Data-Driven Predictions of Potential Leishmania Vectors in the Americas

Gowri M. Vadmal[1*], Caroline K. Glidden[1], Barbara A. Han[2], Bruno M. Carvalho[3], Adrian A. Castellanos[2], Erin A. Mordecai[1]

[1]Department of Biology, Stanford University, Stanford, California, USA
[2]Cary Institute of Ecosystem Studies, Millbrook, New York, USA
[3]Climate and Health Program, Barcelona Institute for Global Health, Barcelona, Spain

*gvadmal@stanford.edu

Financial Disclosure:
This project was supported by the Stanford King Center for Global Development (GMV, EAM), the National Science Foundation (DEB-2011147 with the Fogarty International Center, CKG, EAM; DEB-1717282, BAH, AAC), the National Institutes of Health (R35GM133439, R01AI168097, and R01AI102918 EAM; 5U01AI15180703, BAH, AAC), the Severo Ochoa Center of Excellence Grant (BMC), Spanish Ministry of Science and Innovation & Spanish State Research Agency (CEX2018- 000806-S, BMC). EAM was additionally supported by seed grants from the Stanford Woods Institute for the Environment, King Center on Global Development, Center for Innovation in Global Health, and the Terman Award. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract:

The incidence of vector-borne diseases is rising as deforestation, climate change, and globalization bring humans in contact with arthropods that can transmit pathogens. In particular, incidence of American Cutaneous Leishmaniasis (ACL), a disease caused by parasites transmitted by sandflies, is increasing as previously intact habitats are cleared for agriculture and urban areas, potentially bringing people into contact with vectors and reservoir hosts. Previous evidence has identified dozens of sandfly species that have been infected with and/or transmit *Leishmania* parasites. However, there is an incomplete understanding of which sandfly species transmit the parasite, complicating efforts to limit disease spread. Here, we apply machine learning models (boosted regression trees) to leverage biological and geographical traits of known sandfly vectors to predict potential vectors. Additionally, we generate trait profiles of confirmed vectors and identify important factors in transmission. Our model performed well with an average out of sample accuracy of 86%. The models predict that synanthropic sandflies living in areas with greater canopy height, less human modification, and within an optimal range of rainfall are more likely to be *Leishmania* vectors. We also observed that generalist sandflies that are able to inhabit many different ecoregions are more likely to transmit the parasites. Our results suggest that *Psychodopygus amazonensis and Nyssomia antunesi* are unidentified potential vectors, and should be the focus of sampling and research efforts. Overall, we found that our machine learning approach provides valuable information for *Leishmania* surveillance and management in an otherwise complex and data sparse system.

Author Summary:

American Cutaneous Leishmaniasis (ACL) is a neglected disease caused by sandfly-transmitted parasites in the Americas. There is an incomplete understanding of which sandfly species transmit the parasite, complicating efforts to limit parasite transmission and consequently, disease burden. In this study, the authors created a database of sandfly traits, then used predictive models to determine important factors in parasite transmission and how different climate and environmental variables predict which vectors can transmit the parasites that cause ACL. The models suggest that transmission occurs at the interface between domestic habitats and well-preserved forests. The authors also generate predictions of which sandflies might be transmitting the parasite that are not known vectors at the time, specifically *Psychodopygus amazonensis and Nyssomia antunesi*. This new knowledge can lead to a better understanding of the system of transmission, and can point to possible hotspots of risk. The analysis can also help direct researchers to areas of interest for sampling studies, as well as specific sandflies on which to focus effort.

Introduction:

American cutaneous leishmaniasis (ACL) is a neglected tropical disease caused by parasites in the genus *Leishmania*, and transmitted by sandflies of the subfamily Phlebotominae [1, 2]. The World Health Organization estimates that worldwide there are approximately 1 to 2 million new cases of leishmaniasis each year [3], with 700,000 to 1 million of those cases identified as cutaneous leishmaniasis cases [4]. ACL cases occur across the Americas, with hotspots in northeastern and southeastern states in Mexico, northern Nicaragua, Costa Rica, Brazil, Peru, and at the convergence of the borders of Brazil, Peru, and Bolivia [5, 6]. In some regions, incidence of ACL is increasing among farmers, loggers, hunters, and others working at the forest-human interface. Additionally, although primarily a tropical and subtropical disease [2], cases have also been more recently reported in the southern United States [3, 7], making it an important emerging health problem in temperate regions.

Similar to other disease-causing parasites, *Leishmania* parasites are found in wild and domestic reservoir hosts, which are located across the Americas, and are picked up by sandfly females during their blood meal before laying eggs [2, 8, 9]. In the female sandfly gut, the parasite amastigotes develop into promastigotes, which migrate to the salivary glands and spread to other mammals or to humans during subsequent blood meals [2, 8, 9] (Fig 1). Once a human is infected, the incubation period typically lasts around one to ten weeks (but can last many years) in which promastigotes invade local tissues and transform into amastigotes, entering macrophages through phagocytosis [2, 8]. Clinical symptoms include lesions, rashes, open sores, ulcers, and small bumps covering the skin, which can lead to deformation with possible recurrences [2]. In some cases, ACL can evolve into diffuse or disseminated leishmaniasis, and rarely, into mucocutaneous leishmaniasis which can lead to severe facial mutilations and extensive disfiguring of the face, soft palate, pharynx, and larynx [2, 10, 11]. Overall, the disease is extremely painful and at times, severely debilitating.

