Phylogenetic and biogeographical traits predict unrecognized hosts of zoonotic leishmaniasis

Caroline K. Glidden[1], Aisling Murran[1], Rafaella Albuquerque Silva[2], Adrian A. Castellanos[3], Barbara A. Han[3], Erin Mordecai[1]

[1]Department of Biology, Stanford University, Stanford, California, USA 94305

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

[3]Cary Institute of Ecosystem Studies, Millbrook, New York, USA 12545

Abstract

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 muco-cutaneous) 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 limiting pathogen spread as well as 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 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 global change on local transmission cycles.

Author summary

Leishmaniasis is a neglected tropical disease with a high burden throughout the Americas. Although the World Health Organization has targeted leishmaniasis for elimination and control by 2030, the disease remains a persistent threat. In some areas, risk of infection is increasing in geographic extent. Leishmaniasis is a zoonotic, vector-borne disease; leishmaniasis is cause by

the parasite *Leishmania,* which is maintained in the environment via an animal host and transmitted from wildlife to humans via sandfly vector. The full suite of wildlife hosts has yet to be documented, likely due to resource limitations. Here, we use machine learning to identify probable wildlife hosts of *Leishmania* throughout the Americas. This work aims to facilitate targeted surveillance and management of leishmaniasis.

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 (Figure 1; PAHO, 2022). Further, incidence is rapidly shifting in geographic extent with land-use and climate change (3–7). For instance, hot spots of leishmaniasis have significantly expanded throughout Costa Rica in the last two decades, in concert with agricultural intensification, deforestation, and reforestation (4).

Leishmaniasis is caused by protists within the *Leishmania* genus. There are 14 known species of zoonotic *Leishmania* (i.e., they spill over from animals to humans) in the Americas. These *Leishmania* species are divided between two subgenera: *Leishmania* (*Vianna*) and *Leishmania (Leishmania)*. In the Americas, *Leishmania (Vianna)* consists of ten species, nine of which are zoonotic; *Leishmania (Leishmania*) consists of seven species, five of which are zoonotic (8). *Leishmania (Vianna)* spp. cause cutaneous and mucocutaneous leishmaniasis, which manifests as skin lesions that are susceptible to painful secondary bacterial infections, and/or destruction of the mucus membranes of the nose, mouth, and throat (WHO). In addition to cutaneous and mucocutaneous leishmaniasis, species within the *Leishmania (Leishmania)* subgenera (*L. (L.) infantum* and, less frequently, *L. (L.) mexicana*) cause fatal visceral leishmaniasis. Visceral

leishmaniasis affects internal organs, typically causing enlargement of the spleen and liver (WHO). Treatment for both visceral and cutaneous leishmaniasis have many side effects and are extremely expensive (PAHO; WHO; (9)), thus prevention and control hinges on limiting transmission.

Leishmania parasites are transmitted by sandflies, 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. In the Americas, there are 137 known wildlife hosts (i.e., animals that are naturally exposed and become infected once exposed), of which around 60 are competent hosts (i.e., animals become infectious once infected, and may maintain the pathogen in the environment as reservoir hosts) (10,11). Wild hosts occur across eight orders of mammals including Chiropetra (bats), Carnivora, Cingulata (armadillos), Didelphimorphia (opossums), Lagomorpha (rabbits), Pilosa (sloths and anteaters), Primates, and Rodentia (Figure 2). However, 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 (e.g., (12–15)).

Limited knowledge on the transmission cycle makes it difficult to understand the mechanisms driving *Leishmania* distribution and spillover. Unknown transmission cycles impede effective ecological interventions to reduce spillover risk to people, such as reducing host densities or establishing barriers to wildlife-human contact (16,17). Additionally, global change can alter host dynamics to amplify spillover risk. For example, land-use change may increase host densities and/or push hosts into closer proximity to anthropophilic vectors and humans (16). Climate warming may also shrink or expand the range of hosts, further changing the rate of contact among hosts, vectors, and humans (18). As such, the incomplete classification of possible *Leishmania* hosts could lead to unexpected emergence and hotspots of human leishmaniasis.

The majority of zoonotic *Leishmania* host discovery occurs through opportunistic sampling and broad surveillance of mammalian communities. However, study designs optimal for detecting *Leishmania* hosts, such as systematic sampling of large populations of a diversity of species through space and time, are limited by logistics and cost of sampling. *Leishmania* host discovery could benefit from strong *a priori* hypotheses about which mammals to dedicate sampling effort; more targeted sampling could help to efficiently use time and monetary resources, ultimately expediting the full classification of the *Leishmania* transmission cycle. Through the last decade, machine learning approaches have been used to generate predictions of hosts of a multitude of different pathogens, including betacoronaviruses (19), flaviviruses (20), Rabies virus (21), and *Borrelia burgdoferi* (causative agent of Lyme disease) (22). These methods have facilitated more targeted laboratory and field work of novel pathogens, and they have great potential to improve our understanding of neglected tropical diseases with broad host and vector ranges.

Host status for any parasite or pathogen is driven by two processes: exposure and physiological susceptibility (Combs 2005; Sweeny & Albery, 2022). Exposure depends on the environmental conditions needed to support pathogen transmission and, in the case of vector-borne disease, vector reproduction, biting, and vector survival (24). Following exposure, physiological susceptibility then depends on pathogen interaction with host cells to gain entry and, in some cases, avoid immune attack and replicate (24). 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 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. In particular, extreme gradient boosted regressions can identify non-linear interactions among numerous traits, even in data-sparse systems with non-random patterns of missing data (25). The flexibility of these models allows for pinpointing combinations of mammalian traits unique to the confluence of *Leishmania* exposure and susceptibility, by learning from 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 extreme gradient boosted regression and mammalian traits, including but not limited to biogeographical, phylogenetic, and life-history features, to: (i) describe trait profiles of hosts for zoonotic *Leishmania (Vianna)* and *Leishmania (Leishmania)* parasites; and (ii) use the trait profiles to predict unrecognized wild hosts of zoonotic *Leishmania* within the subgenera *Leishmania (Vianna)* and *Leishmania (Leishmania)*.

We aim for this analysis to identify animal species that are likely to be exposed and infected in the wild; due to extreme data sparsity, we do not employ our methods to identify competent hosts (i.e., animals that become infectious once infected). Our findings can be used to identify hosts that can then be surveyed at larger spatio-temporal scales, investigated for competence, or used

as sentinel species. As such, while our modeling approach does not necessarily identify competent hosts that can transmit zoonotic *Leishmania* on to sandflies and eventually humans, it does identify animals likely to be exposed to and susceptible to zoonotic *Leishmania* infection, facilitating the first step of determining epidemiological importance and fine-tuning surveillance efforts.

Methods

Data collection

Zoonotic Leishmania *host status:* We gathered *Leishmania* host status from all wild, endemic, and invasive terrestrial mammals with ranges in Mexico, Central America, and South America using recent reviews (10,11) followed by a Web of Science query. Host status was defined as a binary trait: an animal has been naturally infected by a *Leishmania* species known to cause leishmaniasis in humans (1), or there is no record of infection by a *Leishmania* species known to cause leishmaniasis in humans (0). The latter category includes animals that have been tested for *Leishmania* and have not been found to be infected as well as those that have not been tested for *Leishmania* infection. Infection was determined via detection of *Leishmania* specific antibodies, genetic material, or culture of live parasites from tissue samples. To ensure all relevant mammals were included in our study, a list of species names of endemic Latin American mammals was retrieved from the International Union for Conservation of Nature (IUCN) Red List Inventory and a list of invasive mammals was retrieved from the Global Invasive Species Database (GISD). We cross-referenced our host status data table with the Global Infectious Disease and Epidemiology Network (GIDEON). We then limited our analysis to wild mammals that had at least a 10% range overlap with reported human cases of leishmaniasis for inclusion in our analysis. Zoonotic *Leishmania* ranges were constructed by outlining a concave polygon

around cases of disease occurrence in humans from (26) (Figure 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 (supplementary table 1).

For life-history and some habitat use traits, we collected traits from panTHERIA (27), mammalDIET (28), and EltonTraits (29). We additionally used IUCN range maps and satellite imagery to extract data associated with species biogeography, including variables related to climate (30,31), land-cover type (32), and land-use intensity (32,33) within each species range (see supplementary table 2 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 (34). *Leishmania* and species ranges were built using the R (v 4.0.2) packages *speciesgeocodeR* (35) and *concaveman* (36). 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 (37) 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* (38). 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* (39) to use mammal trait data to predict the probability a species is a *Leishmania (Leishmania)* or *Leishmania (Vianna)* 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) (25). 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. Prior to analysis we removed traits with > 0.7 correlation to increase interpretability of variable importance. 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 (40). 3-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, 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 (41,42). We repeated the nested cross-validation analysis but randomized the response variables (*Leishmania* host) 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. (Vianna)* 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* (43). 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 (25). 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 *Leismania* 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 (44,45). 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% or model iterations (21).

