oswaldo Cruz. 2008 Sep;103(6):545–8.
76. Berzunza-Cruz M, Rodríguez-Moreno Á, Gutiérrez-Granados G, González-Salazar C, Stephens CR, Hidalgo-Mihart M, et al. *Leishmania (L.) mexicana* Infected Bats in Mexico: Novel Potential Reservoirs. PLoS Negl Trop Dis. 2015 Jan 28;9(1):e0003438.
77. Rosa RB, Costa MS da, Teixeira SC, Castro EF de, Dantas WM, Ferro EAV, et al. *Calomys callosus*: An Experimental Animal Model Applied to Parasitic Diseases Investigations of Public Health Concern. Pathog Basel Switz. 2022 Mar 17;11(3):369.
78. García N, Moreno I, Alvarez J, de la Cruz ML, Navarro A, Pérez-Sancho M, et al. Evidence of *Leishmania infantum* Infection in Rabbits (*Oryctolagus cuniculus*) in a Natural Area in Madrid, Spain. BioMed Res Int. 2014;2014:318254.
79. Karayiannis S, Ntais P, Messaritakis I, Tsirigotakis N, Dokianakis E, Antoniou M. Detection of *Leishmania Infantum* in red foxes (*Vulpes vulpes*) in Central Greece. Parasitology. 2015 Nov;142(13):1574–8.
80. Tsokana CN, Sokos C, Giannakopoulos A, Mamuris Z, Birtsas P, Papaspyropoulos K, et al. First evidence of *Leishmania* infection in European brown hare (*Lepus europaeus*) in Greece: GIS analysis and phylogenetic position within the *Leishmania spp.* Parasitol Res. 2016 Jan;115(1):313–21.
81. Kocher A, Cornuault J, Gantier JC, Manzi S, Chavy A, Girod R, et al. Biodiversity and vector-borne diseases: Host dilution and vector amplification occur simultaneously for Amazonian leishmaniasis. Mol Ecol [Internet]. 2022 [cited 2022 Aug 6];n/a(n/a). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.16341>
82. Curtin JM, Aronson NE. Leishmaniasis in the United States: Emerging Issues in a Region of Low Endemicity. Microorganisms. 2021 Mar 11;9(3):578.
83. González C, Wang O, Strutz SE, González-Salazar C, Sánchez-Cordero V, Sarkar S. Climate change and risk of leishmaniasis in north america: predictions from ecological niche models of vector and reservoir species. PLoS Negl Trop Dis. 2010 Jan 19;4(1):e585.
84. McIlwee BE, Weis SE, Hosler GA. Incidence of Endemic Human Cutaneous Leishmaniasis in the United States. JAMA Dermatol. 2018 Sep 1;154(9):1032–9.

895
896
897
898
899
900
901
902
903
904
905

Supplemental materials legend:

Supplementary text 1. Search methods for recording *Leishmania* host status: We used the search string listed below to run search Web of Science of *Leishmania* hosts. We ran the initial search on February 12, 2021. We ran additional searches that included the species names of zoonotic *Leishmania* on February 22, 2021. Our host database was then updated in June 2021 after the publication of (1).

S1 Table Final traits and sources of traits used in the analysis.

S2 Table Summed Shapley Scores per species for newly predicted species, averages across 100 model iterations (0.05 percentile - 0.95 percentile).

S3 Table Complementary evidence of predicted hosts.

S1 Fig Importance of each PCoA phylogenetic dimension in predicting *L. (Viannia)* host status.

S2 Fig Shapley partial dependence plots showing the effect of PCoA dimensions on *L. (Viannia)* host status and distribution of hosts within each order with PCoA dimension values above or below the threshold.

S3 Fig. Features contributions to predictions for the study effort of mammals for animals included in the *L. (Viannia)* analysis.

S4 Fig Functional response of study effort to mammal traits for animals included in the *L. (Viannia)* analysis.

S5 Fig Features contributions to predictions for the study effort of mammals for animals included in the *L. (Leishmania)* analysis.

S6 Fig Functional response of study effort to mammal traits for animals included in the *L. (Leishmania)* analysis.