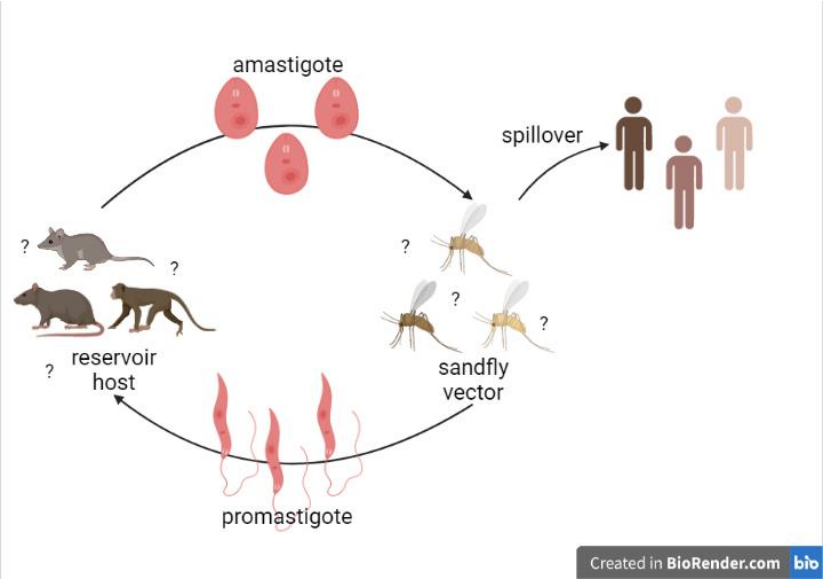


Fig 1: ACL life cycle. Female sandfly vectors pick up *Leishmania* parasites during their blood meal from reservoir hosts. Spillover events occur when a sandfly with *Leishmania* parasites in its salivary glands takes a blood meal from a human, and infects the human with the parasite. Identity of all reservoir hosts and sandfly vectors are still unknown, which makes it hard to model and prevent transmission. Created with BioRender.com.

There is no recognized oral treatment of ACL, and current antimonial treatments are painful, potentially dangerous and expensive [2, 3, 8, 11, 12]. Therefore, ACL is best managed through ecological interventions like controlling vectors and preventing transmission [13].

Climate change, deforestation, travel, and natural disasters are correlated with the spread of the parasite and leishmaniasis [2], yet it is hard to mechanistically predict the effect of global change on leishmaniasis incidence and distribution due to gaps in understanding of the parasite transmission cycle. Specifically, the full suite of reservoir hosts and confirmed vectors has not been fully characterized, making it difficult to model transmission under global change. For example, the incubation period of the parasite in the vector can be longer at lower temperatures and *shorter* at higher temperatures, ultimately impacting the number of hosts one sandfly can infect [2, 14, 15]. However, temperature *responses* likely vary by sandfly species [14, 15, 16], and thus the effect of climate warming on *Leishmania* transmission cannot be predicted precisely as the full range of sandflies that could transmit the parasites have yet to be described. Additionally, since different

sandfly species have different habitat and biting preferences, sandfly species differ in their propensity to respond to land use change and contact infectious reservoir hosts and humans, thereby contributing to the human ACL burden. These are key uncertainties that need to be resolved for a more complete understanding of the transmission cycle, which will help to inform these mechanisms and identify vector and disease management opportunities.

Parasite-vector interactions can be divided into restrictive (sandflies that demonstrate specificity for the *Leishmania* species they can transmit) and permissive (sandflies that show non-specific interactions) groupings under laboratory conditions [17]. While restrictive pairs are well studied and linked together using molecular analysis methods, not much is known about permissive interactions, and they may be under-represented in the known list of sandfly vectors [17]. Thus, the current classification criteria might overlook some sandfly vectors that carry several different species of *Leishmania*. There have been efforts to use modeling approaches to identify environmental factors and vectors at the local level in Colombia [18, 19] and in the Middle East [20], but no such models have been built using data from across the Americas.

In recent years, new machine learning methods have been used to identify potential reservoir hosts, vectors, and important factors in transmission of other vector-borne pathogens [21, 22]. In order to fill in gaps in our understanding of *Leishmania* transmission, we use a similar approach to model the relationships between sandfly biology and vector status. From these models, we generate a list of sandflies predicted to transmit *Leishmania* spp. causing ACL, which we recommend should be empirically tested for vector competence and added to surveillance efforts.

Methods:

Data collection

We first compiled a database on Phlebotomine sandfly vector status as well as behavioral, morphological, taxonomic, and ecological traits that could be used to delineate vectors from non-vectors. Past vector identification has typically relied on the following criteria stated by Killick-Kendrick: (i) epidemiological observation that the sandfly is anthropophilic, (ii) proof that the fly feeds regularly on a relevant reservoir host,

(iii) repeated isolation and identification of the same species of *Leishmania* spp. promastigotes infecting the humans in the surrounding area, (iv) evidence that the fly supports the complete development of the parasite, (v) and experimental evidence the sandfly can transmit the parasite through blood meal bite [17, 23, 24]. We defined a known vector to be a sandfly that is incriminated as a *Leishmania* vector by these criteria, which are generally considered to be the gold standard for identification [23]. We directly used vector status from Akhoundi et al. [24], which used the above criteria for identification. Molecular methods have been discussed as possible evidence, but are not sufficient to incriminate a sandfly species as a proven vector [24]. Here, we will use the term 'vector' to represent the sandflies that are confirmed through the five criteria stated by Killick-Kendrick [23]. We note that our definition of vector is specific to vectors of *Leishmania* spp. that can infect humans. We defined a 'potential vector' to be a sandfly that has been observed to carry *Leishmania* parasites in the wild via molecular diagnosis or dissection but not confirmed to transmit it to humans; it is important to note that we did not consider sandflies that have been experimentally infected with a *Leishmania* parasite to be a potential vector, and that we did not apply positive labels to potential vectors in our primary analysis. Our dataset comprised 512 documented sandfly species across the Americas (sample size n = 512). Thirty-seven of these 512 are confirmed at the time of analysis as vectors of one or more species of *Leishmania* causing ACL, and 35 additional species are potential vectors [10, 24, 25-28].

From the literature we collected thirteen female morphological traits [29-32] including but not limited to wing length, width, and number of teeth (Table S1), as well as vector and infection status [24, 26, 27, 30, 33-37]. We included genus as a variable to include a measure of taxonomic relatedness in the model. Most of the sandfly morphological traits and biting behavior were taken from Young & Duncan's 1994 book [30], with additional biting behavior data taken from sandfly surveys using Disney and Shannon traps [25, 27, 37-78]. We paired occurrence points from the Global Biodiversity Information Facility (GBIF) [79, 80] and published sandfly sampling studies [37-78] with GIS data from Google Earth Engine to describe biogeographical features of sandfly habitat such as temperature, wind speed, and canopy height (Table S1). For each sampling study from the literature, we only included sandflies that made up more than 1% of the sampled species to account for possible misidentification or outliers. Habitat features were calculated for a sandfly species if there were at least 4 occurrence points for the species.