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

Results

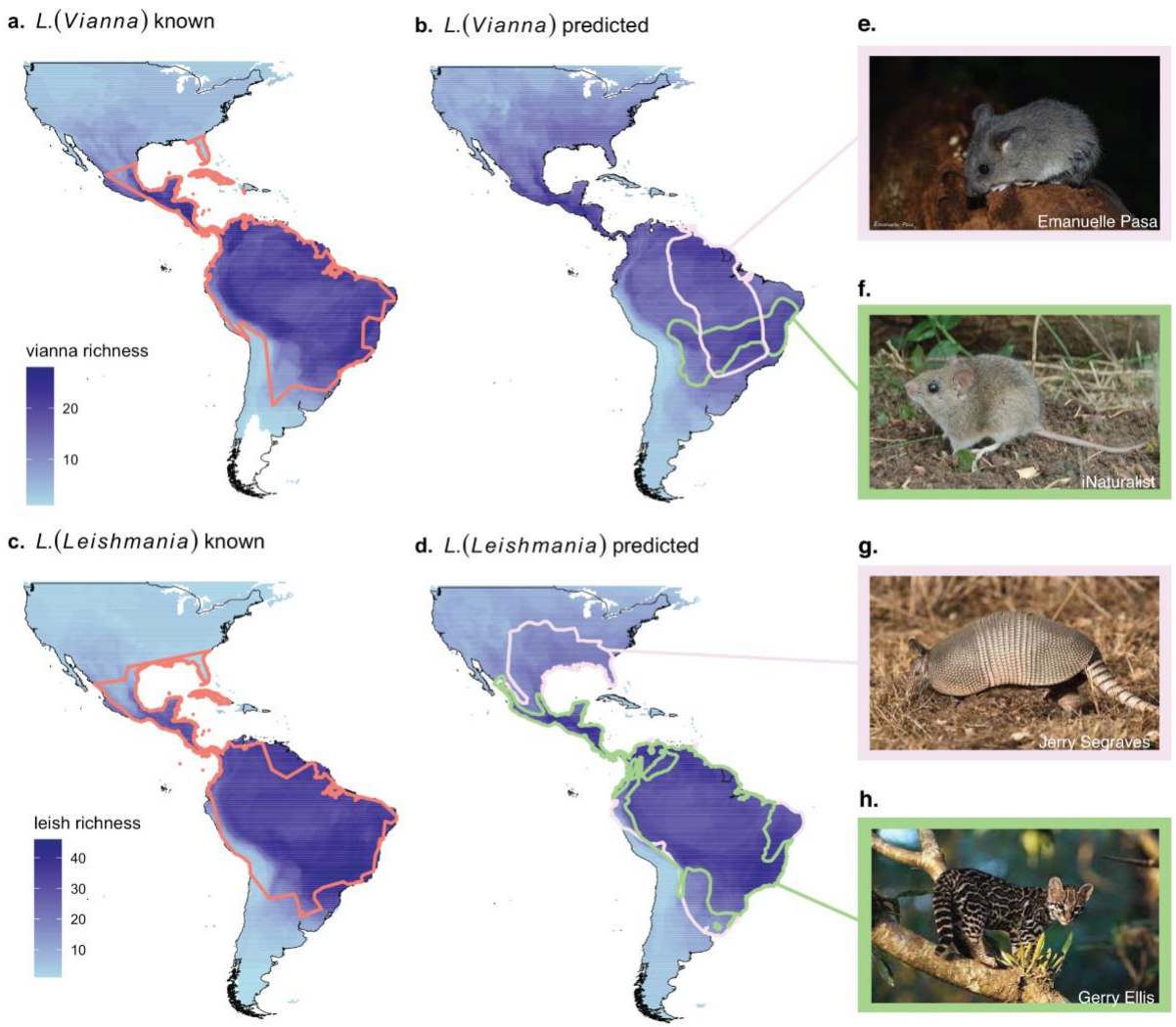


Figure 1. The range of zoonotic *Leishmania* and their hosts extend throughout the Americas. The number of overlapping ranges of known *Leishmania* hosts are depicted in (a & c), with the range of human cases of *L. (Vianna)* and *L. (Leishmania)* outlined in salmon. The number of overlapping ranges of newly predicted hosts are outlined in (b & d). *Oryzomys megacephalus* (e) and *Calomys callosus* (f) are unrecognized hosts with the highest and second highest predictions for *L. (Vianna)* according to Shapley values. Their ranges are outlined in pink and green in (b). *Dasypus novemcinctus* (g) and *Leopardus wiedii* (h) are currently hosts with the highest and second highest Shapley values for *L. (Leishmania).* Their ranges are outlined in pink and green in (d). An exhaustive list of newly predicted hosts are listed in supplementary table 3.

Leishmania (Vianna): Through our literature search, we found 96 known *L. (Vianna)* hosts out of 1460 mammals that occur within the geographic range of *L. (Vianna)* parasites. 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. (Vianna) *trait profiles:* After removing highly correlated traits, 30 features were used to predict *L. (Vianna)* hosts. A mix of biogeography, life-history, phylogeny, and study effort covariates were among the top most important model features when estimating global feature contribution (Figure 3). Phylogenetic distances, reduced into six axes of a PCoA, was, on average, the most important trait for *Leishmania (Vianna)* hosts. Specifically, animals were more likely to be hosts if they were rodents and opposums, and less likely to be hosts if they were non-human primates, carnivores, and even toed-ungulates (supplementary figure 1-2). Similar to other zoonotic pathogen systems, *L. (Vianna)* hosts live at high population densities and demonstrate signatures of fast-paced life-histories, indicated by short gestation periods (Figure 4a). Further, *L. (Vianna)* host ranges are large, have high proportions of crop and urban land cover, and, on average, peak at ~28C during the warmest quarter of the year (Figure 4a). Geographic coordinates of the species' native range was also an important predictor: 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 top most important traits, 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. (Vianna)* host status (supplementary figure 3-4). As such, while *L.* (*Vianna*) hosts are likely understudied, we are not simply reporting the trait profiles of well-studied mammals.

L. (Vianna) *hosts:* Eighty-three animals were identified as likely hosts (Figure 2, supplementary table 2). Out of these newly predicted hosts, the majority were rodents, bats, and opossums. Further, 14 of these hosts are known hosts of *L. (Leishmania)* (supplementary table 3). The newly predicted hosts with the highest summed Shapley scores were the large headed rice rat (*Oryzomys megacephalus*) and the large vesper mouse, *Calomys callosus* (Figure 1e,f). In contrast to known hosts, the range of newly predicted hosts extends to the southern tip of South America and their ranges occur at higher densities in the southern United States (Figure 1a, b).

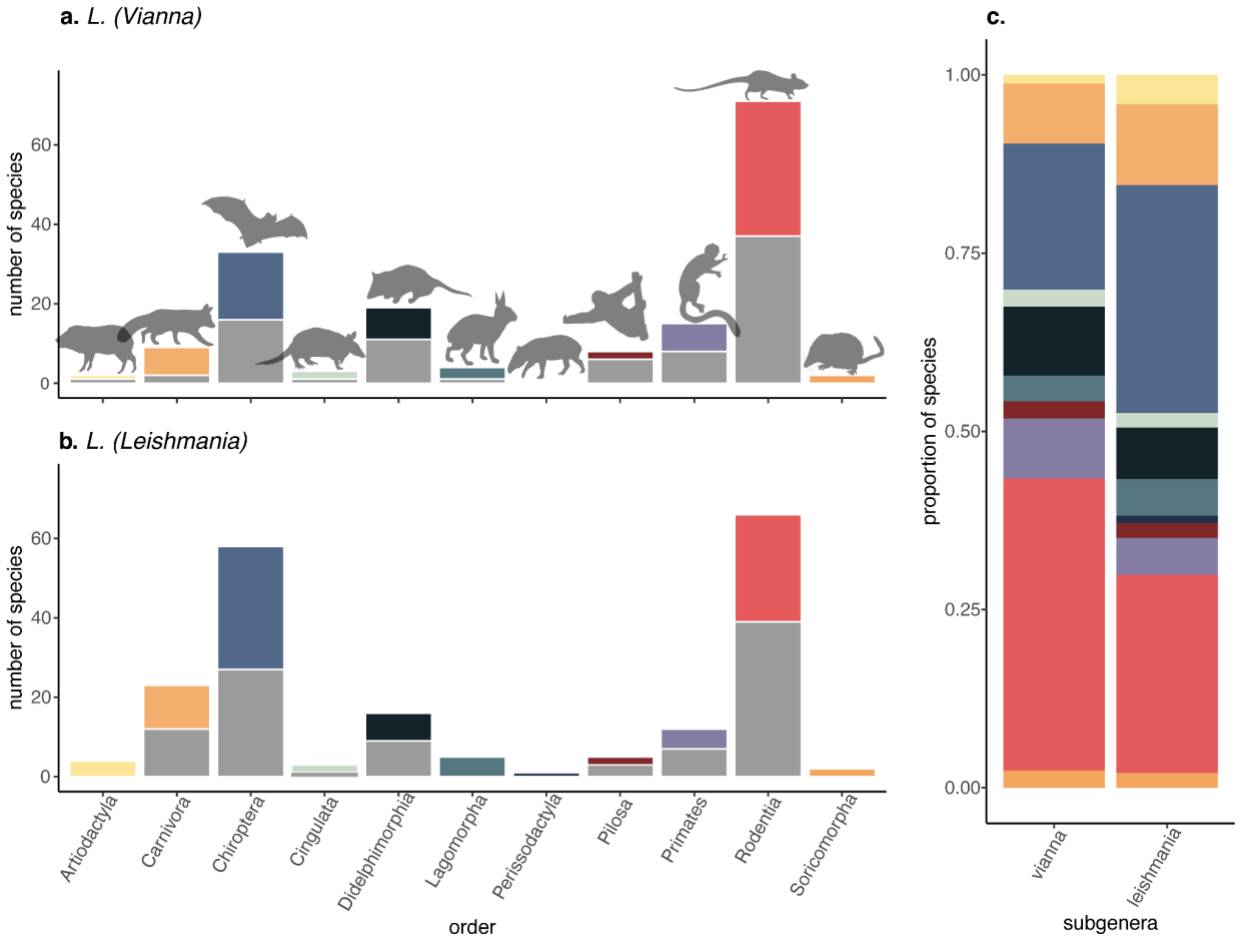


Figure 2. The number of hosts per order. Grey bars represent the number of known hosts within each order, colored bars represent the number of newly predicted hosts per order. Animals were classified as newly predicted if > 95% of mean SHAP values were greater than 0 (a-b). Bars

represent the proportion of newly classified hosts within each order; colors match bar plots in a-b (c).

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. Of these, 40 animals were hosts for both *Leishmania* subgenera. After accounting for target shuffling, our 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 (Figure 3). Similarly to *L. (Vianna)* 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 (Figure 3b). Further, the range of *L. (Leishmania)* hosts are similar to *L. (Vianna)* hosts in that the mean temperature in the warmest quarter peaks around 28C (Figure 4b). On average, geographic coordinates of the species' native range was 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. The latter geographic extent 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 (supplementary figures 5-6).

L. (Leishmania) *hosts*: Using the Shapley classification criterion, 98 animal species that are currently unrecognized hosts were labeled as newly predicted hosts (Figure 2, supplementary table 3). Out of these newly predicted hosts, the majority were rodents, bats, and carnivores. Nine newly-predicted hosts are known hosts of *L. (Vianna)*. Forty-five newly predicted hosts were newly predicted hosts for the *L. (Leishmania)* and *L. (Vianna)* subgeneras (supplementary table 3). The newly predicted hosts with the top summed Shapley scores included the nine-banded armadillo (*Dasypus novemcinctus*) and margay (*Leopardus wiedii*) (Figure 1g,h). The nine-banded armadillo is a known host of *L. (Vianna)* parasites, with a range that extends from northern Argentina to the central United States (Figure 1d).

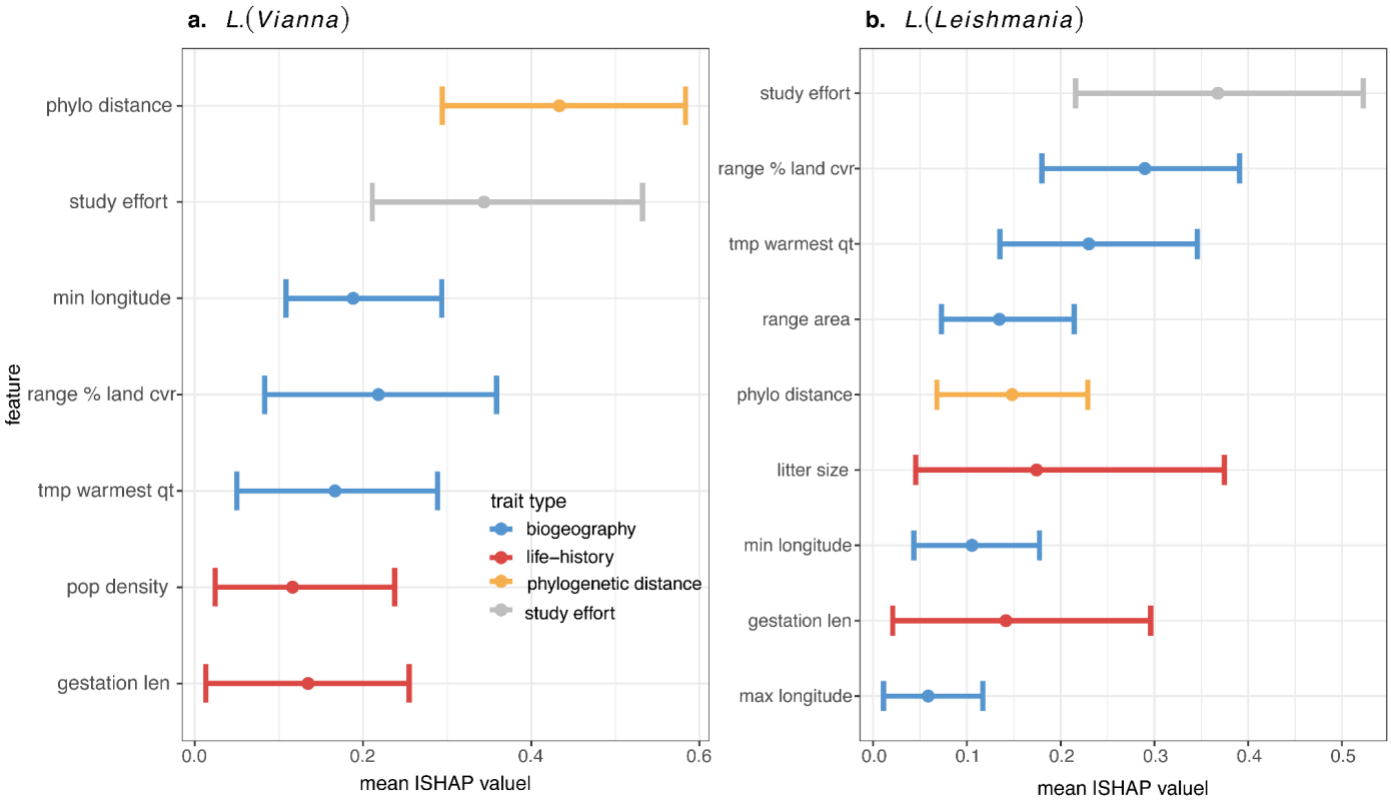
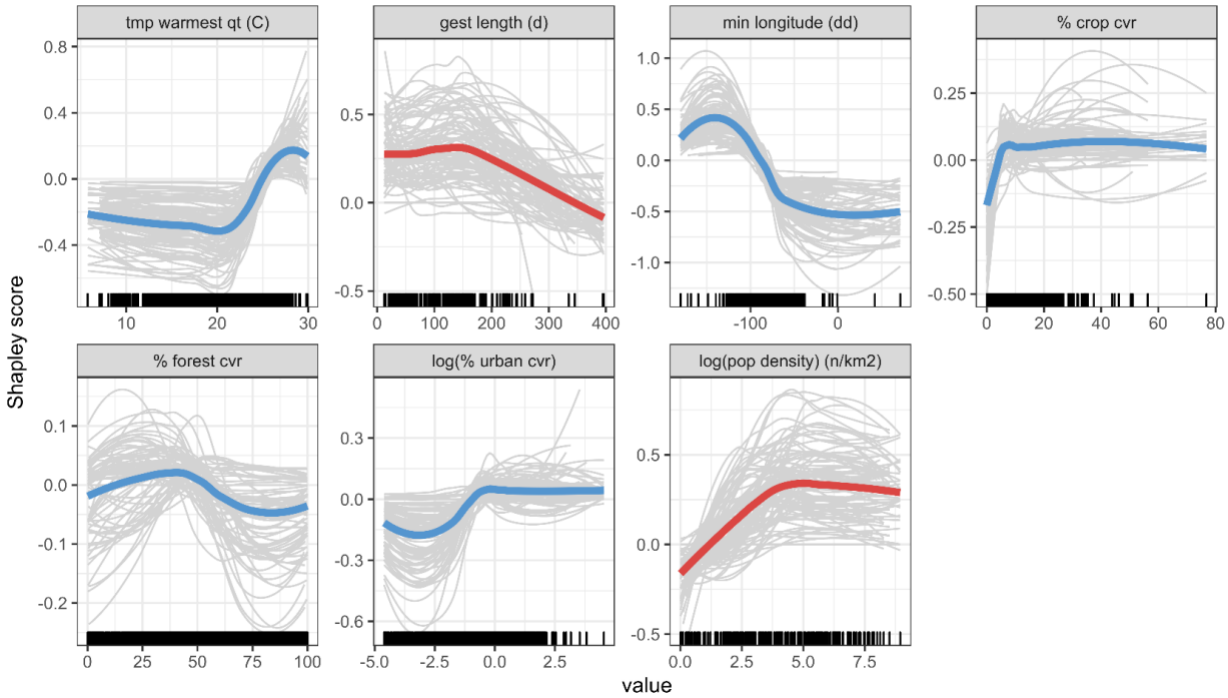


Figure 3. Biogeography, life-history, and phylogenetic traits all significantly contributed to model predictions. Biogeographical traits (minimum longitude of the range, maximum longitude of the range, % cover of land-use/land cover in the species range, average temperature in the warmest quarter in the species range, range area) are colored in blue, life-history traits (population density, gestation length, litter size) are colored in red, phylogenetic distance 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.

a. *L.(Vianna)*



b. *L.(Leishmania)*

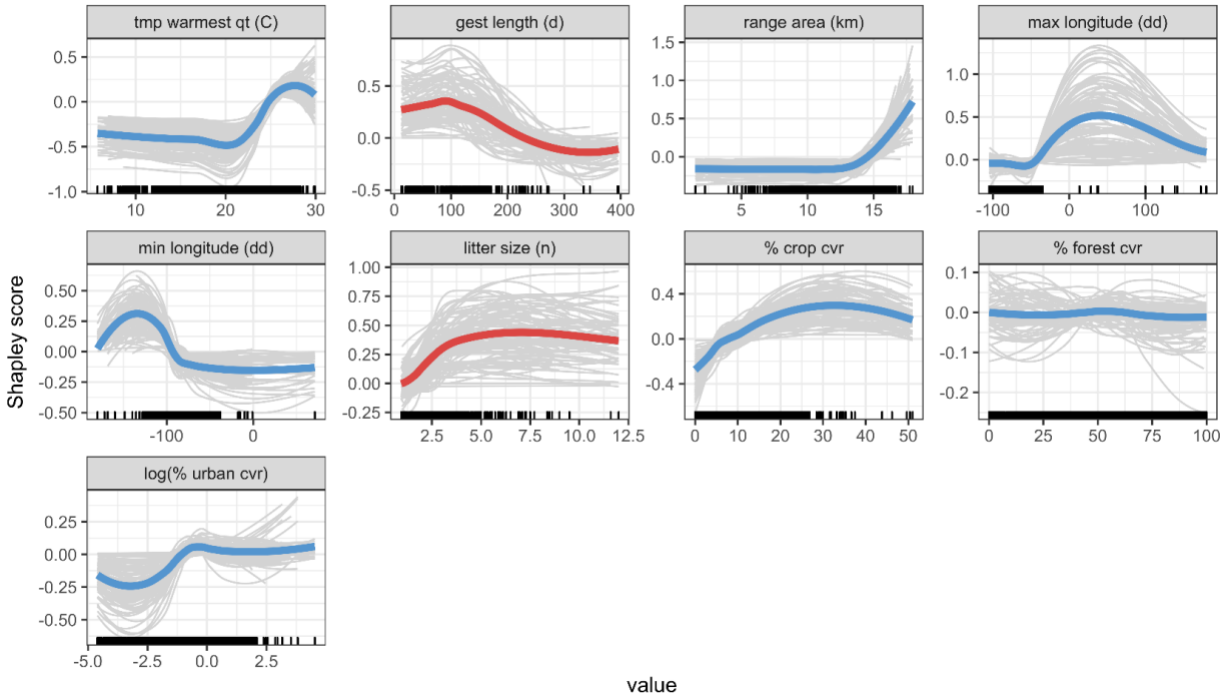


Figure 4. Hosts have fast-paced life-histories and live in proximity to humans. Shapley partial dependence plots showing the effect of each feature on *L.* (*Vianna*) (a) and *L. (Leishmania)* (b) host status after accounting for the average effect of the other features in the model. Colored lines represent the average effect across model iterations, while grey lines show each individual model iteration (model fit with 80% of data) (blue = 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. The majority of these predicted hosts for both subgenera fell within the Rodentia order, with a similar number of predicted *Leishmania (Leishmania)* hosts within the Chiroptera order. 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. Geographic range area was also reported to be a predictive feature shared in common among zoonotic hosts and reservoirs in other systems (Han et al., 2015; Han et al. 2020; Worsley-Tonks et al., 2020). Further, ranges of *L. (Leishmania)* and *L. (Vianna)* 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 (47). Further, our results align with recent literature showing that risk of leishmaniasis is associated with agricultural intensification and urbanization (3,48). *L. (Vianna)* hosts are marginally more likely to have ranges with moderate forest cover, 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 (49), while hosts are more common in disturbed landscapes (50). 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) (6,51). Spillover occurs when reservoir hosts, vectors, and humans overlap in space and time so that forward transmission occurs from animal - sandfly - 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 of some species (Gùzman & Tesh 2000; Hlavacova et al., 2013).

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 (46,53). 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 (21,46). 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 (54). As a result, animals with fast-paced life histories are thought to divert energetic resources away from effective adaptive immunity (55), or utilize responses conducive to tolerance but not resistance (56), so as to allocate maximal resources to short term reproductive costs. As such, animals with fast-paced life-histories may be more susceptible to infections.

However, *L. (Vianna)* 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. (Vianna)* 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 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 (e.g., (14,50,57–61)). 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. Researchers across Latin America are faced with similar, if not more severe, funding constraints. For example, in Brazil, a major hotspot of leishmaniasis, the current administration cut 90% of federal science funding (62). While our model results facilitate more targeted field sampling for particular species, optimized surveillance should consist of an iterative feedback between model predictions to guide field work, and updating the model with new results from field sampling to guide ongoing surveillance and field studies (e.g., (14,19)). Importantly, to improve study effort bias going forward, care should be taken to estimate and survey undersampled areas and species in addition to newly predicted species. 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. Overall, leishmaniasis surveillance and management would benefit from larger funding mechanisms prioritizing understanding the ecology of this group of pathogens.

Models for both subgenera identified several new hosts that can be targeted for future surveillance and research to determine their ecological significance in leishmaniaisis transmission and spillover to humans. In some cases, these predictions are supported by multiple independent but complementary lines of evidence (summarized below and in supplementary table 4). 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 opposum (*Monodelphis domestica*) (10,60,63,64). 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*) (Carreira et al.

2017); 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 oppossum (*Didlephis virginiana*), kinkajou (*Potos flavus*), racoon (*Procyon lotor*), white-nose coati (*Nasua narica*), and the margay (*Leopardus wiedii*) (Stephens et al. 2009; Stephens et al. 2016). 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. (Vianna)* as well as other zoonotic pathogens including the causative agent of leprosy (da Silva et al., 2018).

Host competence (i.e., the ability to become infected and infectious) has been demonstrated in the lab for four species: the Virginia oppossum, the large vesper mouse (*Calomys callosus*), spix's yellow toothed cavy (*Galea spixii*), and the Mexican free-tailed bat (*Tadarida brasiliensis*) (Araujo Carreira et al., 2017; Barbosa et al., 2008; Berzunza-Cruz et al., 2015; Rosa et al., 2022). Our model suggests that these animals are also likely to be naturally infected in the wild; the large vesper mouse is one of the top predicted species for *L. (Vianna).* 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*) (69–71).

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 (72–74). Similarly, the range of predicted hosts for *L. (Vianna)* 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.

Importantly, there are 14 species of zoonotic *Leishmania* parasites in the Americas. While our analysis divides these species into two subgenera we do 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, (49) used machine learning to predict unrecognized vectors of leishmaniasis. Further, advance sequencing of arthropods has started to identify sandfly feeding networks (Kocher et al., 2022; Massey et al., 2022) that may be useful in linking model predictions of vectors and hosts to local transmission cycles.

Overall, we estimate that there are likely many unknown hosts of leishmaniasis throughout the Americas, highlighting the complexity of the transmission cycle and a need to increase study effort on the ecology and epidemiology of this system. Identifying local transmission cycles is key for

reducing leishmaniasis spillover and disease burden in people, especially in the face of land-use and climate change, which likely influence the geographic distribution and extent of vector-host-human contact, and were strongly associated with novel host predictions in our models. Our analysis included a number of potential new hosts, including but not limited to the nine-banded armadillo, the margay, the large vesper mouse, and the large headed rice rat. Future work should further investigate the role of newly identified hosts in the zoonotic *Leishmania* transmission cycle and iteratively update the current analysis as new information is collected, taking careful consideration to not perpetuate the effects of study bias.