We used the Google Earth Engine Python API in Jupyter Notebook to get environmental and geographical traits averaged over each species' distribution, using the smallest spatial resolution possible as sandfly dispersal is very limited [2, 81, 82] (Table S1). Based on known sandfly-parasite interactions, we expect important traits to include temperature, forest integrity, and other environmental change variables [2, 4, 14, 15, 16]. We used Copernicus Climate Change Service's ERA5 datasets of biogeographical data to get the mean monthly temperature, temperature range, mean monthly total rainfall, mean monthly wind speed, and mean elevation [83]. All data from Google Earth Engine datasets was from 2009 to 2019. We used NASA's Terra Vegetation dataset to get an enhanced vegetation index [84] and the Copernicus Global Land Cover dataset for tree, shrub, urban, grass, and water cover [85]. We used NOAA's ETOPO1 dataset for elevation [86] and NASA and JPL's Global Forest Canopy Height for canopy height [87]. We defined a species' ecoregion breadth as the number of different ecoregions it inhabited in the RESOLVE Biodiversity and Wildlife Solutions dataset (i.e., how many unique ecoregions the occurrence points mapped to) [88]. A species' presence in a biome was a binary trait for 10 different biomes [88]. The global human modification trait was the cumulative measure of human modification of terrestrial lands globally at a 1 square-kilometer resolution in the Conservation Science Partners gHM dataset [89]. We used the Forest Landscape Integrity Index for the average forest integrity of an occurrence point, determined by degree of anthropogenic modification [90]. We used temperature variance as an indicator of seasonality, and found the average variance across the years with each month as an observation time point [83]. We used the RISmed package in R to quantify citation counts of each sandfly in the PubMed database. Overall, we collected 12 morphological traits and 25 different ecological and biogeographical traits.

Following (Evans et al. 2017, Han et al. 2019, Fischhoff et al. 2021, Han et al. 2015), variables with less than 10 percent coverage and a correlation factor greater than 0.7 with other variables were not included in the final analysis to avoid overfitting and misestimating the importance of highly correlated variables. The cutoff removes traits for which less than 10 percent of sandflies have data, to filter out variables with low coverage to simplify the models a bit. In theory, a data coverage cutoff is not necessary because of the way boosted regression treats missing data - it views the missingness as a 'common value' to group on. If one were to include all of the data regardless of coverage, low coverage variables that also have "low information" (either

due to low coverage or due to the feature not being consequential for prediction) have low relative importance scores. As such, we used the 10% cutoff to simplify the model by removing variables that are likely to not contribute to model performance and predictions due to "low information". In order to avoid cyclic analysis, we did not include the trait for biting humans in our analysis, as vectors are already defined to be anthropophilic. Traits with skewed distributions were normalized to avoid skewing our analysis with outlying and potentially influential data while training our models [92]. We used one hot encoding, which converts each categorical variable to a new binary variable with either a 0 or 1, to transform categorical variables into binary variables (eg. genus, shape of maxillary tip, or structure of hypopharyngeal teeth) [21].

Data analysis

We used extreme boosting through the XGBoost library in Python to fit a logistic classifier boosted regression tree model (BRT). 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 adding them to a strong classifier (i.e., boosting). 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 [21, 22, 91, 92]. XGBoost also allows the use of regularization parameters to prevent overfitting models to small, unbalanced datasets. XGBoost additionally handles unbalanced data well by weighting positive labels, an advantage when analyzing our data set with relatively few known vectors and sparse feature coverage.

We fit two predictive models for the general *Leishmania* genus as there was not enough data to accurately make separate models for each species of *Leishmania*. For the primary model, only the confirmed vectors were used as positive labels (0: not a confirmed vector, 1: confirmed vector). We include an additional analysis in the supplementary materials that includes both confirmed and potential vectors as positive labels (0: no evidence of the sandfly carrying a *Leishmania* parasite, 1: confirmed vector or observed to carry a *Leishmania* parasite). As such, the secondary analysis indicates the probability a sandfly may be naturally infected with *Leishmania,* but is not necessarily infectious upon infection.

For training and tuning analysis, the data was stratified and split into 80% training and 20% testing sets such that each set had an equal proportion of positive labels. To tune the hyperparameters for our XGBoost model, we used the hyperopt library in Python, which uses Bayesian optimization to find the best performing parameters for the model. We define a search space to include parameters dealing with regularization, depth, and learning rate of the regression trees, then run the optimization algorithm to find the best performing parameters. To ensure the model was generalizable, we used a 3-fold nested cross validation process for parameter tuning, where the training dataset was divided into three folds or subsets. In cross validation, for 3 iterations, a combination of two of those folds was used as the training set, while the remaining fold was used for validation to optimize parameter estimation. The nested cross-validation approach provides conservative estimates of model performance when analyzing small datasets [93]. The training results are averaged over the folds to get the performance score of the model. Due to data sparsity, we opted for 3-fold nested cross validation rather than 10-fold [21]. We used the 10 best performing parameter sets (minimum log-loss) in our BRT models.

Since results of boosted regression tree models are often dependent on test/train splits [24], we used the 10 best performing sets of parameters, and 10 random test train splits, to train 100 total models using the XGBoost library in Python. We evaluated model performance across the 100 model iterations using the aggregate median of the Area Under the Receiver Operator Curve (AUC). Each of the models was used to generate a predicted probability for the sandflies by applying the trained models to the whole dataset of sandflies and their traits. The predicted probability ranged from 0 to 1, and we used the aggregate median generated across the 100 models to assess potential vector status. We define variable importance to be the number of times a variable is selected for splitting a regression tree, weighted by the improvement to the model as a result of that node. Importance was evaluated on a scale of 0 to 1, with higher numbers signifying that the variable had a higher impact on model training, and all the individual importances summing to 1. For the variables we converted from categorical to binary, the relative importance of each binary trait was summed to represent the importance for the overall categorical variable.

Our secondary model, which was trained and fit identically to the primary model, used both confirmed and potential vectors as positive labels. More information on the secondary model can be found in the supplementary materials.