Acknowledgements

This project was supported by the National Science Foundation (DEB-2011147 CKG, EAM; EEID-1717282 BAH), the National Institutes of Health (R35GM133439; EAM), the Stanford King Center for Global Development (EAM), the Stanford Woods Institute for the Environment (EAM), the Stanford Center for Innovation in Global Health (EAM), and the Office of the Vice Provost for Undergraduate Education at Stanford (AM). We thank the Mordecai Lab for providing feedback on manuscript drafts.

References

1. WHO. The road map targets for 2030 [Internet]. The road map targets for 2030. 2022 [cited 2022 Jun 12]. Available from: https://www.who.int/teams/control-of-neglected-tropical-diseases/ending-ntds-together-towards-2030/targets
2. PAHO. Leishmaniasis - PAHO/WHO | Pan American Health Organization [Internet]. 2022 [cited 2022 Jun 12]. Available from: https://www.paho.org/en/topics/leishmaniasis
3. Bayles BR, Rusk A, Christofferson R, Agar G, Pineda MA, Chen B, et al. Spatiotemporal dynamics of vector-borne disease risk across human land-use gradients: examining the role of agriculture, indigenous territories, and protected areas in Costa Rica. Lancet Glob Health. 2020 Apr 1;8:S32.
4. Bayles BR, Rusk AE, Pineda MA, Chen B, Dagy K, Hummel T, et al. Spatiotemporal trends of cutaneous leishmaniasis in Costa Rica. Trans R Soc Trop Med Hyg. 2021 May 8;115(5):569–71.
5. Dutari LC, Loaiza JR. American Cutaneous Leishmaniasis in Panama: a historical review of entomological studies on anthropophilic Lutzomyia sand fly species. Parasit Vectors. 2014 May 11;7(1):218.
6. Gonçalves NV, Miranda C do SC, Costa RJF da, Guedes JA, Matsumura ES de S, Costa

SBN da, et al. Cutaneous leishmaniasis: Spatial distribution and environmental risk factors in the state of Pará, Brazilian Eastern Amazon. J Infect Dev Ctries. 2019 Oct 31;13(10):939–44.

7. Portella TP, Kraenkel RA. Spatial-temporal pattern of cutaneous leishmaniasis in Brazil. Infect Dis Poverty. 2021 Jun;10(03):47–57.
8. Akhoundi M, Kuhls K, Cannet A, Votýpka J, Marty P, Delaunay P, et al. A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies. PLoS Negl Trop Dis. 2016 Mar 3;10(3):e0004349.
9. Chakravarty J, Sundar S. Current and emerging medications for the treatment of leishmaniasis. Expert Opin Pharmacother. 2019 Jul 3;20(10):1251–65.
10. Azami-Conesa I, Gómez-Muñoz MT, Martínez-Díaz RA. A Systematic Review (1990–2021) of Wild Animals Infected with Zoonotic Leishmania. Microorganisms. 2021 May 20;9(5):1101.
11. Roque ALR, Jansen AM. Wild and synanthropic reservoirs of Leishmania species in the Americas. Int J Parasitol Parasites Wildl. 2014 Aug 29;3(3):251–62.
12. López M, Erazo D, Hoyos J, León C, Fuya P, Lugo L, et al. Measuring spatial co-occurrences of species potentially involved in Leishmania transmission cycles through a predictive and fieldwork approach. Sci Rep. 2021 Mar 24;11(1):6789.
13. Medkour H, Davoust B, Levasseur A, Mediannikov O. Molecular Evidence of Leishmania infantum and Leishmania guyanensis in Red Howler Monkey (Alouatta seniculus) from French Guiana. Vector-Borne Zoonotic Dis. 2019 Dec;19(12):896–900.
14. Stephens CR, González-Salazar C, Sánchez-Cordero V, Becker I, Rebollar-Tellez E, Rodríguez-Moreno Á, et al. Can You Judge a Disease Host by the Company It Keeps? Predicting Disease Hosts and Their Relative Importance: A Case Study for Leishmaniasis. PLoS Negl Trop Dis. 2016 Oct 7;10(10):e0005004.
15. Vieira TM, Silva S de O, Lima L, Sabino-Santos G, Duarte ER, Lima SM, et al. Leishmania diversity in bats from an endemic area for visceral and cutaneous leishmaniasis in Southeastern Brazil. Acta Trop. 2022 Apr 1;228:106327.
16. Glidden CK, Nova N, Kain MP, Lagerstrom KM, Skinner EB, Mandle L, et al. Human-mediated impacts on biodiversity and the consequences for zoonotic disease spillover. Curr Biol. 2021 Oct 11;31(19):R1342–61.
17. Sokolow SH, Nova N, Pepin KM, Peel AJ, Pulliam JRC, Manlove K, et al. Ecological interventions to prevent and manage zoonotic pathogen spillover. Philos Trans R Soc Lond B Biol Sci. 2019 Sep 30;374(1782):20180342.
18. Carlson CJ, Albery GF, Merow C, Trisos CH, Zipfel CM, Eskew EA, et al. Climate change increases cross-species viral transmission risk. Nature. 2022 Jul;607(7919):555–62.
19. Becker DJ, Albery GF, Sjodin AR, Poisot T, Bergner LM, Chen B, et al. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. Lancet Microbe [Internet]. 2022 Jan 10 [cited 2022 Jun 12]; Available from: https://www.sciencedirect.com/science/article/pii/S2666524721002457
20. Han BA, Majumdar S, Calmon FP, Glicksberg BS, Horesh R, Kumar A, et al. Confronting data sparsity to identify potential sources of Zika virus spillover infection among primates. Epidemics. 2019 Jun 1;27:59–65.
21. Worsley-Tonks KEL, Escobar LE, Biek R, Castaneda-Guzman M, Craft ME, Streicker DG, et al. Using host traits to predict reservoir host species of rabies virus. PLoS Negl Trop Dis. 2020 Dec 8;14(12):e0008940.
22. Becker DJ, Han BA. The macroecology and evolution of avian competence for Borrelia burgdorferi. Glob Ecol Biogeogr. 2021;30(3):710–24.
23. Sweeny AR, Albery GF. Exposure and susceptibility: The Twin Pillars of infection. Funct Ecol. 2022 May 18;1365-2435.14065.
24. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, et al. Pathways to

zoonotic spillover. Nat Rev Microbiol. 2017 Aug;15(8):502–10.
25. Molnar C. Interpretable Machine Learning [Internet]. 2nd ed. 2022 [cited 2022 Aug 31]. Available from: https://christophm.github.io/interpretable-ml-book/
26. Herrera G, Barragán N, Luna N, Martínez D, De Martino F, Medina J, et al. An interactive database of Leishmania species distribution in the Americas. Sci Data. 2020 Apr 3;7(1):110.
27. Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, et al. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology. 2009;90(9):2648–2648.
28. Kissling WD, Dalby L, Fløjgaard C, Lenoir J, Sandel B, Sandom C, et al. Establishing macroecological trait datasets: digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. Ecol Evol. 2014;4(14):2913–30.
29. Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. Ecology. 2014;95(7):2027–2027.
30. Brun P, Zimmermann NE, Hari C, Pellissier L, Karger DN. Global climate-related predictors at kilometre resolution for the past and future. Earth Syst Sci Data Discuss. 2022 Jun 27;1–44.
31. Karger DN, Schmatz DR, Dettling G, Zimmermann NE. High-resolution monthly precipitation and temperature time series from 2006 to 2100. Sci Data. 2020 Jul 23;7(1):248.
32. Buchhorn M, Lesiv M, Tsendbazar NE, Herold M, Bertels L, Smets B. Copernicus Global Land Cover Layers—Collection 2. Remote Sens. 2020 Jan;12(6):1044.
33. Kennedy CM, Oakleaf JR, Theobald DM, Baruch-Mordo S, Kiesecker J. Managing the middle: A shift in conservation priorities based on the global human modification gradient. Glob Change Biol. 2019;25(3):811–26.
34. Rosa CA da, Ribeiro BR, Bejarano V, Puertas FH, Bocchiglieri A, Barbosa AL dos S, et al. NEOTROPICAL ALIEN MAMMALS: a data set of occurrence and abundance of alien mammals in the Neotropics. Ecology. 2020;101(11):e03115.
35. Töpel M, Zizka A, Calió MF, Scharn R, Silvestro D, Antonelli A. SpeciesGeoCoder: Fast Categorization of Species Occurrences for Analyses of Biodiversity, Biogeography, Ecology, and Evolution. Syst Biol. 2017 Mar 1;66(2):145–51.
36. Gombin J, Vaidyanathan R, Agafonkin V. concaveman. R. 2020;
37. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol. 2017 Jul 1;34(7):1812–9.
38. Fantini D. easyPubMed – Data Pulse [Internet]. 2019 [cited 2022 Aug 31]. Available from: https://www.data-pulse.com/dev_site/easypubmed/
39. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. San Francisco California USA: ACM; 2016 [cited 2022 Aug 31]. p. 785–94. Available from: https://dl.acm.org/doi/10.1145/2939672.2939785
40. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLOS ONE. 2019 Nov 7;14(11):e0224365.
41. Fischhoff IR, Castellanos AA, Rodrigues JPGLM, Varsani A, Han BA. Predicting the zoonotic capacity of mammals to transmit SARS-CoV-2. Proc Biol Sci. 2021 Nov 24;288(1963):20211651.
42. Martin JT, Fischhoff IR, Castellanos AA, Han BA. Ecological Predictors of Zoonotic Vector Status Among *Dermacentor* Ticks (Acari: Ixodidae): A Trait-Based Approach. Gaff H, editor. J Med Entomol. 2022 Sep 5;tjac125.
43. Liu Y, Just A. SHAPforxgboost: SHAP Plots for "XGBoost". R Package Version 010. 2020;
44. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions [Internet]. arXiv; 2017 [cited 2022 Sep 21]. Available from: http://arxiv.org/abs/1705.07874
45. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. Explainable AI for

Trees: From Local Explanations to Global Understanding [Internet]. arXiv; 2019 [cited 2022 Sep 21]. Available from: http://arxiv.org/abs/1905.04610

46. Han BA, Schmidt JP, Bowden SE, Drake JM. Rodent reservoirs of future zoonotic diseases. Proc Natl Acad Sci. 2015 Jun 2;112(22):7039–44.
47. Gibb R, Redding DW, Chin KQ, Donnelly CA, Blackburn TM, Newbold T, et al. Zoonotic host diversity increases in human-dominated ecosystems. Nature. 2020 Aug;584(7821):398–402.
48. Lima ÁLM, de Lima ID, Coutinho JFV, de Sousa ÚPST, Rodrigues MAG, Wilson ME, et al. Changing epidemiology of visceral leishmaniasis in northeastern Brazil: a 25-year follow-up of an urban outbreak. Trans R Soc Trop Med Hyg. 2017 Oct 1;111(10):440–7.
49. Vadmal GM, Glidden CK, Han BA, Carvalho B, Castellanos AA, Mordecai EA. Data-Driven Predictions of Potential Leishmania Vectors in Latin America [Internet]. bioRxiv; 2022 [cited 2022 Oct 10]. p. 2022.08.22.504897. Available from: https://www.biorxiv.org/content/10.1101/2022.08.22.504897v1
50. 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 leishmaniases. Mol Ecol [Internet]. 2022 [cited 2022 Aug 6];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.16341
51. Silva J, Queiroz A, Moura I, Sousa RS, Guimarães LH, Machado PRL, et al. Dynamics of American tegumentary leishmaniasis in a highly endemic region for Leishmania (Viannia) braziliensis infection in northeast Brazil. PLoS Negl Trop Dis. 2017 Nov 2;11(11):e0006015.
52. Hlavacova J, Votypka J, Volf P. The effect of temperature on Leishmania (Kinetoplastida: Trypanosomatidae) development in sand flies. J Med Entomol. 2013 Sep;50(5):955–8.
53. Ostfeld RS, Levi T, Jolles AE, Martin LB, Hosseini PR, Keesing F. Life history and demographic drivers of reservoir competence for three tick-borne zoonotic pathogens. PloS One. 2014;9(9):e107387.
54. Demas N, Nelson R. Ecoimmunology. Oxford, New York: Oxford University Press; 2011. 656 p.
55. Martin LB, Weil ZM, Nelson RJ. Immune defense and reproductive pace of life in Peromyscus mice. Ecology. 2007 Oct;88(10):2516–28.
56. Schwanz LE. Persistent Effects of Maternal Parasitic Infection on Offspring Fitness: Implications for Adaptive Reproductive Strategies When Parasitized. Funct Ecol. 2008;22(4):691–8.
57. Caldart ET, Pinto-Ferreira F, Matos AMRN de, Pascoal ATP, Bertão-Santos A, Mitsuka-Breganó R, et al. Evaluation of an active and early surveillance methodology for visceral leishmaniasis by molecular detection in road-killed wild fauna. Rev Bras Parasitol Veterinária. 2021;30(2):e027920.
58. Massey AL, Bronzoni RV de M, da Silva DJF, Allen JM, de Lázari PR, dos Santos-Filho M, et al. Invertebrates for vertebrate biodiversity monitoring: Comparisons using three insect taxa as iDNA samplers. Mol Ecol Resour. 2022;22(3):962–77.
59. Patiño LH, Castillo-Castañeda AC, Muñoz M, Jaimes JE, Luna-Niño N, Hernández C, et al. Development of an Amplicon-Based Next-Generation Sequencing Protocol to Identify Leishmania Species and Other Trypanosomatids in Leishmaniasis Endemic Areas. Microbiol Spectr. 2021 Oct 31;9(2):e0065221.
60. Richini-Pereira VB, Marson PM, Hayasaka EY, Victoria C, da Silva RC, Langoni H. Molecular detection of Leishmania spp. in road-killed wild mammals in the Central Western area of the State of São Paulo, Brazil. J Venom Anim Toxins Trop Dis. 2014 Jun 16;20:27.
61. Shender LA, De Los Santos M, Montgomery JM, Conrad PA, Ghersi BM, Razuri H, et al. Native Rodent Species Are Unlikely Sources of Infection for Leishmania (Viannia) braziliensis along the Transoceanic Highway in Madre de Dios, Peru. PLoS ONE. 2014 Jul 25;9(7):e103358.

62. Kowaltowski AJ. Brazil's scientists face 90% budget cut. Nature. 2021 Oct 25;598(7882):566–566.
63. Araujo Carreira JC, Magalhães M de AFM, Brazil RP, da Silva AVM. Leishmania in Marsupials—An Overview of Infection Records in the Americas and Australia. Open J Anim Sci. 2017;07(03):315–43.
64. 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.
65. 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.
66. 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.
67. 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.
68. 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.
69. 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.
70. 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.
71. 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.
72. Curtin JM, Aronson NE. Leishmaniasis in the United States: Emerging Issues in a Region of Low Endemicity. Microorganisms. 2021 Mar 11;9(3):578.
73. 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.
74. 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.

Supplemental materials

Supplementary Table 1. Final traits and sources of traits used in the analysis.

trait	***definition***	***source***
neonate body mass (g)	mass of neonates	panTHERIA
adult forearm length (mm)	total length from elbow to wrist	panTHERIA
gestation length (d)	length of time of non-inactive fetal growth	panTHERIA
litter size	average # of offspring in a litter	panTHERIA
litters per year	average # of litters per year	panTHERIA
max longevity (m)	maximum adult age	panTHERIA
social group size (n)	# of individuals in a group	panTHERIA
weaning age (d)	age when primary nutritional dependency on the mother ends and independent foraging begins to make a major contribution to the offspring's energy	panTHERIA
activity cycle	activity cycle (nocturnal, crepuscular, diurnal)	panTHERIA
dominate diet	primary food group consumed by the animal, including groups related to sandfly habitat and exposure (insectivore, frugivore, granivore, folivore)	mammalDiet
diet breadth	number of dietary categories eaten by each	panTHERIA

	species	
home range size (km^2)	size of the area within which everyday activities of individuals or groups (of any type) are typically restricted	panTHERIA
population density (n/km^2)	population size per km^2	panTHERIA
foraging strata	assignment to one of five foraging stratum categories (ground level, scansorial, arboreal, aerial)	EltonTraits
trophic strata	assignment to one of three trophic levels: carnivore, omnivore, herbivore	EltonTraits
temperature seasonality		CHELSA
temperature in the warmest quarter		CHELSA
precipitation in the driest quarter		CHELSA
humidity range		CHELSA
climate moisture index range		CHELSA
main habitat	Main habitat defined by IUCN, includes forest, savanna, shrubland, grassland, wetland, rocky area, desert, artificial	IUCN
habitat breadth	number of IUCN defined main habitats	
percent crop cover	average percent of the range covered by crop land from 2015-2019	IUCN, gee[a]: COPERNICUS/Landcover/100m/Proba-V-C3/Global
percent tree cover	average percent of the range covered by trees from 2015-2019	IUCN, gee[a]: COPERNICUS/Landcover/100m/Proba-V-C3/Global
percent urban cover	average percent of the range covered by urban landscape from 2015-2019	IUCN, gee[a]: COPERNICUS/Landcover/100m/Proba-V-C3/Global
global human modification index	aggregate measure of the human modification; mean gHm within an animal's range in 2016	IUCN, gee[a]: CSP/HM/GlobalHumanModification

range area (km^2)		panTHERIA
maximum longitude of range (dd)		panTHERIA
minimum longitude of range (dd)		panTHERIA
median latitude of range (dd)		panTHERIA
phylogenetic distance	Pairwise divergence time, averaged across all available sources for species pairs, reduced into five dimensions via PCoA	TimeTree
invasive	invasive in Mexico, Central America, or South America	GISD
zoonotic host	reservoir host of any zoonotic pathogen	GIDEON
citation count	number of citations found on pubmed	pubmed
[a] google earth engine		

Supplementary Table 2. Summed Shapley Scores per species for newly predicted species, averages across 100 model iterations (0.05 percentile - 0.95 percentile). Grey boxes represent animals that are already known hosts for the subgenera.

Binomial	**Order**	***L. (Vianna)* scores**	***L. (Leishmania)* scores**
Odocoileus virginianus	Artiodactyla	1.40 (0.01-2.58)	1.46 (0.54-2.17)
Pecari tajacu	Artiodactyla	--	2.28 (1.84-2.83)
Sus scrofa	Artiodactyla	--	1.49 (0.38-2.34)
Tayassu pecari	Artiodactyla	--	2.22 (1.70-2.92)
Bassariscus astutus	Carnivora	--	0.74 (0.31-1.75)
Canis latrans	Carnivora	--	1.39 (0.75-1.92)
Canis lupus	Carnivora	0.86 (0.18-1.45)	1.84 (1.02-2.73)
Leopardus wiedii	Carnivora	0.89 (0.26-1.61)	2.74 (2.22-3.14)
Lontra longicaudis	Carnivora	--	1.74 (1.05-2.35)
Mustela frenata	Carnivora	--	1.64 (0.88-2.44)
Nasua narica	Carnivora	1.29 (0.10-2.30)	1.48 (0.70-2.19)
Potos flavus	Carnivora	1.77 (1.02-2.37)	--
Procyon cancrivorus	Carnivora	--	2.34 (1.57-3.01)
Procyon lotor	Carnivora	1.02 (0.19-2.25)	1.95 (1.21-2.65)
Puma concolor	Carnivora	0.89 (0.23-1.69)	--
Puma yagouaroundi	Carnivora	0.86 (0.24-1.57)	2.20 (1.57-2.82)
Vulpes vulpes	Carnivora	--	1.91 (1.31-2.63)
Anoura caudifer	Chiroptera	--	1.49 (0.43-2.23)
Anoura geoffroyi	Chiroptera	1.23 (0.21-2.31)	2.15 (1.48-2.86)
Artibeus phaeotis	Chiroptera	1.05 (0.54-1.73)	0.97 (0.27-1.45)

Carollia perspicillata	Chiroptera	2.16 (1.43-2.87)	--
Chiroderma villosum	Chiroptera	--	0.98 (0.22-1.63)
Choeronycteris mexicana	Chiroptera	0.60 (0.10-1.29)	--
Chrotopterus auritus	Chiroptera	--	1.75 (0.57-2.55)
Diclidurus albus	Chiroptera	1.11 (0.37-1.94)	1.43 (0.45-2.34)
Diphylla ecaudata	Chiroptera	--	1.79 (0.82-2.42)
Eptesicus brasiliensis	Chiroptera	--	0.95 (0.23-1.73)
Eptesicus furinalis	Chiroptera	1.01 (0.09-1.87)	--
Eptesicus fuscus	Chiroptera	1.11 (0.31-1.86)	1.41 (0.72-1.94)
Lasionycteris noctivagans	Chiroptera	1.41 (0.73-2.09)	1.10 (0.28-1.70)
Lasiurus borealis	Chiroptera	1.04 (0.21-1.99)	1.24 (0.26-2.17)
Lasiurus cinereus	Chiroptera	--	1.33 (0.64-1.89)
Lasiurus ega	Chiroptera	--	2.11 (1.36-2.86)
Lasiurus intermedius	Chiroptera	--	0.84 (0.18-1.78)
Leptonycteris yerbabuenae	Chiroptera	0.73 (0.07-1.38)	--
Macrophyllum macrophyllum	Chiroptera	--	1.04 (0.59-1.64)
Macrotus waterhousii	Chiroptera	--	0.83 (0.15-1.49)
Micronycteris buriri	Chiroptera	--	0.86 (0.38-1.39)
Mormoops megalophylla	Chiroptera	--	1.16 (0.51-1.93)
Myotis albescens	Chiroptera	--	1.66 (0.56-2.38)
Myotis riparius	Chiroptera	--	0.88 (0.31-1.64)
Natalus stramineus	Chiroptera	--	1.08 (0.23-1.97)
Noctilio albiventris	Chiroptera	--	1.86 (1.12-2.54)
Noctilio leporinus	Chiroptera	--	1.21 (0.48-2.12)

Nyctinomops aurispinosus	Chiroptera	--	1.07 (0.51-1.87)
Peropteryx macrotis	Chiroptera	--	1.14 (0.29-1.78)
Phyllostomus hastatus	Chiroptera	1.34 (0.54-2.08)	--
Pipistrellus subflavus	Chiroptera	1.41 (0.65-1.98)	1.25 (0.08-1.88)
Pteronotus parnellii	Chiroptera	1.43 (0.66-2.19)	--
Pteronotus quadridens	Chiroptera	0.82 (0.07-1.52)	--
Saccopteryx bilineata	Chiroptera	--	1.88 (1.08-2.51)
Sturnira lilium	Chiroptera	2.19 (1.19-3.02)	--
Sturnira tildae	Chiroptera	--	1.12 (0.25-2.11)
Tadarida brasiliensis	Chiroptera	1.70 (0.86-2.53)	2.33 (1.93-2.73)
Thyroptera tricolor	Chiroptera	--	1.25 (0.20-2.25)
Trachops cirrhosus	Chiroptera	--	0.77 (0.13-1.58)
Uroderma bilobatum	Chiroptera	1.28 (0.31-2.06)	1.82 (0.93-2.54)
Dasypus novemcinctus	Cingulata	--	3.05 (2.57-3.42)
Euphractus sexcinctus	Cingulata	0.90 (0.22-1.61)	1.98 (1.32-2.69)
Priodontes maximus	Cingulata	0.82 (0.08-1.40)	--
Caluromys derbianus	Didelphimorphia	2.04 (1.06-2.85)	1.23 (0.67-1.82)
Caluromys philander	Didelphimorphia	--	2.06 (1.33-2.68)
Chironectes minimus	Didelphimorphia	--	1.28 (0.37-1.99)
Didelphis virginiana	Didelphimorphia	2.28 (1.52-3.21)	1.77 (0.91-2.62)
Gracilinanus agilis	Didelphimorphia	--	1.17 (0.28-2.00)
Lutreolina crassicaudata	Didelphimorphia	--	1.41 (0.18-2.22)
Marmosa murina	Didelphimorphia	1.51 (0.32-2.38)	--
Metachirus nudicaudatus	Didelphimorphia	1.67 (0.76-2.54)	--