To determine whether we were identifying traits of vectors and not only traits of well-studied sandflies (i.e., our model was not biased by study effort), we ran a citation prediction model. Using the same hyperparameter tuning technique, a gradient boosted regressor model, and citation count as the target variable, we generated predictions of citations and a trait profile of well-studied species. We then compared the trait profiles between our vector model and citation count model to determine bias due to study effort. Next, to test whether our model was overfitting and fitting spurious correlations in the data, we performed target shuffling for 50 iterations (i.e., randomly shuffled the response variable, vector status, for 50 model iterations) and got the average performance score. Target shuffling is a way to test the statistical accuracy of a model, and avoid identifying false positives through false patterns in the data [94]. The model is considered overfit if it identifies shuffled labels with a greater accuracy than with a coin flip (i.e., AUC $\leq$ 0.5).

Results:

Our models used data on known vectors to identify which sandfly species are the most likely to carry and transmit *Leishmania* parasites causing ACL. We trained our ensemble of boosted regression tree (BRT) models on the trait profile of the confirmed vectors, and predicted which species might be potential vectors by leveraging trait similarities among species. The models achieved a high aggregate median out-of-sample AUC of 0.86 with a standard error of 0.008 across the 100 model iterations (Fig S1); therefore, on average, our models classified 86% of our observations correctly.

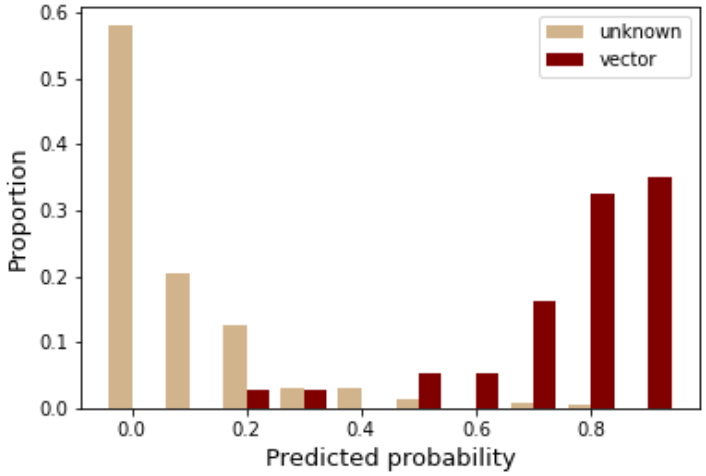


Fig 2: The model accurately classifies known vectors and identifies relatively few species of unknown status as likely vectors. A distribution of predicted probabilities of sandfly species separated by vector status, and scaled by percentage. Red bars indicate the proportion of confirmed vectors that were predicted at that probability, while beige bars indicate the proportion of sandfly species not previously identified as vectors that were predicted at that probability.

For each sandfly species, we generated an aggregate median predicted probability score of how likely it is to be a vector of ACL, and the percentile rank of that possibility. We consider sandfly species to be potential vectors if our models assigned them a predicted probability over 0.5 on a scale of 0 to 1, where 1 indicates that the species has a highest probability of being a vector. The models assigned 35 of 37 confirmed vectors with a median predicted probability above 0.5 and 13 of 475 sandfly species of unknown vector status with a median probability above 0.5 (Tables 1 and S2 and Fig 2). Two confirmed vectors (*Pintomyia youngi*, *Pintomyia ovallesi*) had lower probability scores, which could be due to sparsity of data for those species, thereby limiting our models from predicting species that otherwise might have been vectors. All of the 13 unknown sandflies predicted above a 0.5 probability were above the 90th percentile, and 8 of those 13 were potential vectors that have been observed to carry *Leishmania* parasites but have not yet been confirmed as vectors (Table 1). The full list of sandflies and their predicted probabilities, percentile rank, and status can be found in supplementary materials, and a map of confirmed and predicted vector species occurrence points appears in Fig 3. For our secondary model, which was trained on both proven and potential vectors, we generated the same types of

predictions, resulting in a mean AUC of 0.86, and many of the same top predicted sandflies (Table S4, Fig S6 and S7).

Table 1: The median predicted probability, standard deviation, and percentile for sandfly species of unknown vector status with greater than 0.5 predicted probability of being a vector. The infection status column indicates whether the sandfly is a potential vector (has been observed carrying the parasite, but not confirmed as transmitting it to humans; 'potential'), or that the sandfly has not yet been found infected with *Leishmania* in the wild ('unknown').

species	probability	std	percentile	infection status
Psychodopygus amazonensis	0.868	0.122	0.969	potential
Nyssomyia antunesi	0.852	0.122	0.963	potential
Psychodopygus claustrei	0.838	0.109	0.957	unknown
Psychodopygus guyanensis	0.748	0.133	0.938	unknown
Pintomyia (Pintomyia) pessoai	0.739	0.199	0.936	potential
Psathyromyia (Psathyromyia) bigeniculata	0.725	0.15	0.93	unknown
Trichophoromyia auraensis	0.702	0.195	0.928	potential
Psychodopygus chagasi	0.592	0.205	0.922	unknown
Trichophoromyia castanheirai	0.584	0.213	0.92	unknown
Sciopemyia sordellii	0.578	0.186	0.918	potential
Psathyromyia (Psathyromyia) lanei	0.556	0.182	0.916	unknown
Evandromyia (Evandromyia) infraspinosa	0.541	0.207	0.914	unknown
Warileya rotundipennis	0.511	0.217	0.908	unknown

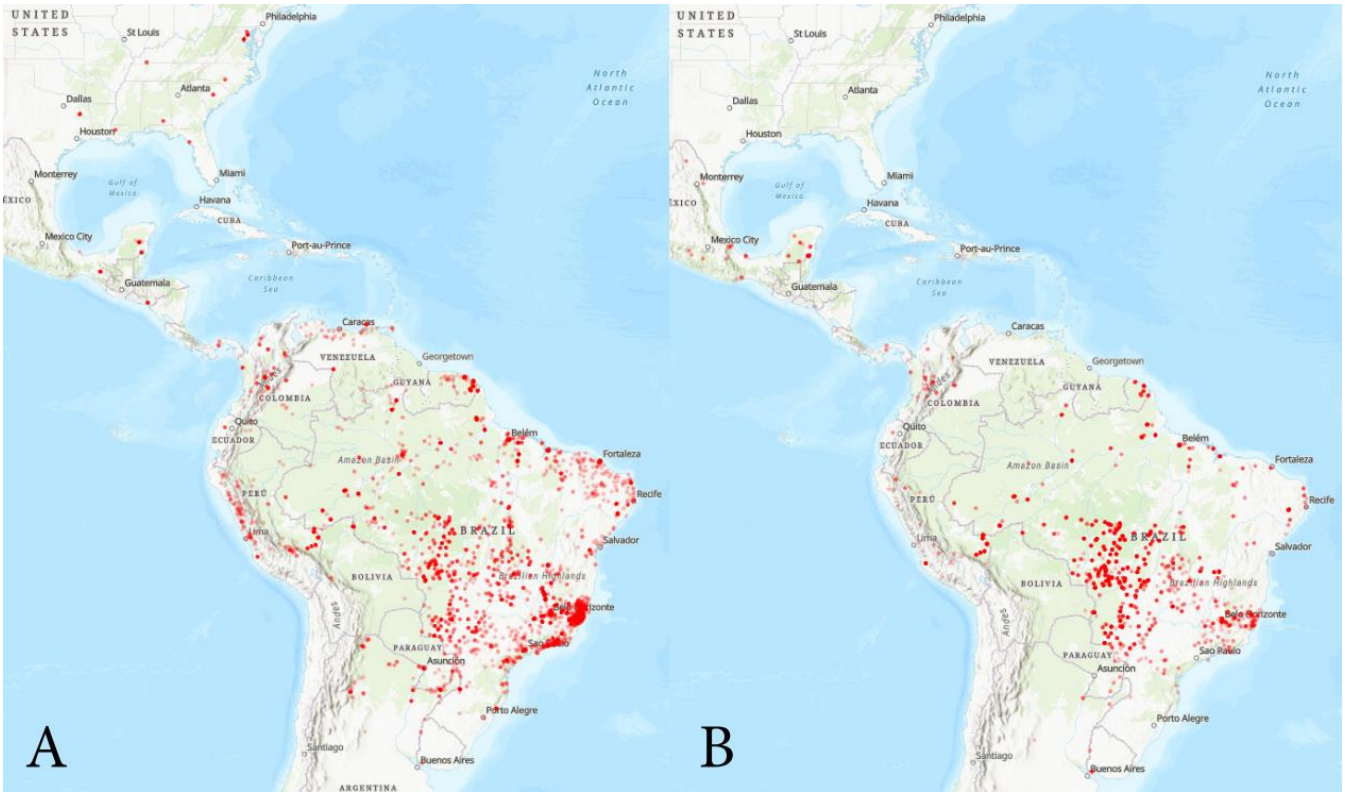


Fig 3: Confirmed (A) and newly-predicted (B) vectors occur throughout the Americas. (A) Observed occurrences of confirmed vectors of *Leishmania spp*. that cause ACL, taken from GBIF and plotted in arcGIS (Esri, USGS | Esri, Garmin, FAO, NOAA, USGS) [79, 80]. (B) Observed occurrences of sandflies of unknown vector status that our models assigned a predicted probability above 0.5. Most predicted vectors are in Brazil due to more extensive survey efforts and availability of public data [79, 80]. Maps showing species richness and vector distribution for each species of *Leishmania spp.* can be found in the supplementary materials (Fig S4 and S5).

The top two unknown sandflies predicted by the models were *Psychodopygus amazonensis* (mean probability = 0.868), which has been observed carrying *L. naiffi* in the wild, and *Nyssomyia antunesi* (mean probability = 0.852), which has been observed carrying *L. lindenbergi* (Table 1). The model also predicted *Psychodopygus claustrei* and *Psychodopygus guyanensis*, both of which have not been observed to carry any species of *Leishmania*, and *Pintomyia pessoai,* which can carry *L. braziliensis*. As such, our model suggests that not only can these sandflies become infected with *Leishmania* spp. that cause ACL, but they can also transmit the parasites and may be important vectors transmitting *Leishmania* spp. to humans, as well as among reservoir hosts. Our secondary model, which predicted the probability that sandflies can be naturally infected with

zoonotic *Leishmania,* predicted *Psychodopygus amazonensis* and *Nyssomyia antunesi* with probabilities above 0.93, along with assigning *Psychodopygus claustrei*, *Psychodopygus guyanensis,* and *Pintomyia pessoai* probability scores above 0.75 (Table S5).

The four most important features in our primary model (i.e., trained on confirmed vectors) are (i) the number of citations in PubMed, (ii) the genus of the sandfly, (iii) number of ecoregions the sandfly inhabits , and (iv) mean canopy height. Partial dependence plots (Fig 4B) indicate that sandflies that have greater study effort, live in areas with greater canopy height, and inhabit many different ecoregions are more likely to be *Leishmania* spp. vectors. Relative importance for the top twenty features and partial dependence plots for the top sixteen features are in the supplementary materials (Table S3, Fig S2, Fig S3). The trait profiles (partial dependence plots) of human modification index and forest land integrity index displayed opposite directionality, and our model assigned greater vector probabilities to sandflies that inhabit environments with less human modification and a greater land integrity index (Fig S2). Other important variables included synanthropy (the tendency of an organism to live close to people and benefit from domestic habitats), main habitat, number of lateral teeth, along with environmental variables like rainfall, temperature, and temperature range, which are worth noting as climate drivers of sandfly development and survival [2, 14, 15]. The top features of our secondary model (i.e., trained on confirmed and potential vectors) were (i) the genus of the sandfly, (ii) the sandfly's main habitat, (iii) synanthropy, and (iv) number of ecoregions the sandfly inhabits (Table S6); the full list can be found in the supplementary materials.