Monodelphis brevicaudata	Didelphimorphia	1.53 (0.38-2.54)	--
Monodelphis domestica	Didelphimorphia	1.56 (0.55-2.57)	1.85 (0.89-2.59)
Thylamys karimii	Didelphimorphia	0.95 (0.00-1.89)	--
Tlacuatzin canescens	Didelphimorphia	1.03 (0.14-1.91)	--
Lepus europaeus	Lagomorpha	--	1.25 (0.24-2.07)
Oryctolagus cuniculus	Lagomorpha	0.95 (0.20-2.07)	0.95 (0.18-1.66)
Romerolagus diazi	Lagomorpha	0.96 (0.13-1.90)	--
Sylvilagus audubonii	Lagomorpha	1.36 (0.54-2.12)	1.20 (0.49-2.06)
Sylvilagus brasiliensis	Lagomorpha	--	1.67 (0.63-2.29)
Sylvilagus palustris	Lagomorpha	--	0.79 (0.11-1.56)
Tapirus terrestris	Perissodactyla	--	1.22 (0.54-1.81)
Bradypus variegatus	Pilosa	--	1.85 (1.15-2.81)
Cyclopes didactylus	Pilosa	2.14 (1.15-2.99)	1.66 (0.86-2.37)
Myrmecophaga tridactyla	Pilosa	1.36 (0.42-2.22)	--
Ateles geoffroyi	Primates	1.88 (1.00-2.61)	0.97 (0.16-1.69)
Callithrix jacchus	Primates	1.10 (0.24-1.84)	--
Cebus apella	Primates	--	0.91 (0.07-1.99)
Cebus capucinus	Primates	1.52 (0.30-2.63)	--
Erythrocebus patas	Primates	--	1.11 (0.33-2.16)
Macaca mulatta	Primates	--	1.10 (0.31-2.09)
Saguinus fuscicollis	Primates	0.92 (0.14-1.67)	--
Saguinus oedipus	Primates	1.67 (0.72-2.65)	1.04 (0.12-2.20)
Saimiri oerstedii	Primates	1.15 (0.12-2.30)	--
Saimiri sciureus	Primates	1.00 (0.38-1.68)	--

Agouti paca	Rodentia	--	2.45 (1.77-3.18)
Calomys callosus	Rodentia	2.32 (1.61-3.09)	2.15 (1.24-3.02)
Calomys expulsus	Rodentia	0.84 (0.09-1.69)	--
Cavia aperea	Rodentia	1.11 (0.52-2.03)	--
Dasyprocta leporina	Rodentia	0.87 (0.03-2.09)	1.46 (0.57-2.22)
Dasyprocta punctata	Rodentia	1.54 (0.57-2.82)	0.73 (0.07-1.41)
Dipodomys merriami	Rodentia	--	1.68 (0.86-2.41)
Dipodomys ordii	Rodentia	1.46 (0.59-2.22)	1.45 (0.63-2.43)
Galea spixii	Rodentia	--	2.34 (1.76-2.96)
Glaucomys volans	Rodentia	1.37 (0.64-2.07)	0.77 (0.01-1.68)
Holochilus brasiliensis	Rodentia	--	0.43 (0.07-0.89)
Hydrochoerus hydrochaeris	Rodentia	1.74 (0.63-2.59)	--
Kerodon rupestris	Rodentia	--	1.30 (0.50-2.01)
Liomys salvini	Rodentia	1.53 (0.73-2.53)	1.77 (1.02-2.45)
Myocastor coypus	Rodentia	0.78 (0.18-1.48)	0.79 (0.22-1.45)
Myoprocta acouchy	Rodentia	--	0.94 (0.03-1.96)
Myoprocta pratti	Rodentia	1.54 (0.52-2.29)	--
Neotoma mexicana	Rodentia	1.17 (0.29-1.78)	--
Neotomodon alstoni	Rodentia	1.11 (0.41-1.96)	--
Ochrotomys nuttalli	Rodentia	2.08 (1.38-2.73)	1.66 (1.22-2.33)
Oligoryzomys fornesi	Rodentia	1.21 (0.54-2.00)	--
Oligoryzomys fulvescens	Rodentia	1.63 (0.84-2.48)	1.52 (0.90-2.25)
Oligoryzomys nigripes	Rodentia	0.75 (0.13-1.36)	--
Oryzomys couesi	Rodentia	1.42 (0.72-2.24)	1.22 (0.39-2.00)

Oryzomys megacephalus	Rodentia	2.64 (1.72-3.29)	2.44 (1.90-3.09)
Oryzomys palustris	Rodentia	2.16 (1.40-2.83)	1.70 (1.13-2.35)
Oryzomys subflavus	Rodentia	--	1.40 (0.13-2.27)
Otonyctomys hatti	Rodentia	0.83 (0.12-1.74)	--
Peromyscus aztecus	Rodentia	0.91 (0.09-2.08)	--
Peromyscus gossypinus	Rodentia	2.11 (1.75-2.65)	1.62 (1.20-2.09)
Proechimys guairae	Rodentia	1.23 (0.52-1.92)	1.69 (0.82-2.30)
Reithrodontomys fulvescens	Rodentia	1.12 (0.26-1.94)	1.01 (0.29-1.65)
Reithrodontomys humulis	Rodentia	1.79 (1.10-2.74)	1.43 (0.74-2.38)
Reithrodontomys megalotis	Rodentia	1.39 (0.65-2.03)	0.89 (0.43-1.69)
Rhipidomys mastacalis	Rodentia	1.39 (0.81-2.24)	--
Sciurus granatensis	Rodentia	1.43 (0.60-2.40)	--
Sciurus variegatoides	Rodentia	--	0.93 (0.14-1.76)
Sciurus vulgaris	Rodentia	1.48 (0.46-2.25)	--
Scotinomys teguina	Rodentia	0.94 (0.29-1.92)	--
Spermophilus variegatus	Rodentia	0.91 (0.16-1.84)	--
Tylomys nudicaudus	Rodentia	1.58 (0.50-2.29)	0.84 (0.12-1.67)
Zygodontomys brevicauda	Rodentia	1.87 (1.16-2.68)	1.07 (0.27-1.84)
Cryptotis parva	Soricomorpha	1.40 (0.50-2.51)	0.99 (0.31-1.59)
Scalopus aquaticus	Soricomorpha	1.78 (1.19-2.46)	1.64 (0.84-2.25)

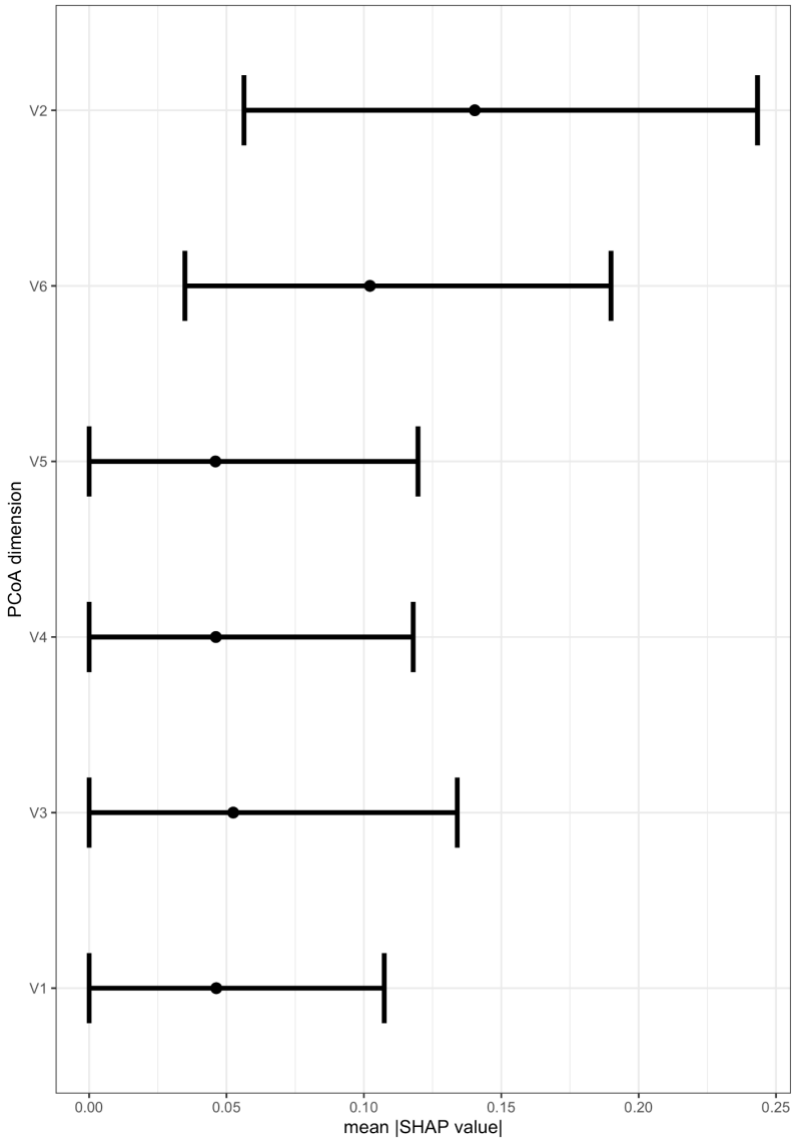
802

Supplementary Table 3. Complementary evidence of predicted hosts. Complementary evidence is defined as published literature that suggests animals may be exposed to and/or infected by zoonotic *Leishmania*, but was not robust enough to include the animal in the analysis as a known host. P = predicted by our model; K = known host (the animal has been naturally infected by a *Leishmania* species known to cause leishmaniasis in humans).

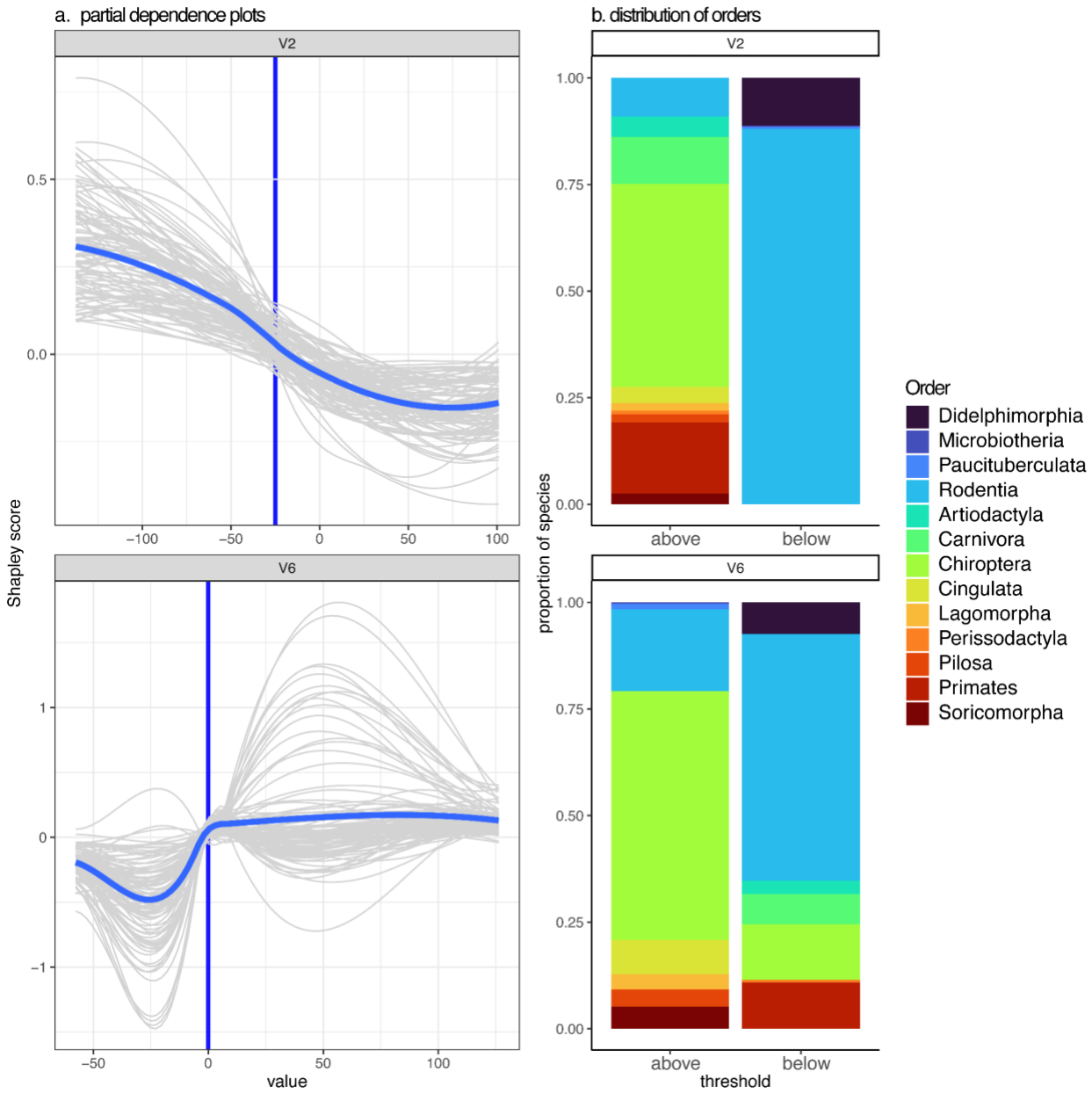
Binomial	Common name	*L. (V)* host	*L. (L) host*	complementary evidence	Ref
Calomys callosus	Large vesper mouse	P	P	experimental infection	Rosa et al. 2022
Galea spixii	Spix's yellow-toothed cavy	--	P		Barbosa et al. 2008
Tadarida brasiliensis	Mexican free-tailed bat	P	P		Berzunza-Crus et al. 2015
Didelphis virginiana	Virginia opossum	P	P	experimental infection; significant overlap w/ *L. mexicana* vectors	Carreria et al. 2017; Stephens et al. 2009, 2016
Vulpes vulpes	Red fox	--	P	*L. infantum* host in Europe	Karayiannis et al. 2015
Lepus europaeus	European hare	--	P		Tsokana et al. 2016
Oryctolagus cuniculus	European rabbit	P	P		García et al. 2014
Sus scrofa	Wild boar	--	P	leishmania infection but not confirmed as zoonotic	
Canis latrans	Coyote	--	P		Rosypal et al. 2013
Lontra longicaudis	Neotropical otter	--	P		Azami-Conesa et al. 2021
Procyon cancrivorus	Crab-eating racoon	--	P		Richini-Pereira et al. 2014

Monodelphis domestica	Gray short-tailed opossum	P	P		Carreria et al. 2017
Euphractus sexcinctus	Six-banded armadillo	P	P		Barbosa et al. 2008
Cavia aperea	Brazilian guinea pig	P	K		Caldart et al. 2021
Hydrochoerus hydrochaeris	Capybara	P	K		Caldart et al. 2021
Sciurus granatensis	Red-tailed squirrel	P	--		Júnior et al. 1992
Caluromys derbianus	Derby's wooly opossum	P	P	leishmania infection but not confirmed as zoonotic; vector feeds on animal	Carreria et al. 2017
Procyon lotor	Common racoon	P	P	significant overlap w/ L. mexicana vectors	Stephens et al. 2009, 2016
ma villosum	Hairy big-eyed bat	--	P		
Eptesicus brasiliensis	Brazilian brown bat	--	P		
Saccopteryx bilineata	Greater sac-winged bat	--	P		
Uroderma bilobatum	Tent-making bat	P	P		
Agouti paca	Lowland paca	K	P		
Neotoma mexicana	Mexican woodrat	P	--		
Oligoryzomys fulvescens	Fulvous pygmy rice rat	P	P		
Oryzomys nigripes		P	--		
Oryzomys couesi	Coues's rice rat	P	P		

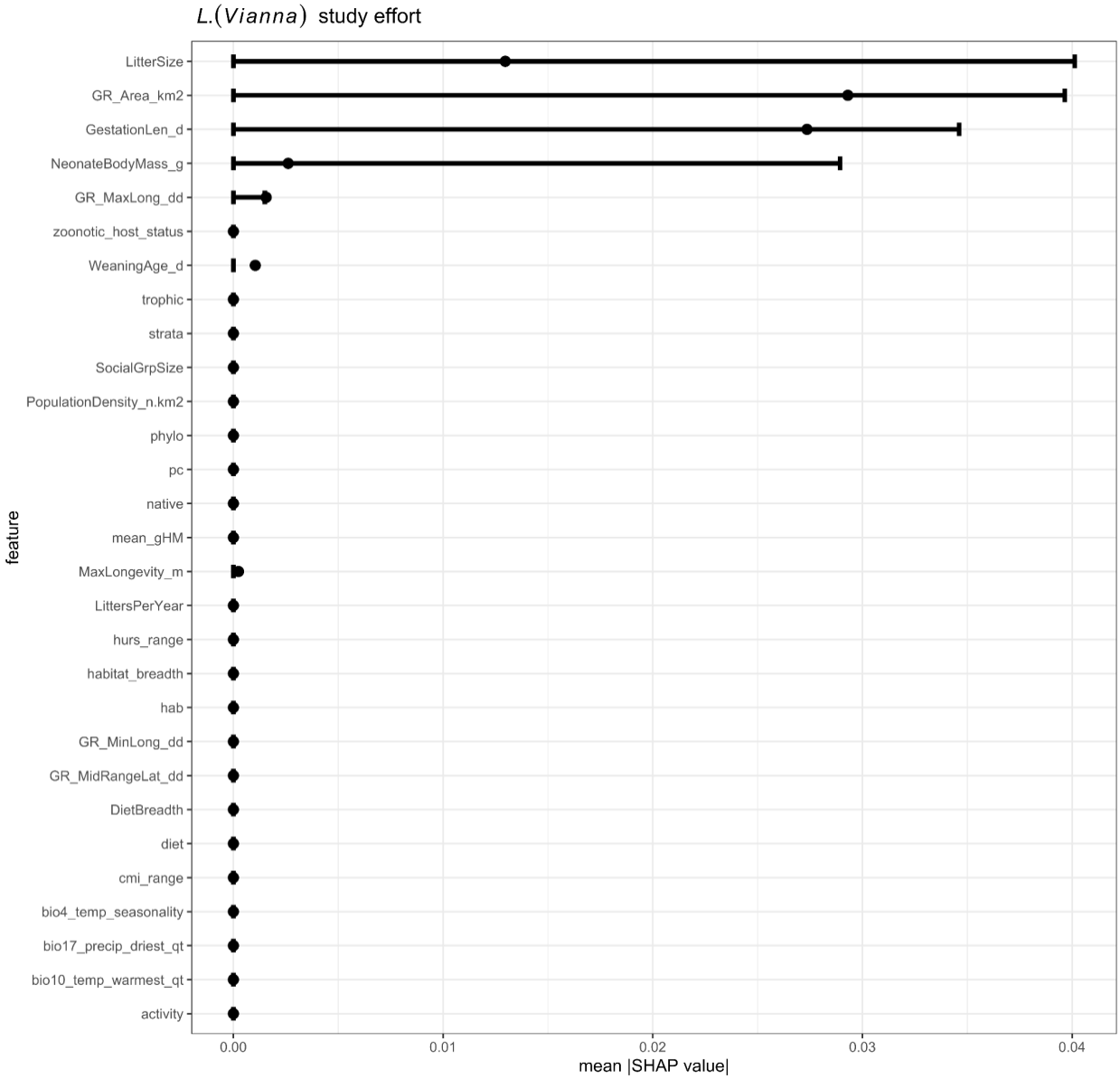
Reithrodontomys fulvescens	Fulvous harvest mouse	P	P		
Potos flavus	Kinkajou	P	K		
Nasua narica	White-nosed coati	P	P		
Leopardus wiedii	Margay	P	P		
Dasyprocta leporina	Red-rumped agouti	P	P	vector feeds on animal	dos Santos et al. 2018



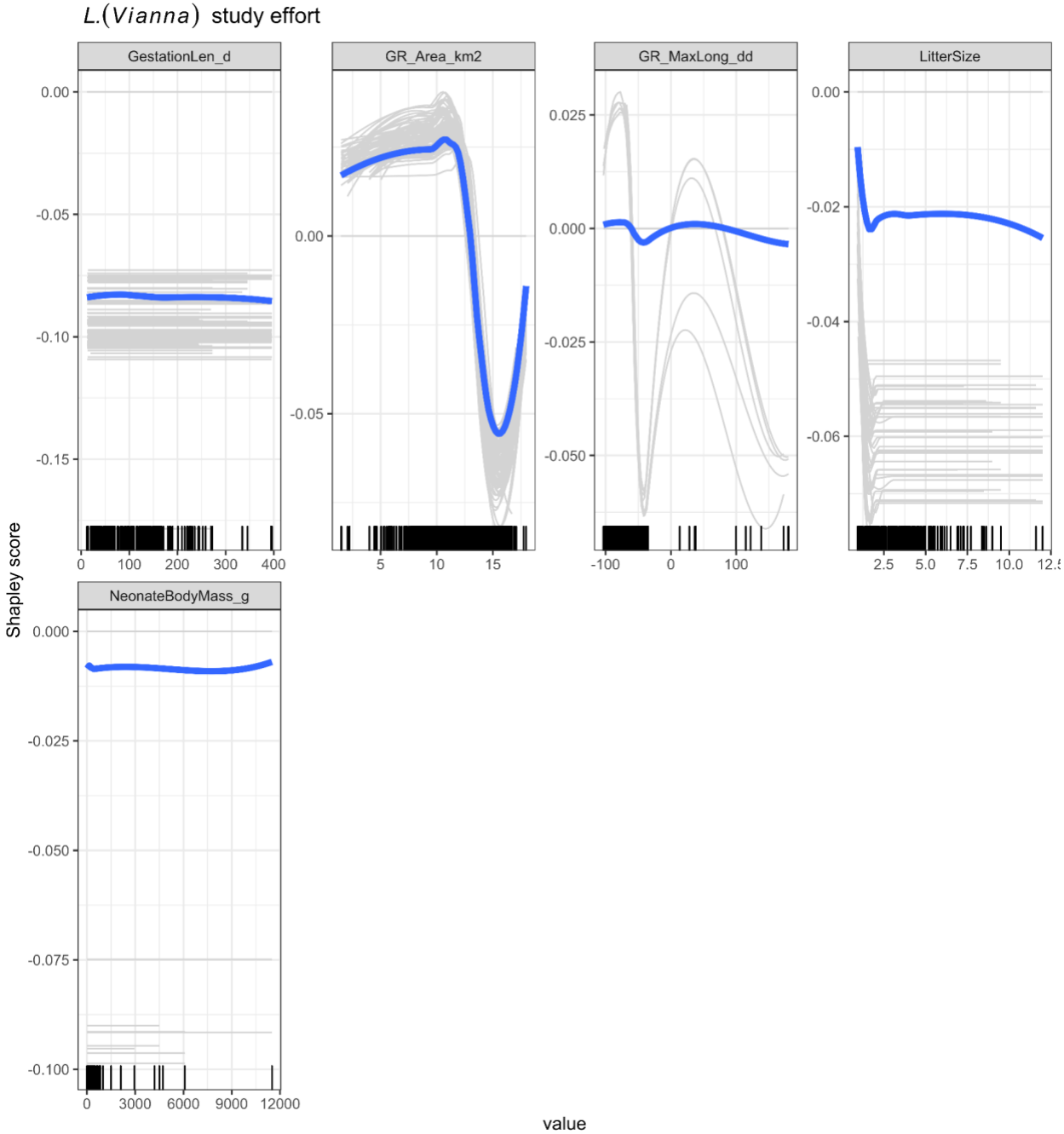
Supplementary figure 1. Importance of each PCoA phylogenetic dimension in predicting *L. (Vianna)* host status. Points are the mean Shapley importance for the PCoA dimension across all mammals, bars represent the 0.05-0.95 percentiles.