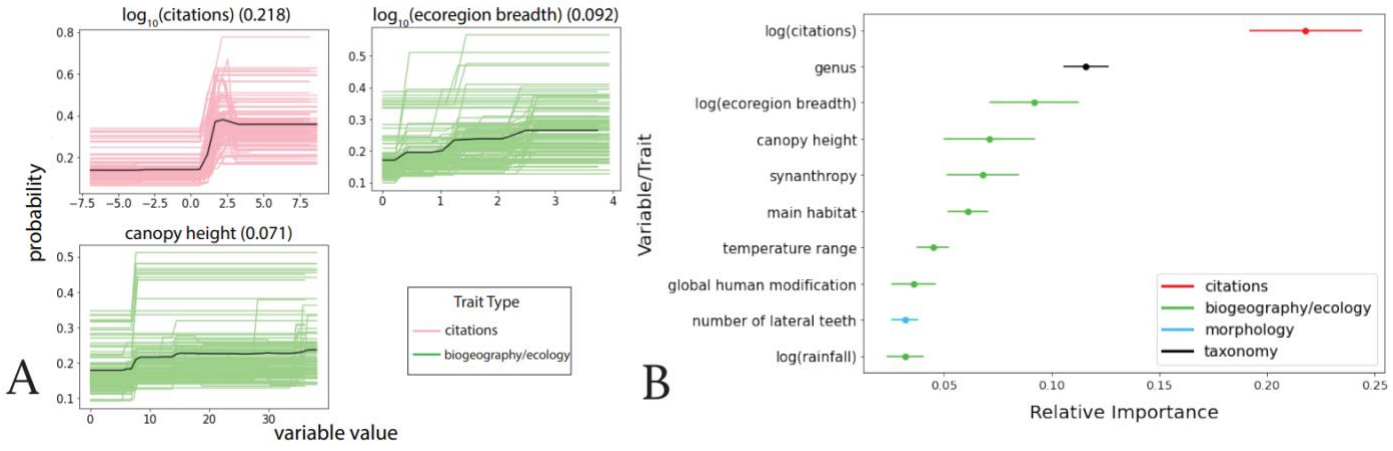


Fig 4: Human biting, study effort, and canopy height were the most important features for predicting vector status. (A) Partial dependence plots of the top three variables from the BRT analysis showing the marginal dependence of each trait (shown in order of importance) on the probability of being a vector of ACL. The variable along with its average importance (on a scale of 0-1) are above each plot, the trait value is shown on the x axis, and the effect on probability is shown on the y-axis. The colored lines represent the marginal dependence of the trait from the 100 BRT models, while the solid black line represents the average dependence. The definition of each variable can be found in Table S1. (B) Variable importance, scaled from 0-1, for the top 10 most important variables with 95% confidence intervals. Points represent mean gain value across 100 iterations. The importances for binary variables were summed up to obtain a single value for the entire categorical variable.

Our target shuffling subanalysis returned a performance score of 0.50 (i.e., the model was no better at predicting shuffled labels than a coin-flip), indicating that our model is not overfit and simply finding spurious correlations in the data [94]. The citation prediction model was able to predict citation count to some extent ($R^2 = 0.18$), but has a different trait profile than that of our original vector models, suggesting that our predictions of *Leishmania* vectors do not simply reflect study bias (Table S3, Table S4).

Discussion:

Our primary model leveraged ecological, behavioral, taxonomic, and biogeographic characteristics of sandfly species found across the Americas to predict the probability of a sandfly being a vector of ACL. The group of 100 BRT models was able to classify sandfly vectors with 86% accuracy and identified several previously unknown species with a relatively high probability of being a vector. Similarly, our secondary model was able to determine which sandflies might be capable of carrying *Leishmania* with high accuracy. Overall, we found that the ecology and taxonomic features of a sandfly are most important in determining whether it has the potential to be a vector, followed by behavioral and morphological features. While citation count was the most important factor in predicting a vector, our citation count subanalysis suggested that our results were not primarily driven by study bias, as it performed poorly and had a different trait profile than our vector models. Study effort may indicate that *Leishmania* vectors are undersampled, but it may also indicate that vectors have a broad range

that interfaces with human settlements, as it would be sampled in many different field surveys across the Americas. We found that both an increase in citation count and ecoregion breadth, i.e., the number of unique ecoregions in which the sandfly has been observed, were associated with a higher likelihood of a sandfly being a vector. This may suggest that the higher the species' propensity to adapt to and live in many different environments, the more likely it is able to survive—and be captured and sampled—in an environment with human inhabitants and transmit the *Leishmania* parasite.

Recent studies have shown that reservoir host relative abundance increases but overall mammal diversity decreases with human modification, while sandfly density increases with mammal diversity and decreases with human modification [95]. Our results support these findings as our model shows that sandflies that occupy areas with lower human modification and higher land integrity are more likely to be vectors. Since the opposite is true for the effect of human modification on reservoir host communities, this may indicate that the highest risk of *Leishmania* spillover lies at the interface between human modification and intact forests [96]. Indeed, synanthropic sandfly species that live in domestic habitats, but are associated with higher-integrity land cover, were more likely to be vectors. Interestingly, canopy height, which can be indicative of forest intactness, was one of the most important features in our model. An increase in canopy height suggests older forest and trees, as well as more space along the bark for sandflies to breed and live. Based on our model, sandflies found in these preserved areas with high forest integrity and low human modification are more likely to be vectors of *Leishmania*. Since sandflies in general are weak fliers and typically prefer to stay within 30 to 300 meters of their breeding and living environment [2, 96], our analyses support previous hypotheses that transmission can happen when people enter previously undisturbed areas and come into contact with sandflies [2]. Since synanthropic sandflies are also more likely to be vectors due to their proximity to humans, this contact and transmission is thought to occur at interfaces between intact forests and domestic settlements, as previously suggested. Additionally, canopy height was correlated with sandfly wing width, which was removed for model training, yet indicates that our model accounts for sandfly biology and relevant environmental interactions. It is difficult, however, to track spillover that occurs at these interfaces as temporal and spatial differences in vector and host habitat make it challenging to observe the alignment of human, reservoir host, and sandfly dynamics conducive to wildlife-sandfly-human transmission [97]. Further, overall data sparsity of sandfly

occurrence points and traits, as well as vector and host interactions, makes modeling ACL transmission even more complicated.

The second most important predictor of vector status was a sandfly's genus, indicating that sandflies from certain genera can be more or less inclined to be a vector. This suggests certain traits that occur in these parts of the sandfly phylogenetic tree are important for vector status, but full sandfly phylogenies were unavailable, so we were unable to quantitatively asses the effect of genetic distance. This highlights the importance of careful taxonomic work as well as high quality genetic data for as many species as possible.

We also found non-linear relationships between vector probability and climate variables. Vectors that occur in habitats with high temperatures and temperature ranges were less likely to be vectors, indicating that there is an optimal temperature range in which vector transmission occurs. Sandflies in environments with higher rainfall are also more likely to be vectors, indicating an optimal range of precipitation that might also support high sandfly population abundance. This suggests there is a landscape of differential vector transmission success, which is worth investigating to determine risk score across different habitats.

The analysis helped to confirm that some sandfly species observed to carry *Leishmania* in the wild but have not yet been confirmed as capable of transmitting to humans (i.e., potential vectors) are likely to be infectious. These predicted vectors should be empirically tested to determine if they are indeed vectors, and using genomic blood meal analyses [98], determine what reservoir hosts they feed on. In particular, efforts should focus on *Psychodopygus amazonensis* and *Nyssomia antunesi*. *Psychodopygus amazonensis* has been observed to carry *L. naiffi*, and shares a genus with eleven sandflies that are proven vectors of *L. braziliensis* and *L. naiffi* [22], with the model assigning three new sandflies of genus *Psychodopygus* probability scores above 0.5. *Nyssomyia antunesi* is an anthropophilic sandfly observed to carry *L. lindenbergi*, and there is strong evidence that it is a vector of ACL [24, 99]. It is additionally taxonomically related to *Nyssomyia whitmani* and *Nyssomyia umbratilis*, both vectors of *L. braziliensis* and *L. guyanensis* [22, 100, 101].

Our secondary model trained on identifying potential vectors supported the sandfly predictions generated by our primary model. Sandflies predicted by the primary model were ranked highly by our secondary model. However, not all potential vectors (sandflies that have been observed to carry but not transmit *Leishmania*) were predicted to be highly likely vectors to humans, indicating that the ability to carry the parasite, while important, may not be representative of the sandfly's ability to be a human vector of ACL. Rather, there are additional biological variables like forest integrity, canopy height, and temperature range that play important roles in *Leishmania* transmission to humans, as indicated in our analysis. Importantly, while these sandflies may not be vectors involved in human transmission, they may still be involved in transmitting the pathogen among reservoir hosts, maintaining a reservoir community, and warranting further research.

The majority of occurrence points of newly identified sandfly vector species were found in Brazil. This could be because Brazil has the most sampling studies done on sandflies, but incidence of ACL is also highest in Brazil according to Pan American Health Organization (PAHO) [6]. So while Brazil is generally better studied, there is also high vector species richness and abundance, specifically with predicted vectors concentrated in central Brazil in Mato Grosso and Mato Grosso Do Sul. Certain areas in Brazil are well sampled, but there are many regions that are neglected and ought to be the focus of new sampling efforts in order to identify new vectors. Future sandfly sampling efforts should be concentrated in these areas predicted by the model, in addition to poorly explored regions with not many sampling surveys or efforts, such as the Caatinga biome (northeastern Brazil) and the western Amazon. Then, to assess vector competence and incrimination according to the Killick-Kendrick criteria, entomological studies should be carried out in disease hotspots, followed by human disease notification and prevention efforts [98]. We additionally identified predicted vector occurrence points in Madre de Dios, Peru, another hotspot of ACL transmission, the eastern coast of French Guiana, and northwestern Colombia. Our analysis suggests new vectors (*Psy. amazonensis, N. antunesi, Psy. claustrei, Psy. guyanensis, P. pessoai, Pa. bigeniculata, T. auraensis, Psy. chagasi, T. castanheirai, S. sordelli, Pa. lanei, Ev. infrapsinosa, W. rotundipennis*) to be incorporated into surveillance efforts in these regions.

Due to data sparsity and low coverage, we were unable to generate *Leishmania* species-specific models, as some species had only one or two confirmed vectors. Instead, we opted for a genus-wide model that includes all

Leishmania parasites that cause cutaneous leishmaniasis in the Americas. Although it is more informative and reliable to have more data for training the model, we lose the *Leishmania* species-specific vector transmission information in this general model. Additional studies into the *Leishmania* parasites themselves as well as vectors that transmit them would increase the model accuracy for predicting specific *Leishmania* transmission cycles. In addition, some traits had to be removed from the final model due to low coverage (< 10%). These included activity throughout the day, lifespan, and which taxa they feed on, which would be valuable to future work in sandfly vector transmission of *Leishmania*. Data sparsity along with gaps in basic knowledge about ACL transmission contributed to model uncertainty. For instance, low trait coverage for some species could affect how they are predicted. While we are relatively confident in sandfly species predicted with high probability, additional data are required to reach similar predictive confidence for species that are currently predicted with low probability. We are more likely to have failed to identify an unknown and understudied species as a vector than an unknown but well-studied species. Additionally, although North American sandflies were included in the models, there were few occurrence points for these species, and they did not have high vector prediction scores. This is another limitation of our analysis considering the rise in North American cases of leishmaniasis and potential range shifts of sandfly vectors. Specifically, more data about species occurrence, behavior, and morphology are necessary to deepen our understanding of ACL vectors and spillover transmission in human populations.

While our model generally performed well, it assigned two known sandfly vectors a probability score lower than 0.5. One of these sandflies was *Pintomyia youngi*, which is a confirmed vector of *L. braziliensis*, and a potential vector of *L. amazonensis*. Due to difficulty in taxonomic identification procedures, *Pintomyia youngi* could have been misidentified as *Lutzomyia townsendi* [31, 102, 103], which means that covariate and/or vector status data might be assigned to the wrong species. This can lead to misrepresentation in the model (that already relies on sparse data), which might explain the low probability score assigned to *Pintomyia youngi*. A more stringent taxonomic identification criteria will help with *Leishmania* studies worldwide, as well as with modeling efforts. The other species with a low model-predicted probability was *Pintomyia ovallesi*, a lesser-studied vector of ACL in Central America [24, 104, 105]. It is possible that our model failed to identify them due to a lack of data and bias towards Brazilian sandfly vectors.