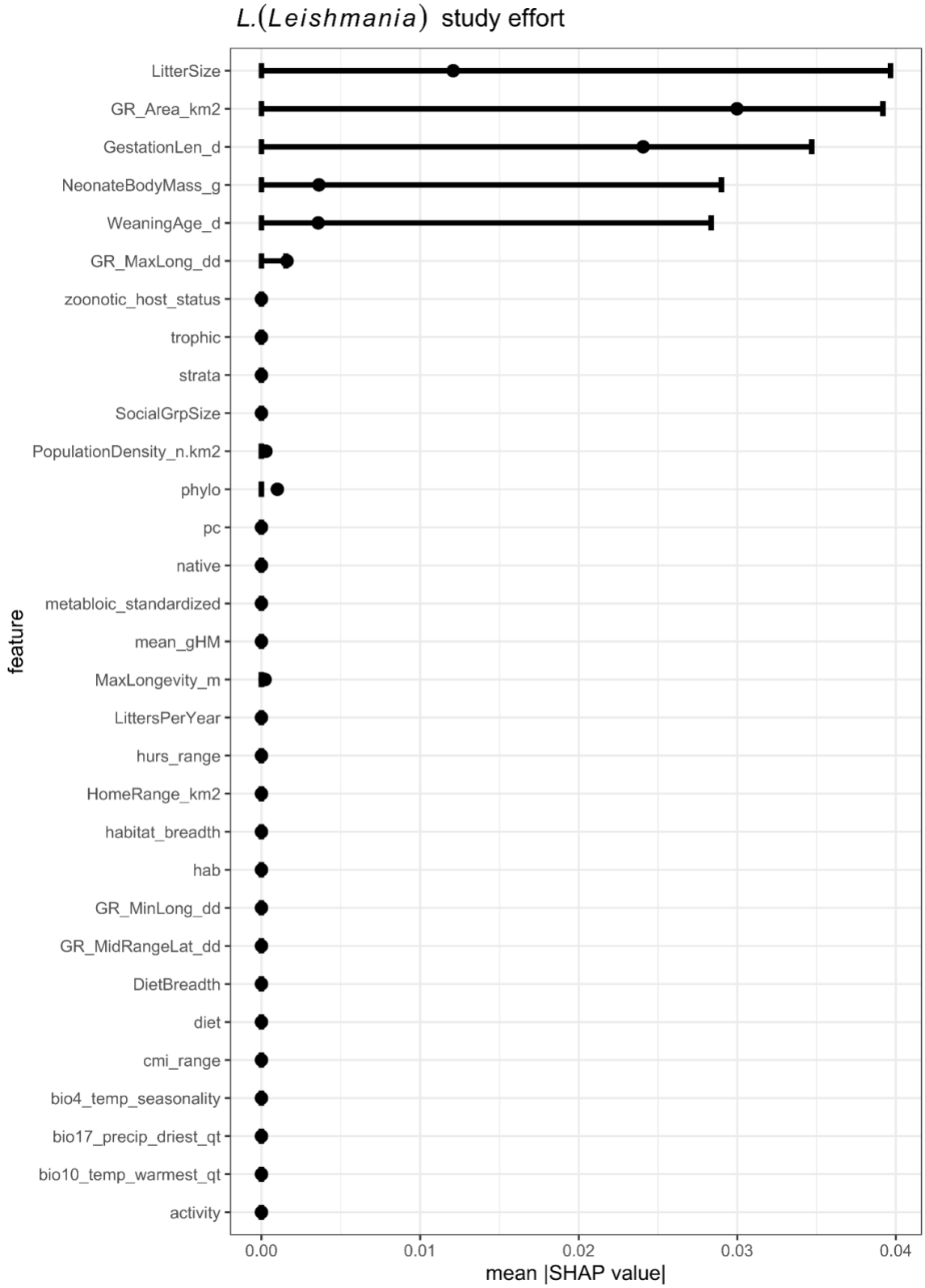
Supplementary figure 2. (a) Shapley partial dependence plots showing the effect of PCoA dimensions on *L.* (*Vianna*) host status. Blue non-linear lines represent the average effect across model iterations , while grey lines show each individual model iteration (model fit with 70% of data). Dark blue vertical line represents a threshold value, where the dimension switches from having a negative effect on host status to a positive effect on host status. Only dimensions with importance scores > 0 for > 95% of model iterations are shown. **(b)** distribution of hosts within each order with PCoA dimension values above or below the threshold.



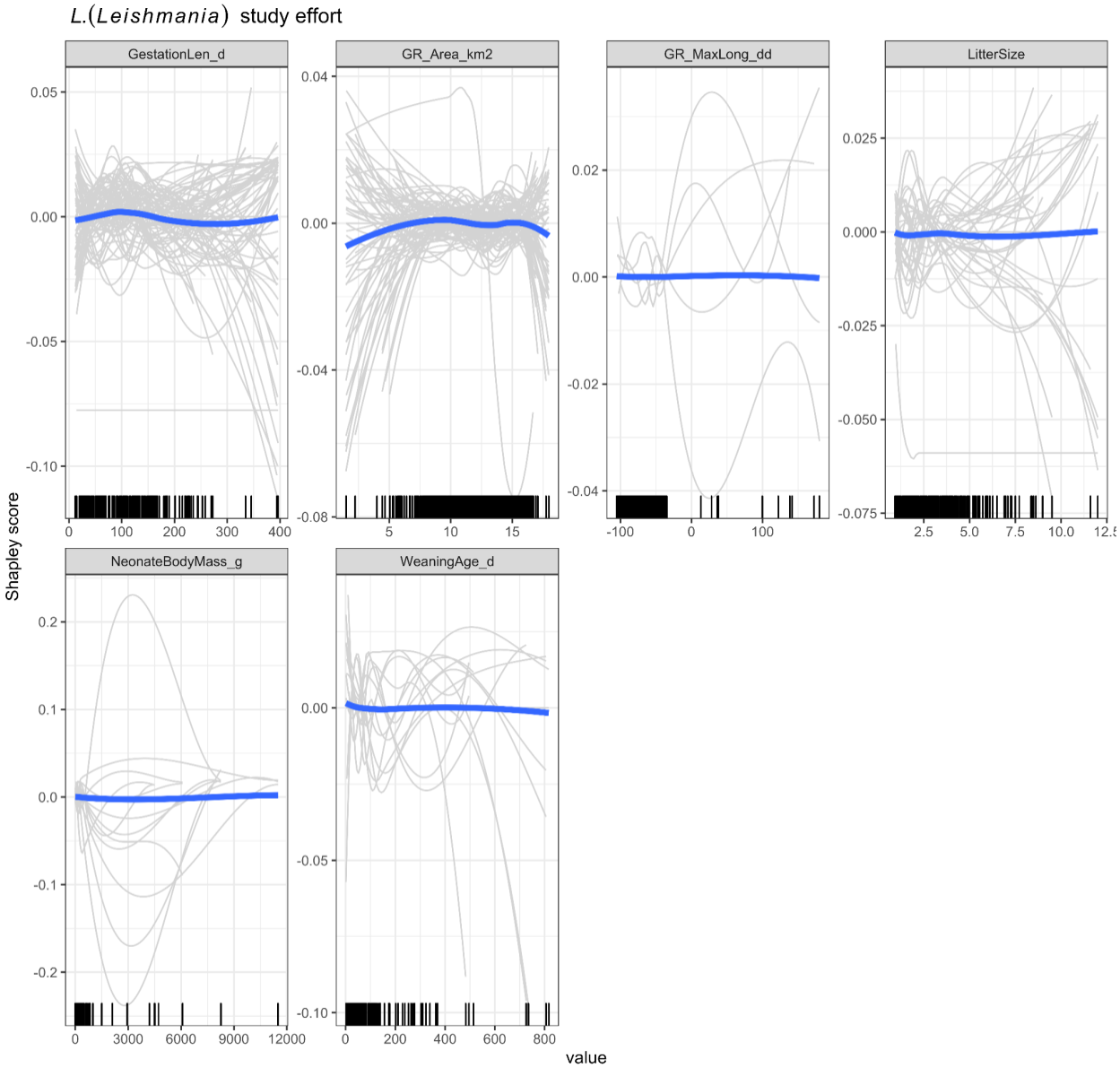
Supplementary figure 3. Features inconsistently contribute to predictions for study effort of mammals included in the *L. (Vianna)* model. Global feature contribution for traits predicting study effort of mammals included in the *L. (Vianna)* model Points are the absolute value of the mean Shapley importance (mean |SHAP value|) for the trait across all mammals, bars represent the absolute values of the 0.05-0.95 percentiles. No trait consistently contributes to model predictions across all model bootstrapping iterations (i.e., there is no trait where the 0.05 percentile is greater than 0). Further, traits with average values > 0 do not correspond to traits that consistently predict *L. (Vianna)* host status.



Supplementary figure 4. Response of study effort predictions to mammal traits are inconsistent and do not mirror functional forms of response of *L. (Vianna)* host status to mammal traits. Shapley partial dependence plots showing the effect of mammal traits on study effort of mammals included in the *L. (Vianna)* model. Blue non-linear lines represent the average effect across model iterations, while grey lines show each individual model iteration (model fit with 70% of data). Functional form varies greatly across model iterations. Only traits where the absolute value of average contribution was > 0 are shown. Rug plots on the x-axis indicate the distribution of the data.



Supplementary figure 5. Features inconsistently contribute to predictions for study effort of mammals included in the *L. (Leishmania)* model. Global feature contribution for traits predicting study effort of mammals included in the *L. (Leishmania)* model. Points are the absolute value of the mean Shapley importance (mean |SHAP value|) for the trait across all mammals, bars represent the absolute values of the 0.05-0.95 percentiles. No trait consistently contributes to model predictions across all model bootstrapping iterations (i.e., there is no trait where the 0.05 percentile is greater than 0). Further, traits with average values > 0 do not correspond to traits that consistently predict *L. (Leishmania)* host status.



Supplementary figure 6. Response of study effort predictions to mammal traits are inconsistent and do not mirror functional forms of response of *L. (Leishmania)* host status to mammal traits. Shapley partial dependence plots showing the effect of mammal traits on study effort of mammals included in the *L. (Leishmania)* model. Blue non-linear lines represent the average effect across model iterations, while grey lines show each individual model iteration (model fit with 70% of data). Functional form varies greatly across model iterations. Only traits where the absolute value of average contribution was > 0 are shown. Rug plots on the x-axis indicate the distribution of the data.