By understanding the environmental features that promote specific sandfly vector species, we can better (i) understand ACL transmission cycles as they impact human risk, (ii) understand the potential impacts of human modifications such as land use change on ACL transmission, and (iii) predict how ACL transmission cycles will respond to global change in the future. Based on the traits most predictive of vector status, we hypothesize that human risk peaks at the interface of human, vector, and host communities in intact forest areas with high canopies and relatively low temperature variance. Further analysis of epidemiological data and surveillance data on hosts, parasites, and vectors could be used to test this hypothesis. In addition, predicted but not yet confirmed vectors of ACL, i.e., *Psychodopygus amazonensis, Nyssomyia antunesi, Psychodopygus claustrei, Psychodopygus guyanensis,* and *Pintomyia pessoai,* should be empirically tested for competence in the laboratory. If confirmed, vector control methods should be expanded to account for the new vectors. Similarly, sampling and public health efforts should target central-west Brazil, where predicted vectors are concentrated, as well as lesser studied areas such as northeastern Brazil and the western Amazon. As ACL increases across the Americas, sandfly species that fit the trait profile of other confirmed vectors—including dwelling in forest with high canopy and high integrity, biting humans, and spanning many ecoregions—are important targets for *Leishmania* surveillance to better identify reservoir host transmission cycles and risk factors for human spillover and to cost-effectively target the most likely potential vectors.

Acknowledgements:

The authors thank the Mordecai Lab at Stanford University for their feedback.

References:

1. Chaves LF, Hernandez M-J. Mathematical modelling of American Cutaneous Leishmaniasis: incidental hosts and threshold conditions for infection persistence. Acta Tropica. 2004;92: 245–252. doi:10.1016/j.actatropica.2004.08.004
2. Torres-Guerrero E, Quintanilla-Cedillo M, Ruiz-Esmenjaud J, Arenas R. Leishmaniasis: A review. F1000Research. 2017;6: 750. doi:10.12688/f1000research.11120.1
3. Royer M, Crowe M. American Cutaneous Leishmaniasis: A Cluster of 3 Cases During Military Training in Panama. Archives of Pathology & Laboratory Medicine. 2002;126: 471–473. doi:10.5858/2002-126-0471-ACL
4. Prevention C-C for DC and. CDC - Leishmaniasis - Epidemiology & Risk Factors. 18 Feb 2020 [cited 10 Apr 2022]. Available: https://www.cdc.gov/parasites/leishmaniasis/epi.html
5. Mexico: Leishmaniasis | IAMAT. [cited 10 Apr 2022]. Available: https://www.iamat.org/country/mexico/risk/leishmaniasis
6. Epidemiological Report of the Americas. PAHO, WHO. 2021. Available: https://iris.paho.org/handle/10665.2/55368
7. Curtin J, Aronson N. Leishmaniasis in the United States: Emerging Issues in a Region of Low Endemicity. Microorganisms. 2021;9: 578. doi:10.3390/microorganisms9030578
8. Freeman K. American cutaneous leishmaniasis. J R Army Med Corps. 1983;129: 167–173. doi:10.1136/jramc-129-03-09
9. Pimenta PFP, de Freitas VC, Monteiro CC, Pires ACMA, Secundino NFC. Biology of the Leishmania–Sand Fly Interaction. In: Rangel EF, Shaw JJ, editors. Brazilian Sand Flies : Biology, Taxonomy, Medical Importance and Control. Cham: Springer International Publishing; 2018. pp. 319–339. doi:10.1007/978-3-319-75544-1_6
10. Goto H, Lauletta Lindoso JA. Cutaneous and mucocutaneous leishmaniasis. Infect Dis Clin North Am. 2012;26: 293–307. doi:10.1016/j.idc.2012.03.001
11. Treatment of American Cutaneous Leishmaniasis with Miltefosine, an Oral Agent | Clinical Infectious Diseases | Oxford Academic. [cited 10 Apr 2022]. Available: https://academic.oup.com/cid/article/33/7/e57/433907
12. Rodríguez DE, Sebastian MS, Pulkki-Brännström A-M. "Cheaper and better": Societal cost savings and budget impact of changing from systemic to intralesional pentavalent antimonials as the first-line treatment for cutaneous leishmaniasis in Bolivia. PLOS Neglected Tropical Diseases. 2019;13: e0007788. doi:10.1371/journal.pntd.0007788
13. 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;374: 20180342. doi:10.1098/rstb.2018.0342
14. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, et al. Thermal biology of mosquito-borne disease. Ecol Lett. 2019;22: 1690–1708. doi:10.1111/ele.13335
15. Hlavacova J, Votypka J, Volf P. The Effect of Temperature on Leishmania (Kinetoplastida: Trypanosomatidae) Development in Sand Flies. Journal of Medical Entomology. 2013;50: 955–958. doi:10.1603/ME13053

16. Villena OC, Ryan SJ, Murdock CC, Johnson LR. Temperature impacts the environmental suitability for malaria transmission by Anopheles gambiae and Anopheles stephensi. Ecology. 2022; e3685. doi:10.1002/ecy.3685
17. Cecílio P, Cordeiro-da-Silva A, Oliveira F. Sand flies: Basic information on the vectors of leishmaniasis and their interactions with Leishmania parasites. Commun Biol. 2022;5: 1–12. doi:10.1038/s42003-022-03240-z
18. Spatial modeling of cutaneous leishmaniasis in the Andean region of Colombia - PMC. [cited 10 Apr 2022]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4957495/
19. King RJ, Campbell-Lendrum DH, Davies CR. Predicting Geographic Variation in Cutaneous Leishmaniasis, Colombia. Emerg Infect Dis. 2004;10: 598–607. doi:10.3201/eid1004.030241
20. Ali Hanafi-Bojd A, Yaghoobi-Ershadi MR, Haghdoost AA, Akhavan AA, Rassi Y, Karimi A, et al. Modeling the Distribution of Cutaneous Leishmaniasis Vectors (Psychodidae: Phlebotominae) in Iran: A Potential Transmission in Disease Prone Areas. Journal of Medical Entomology. 2015;52: 557–565. doi:10.1093/jme/tjv058
21. Han Barbara, 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;27: 59–65. doi:10.1016/j.epidem.2019.01.005
22. Evans MV, Dallas TA, Han BA, Murdock CC, Drake JM. Data-driven identification of potential Zika virus vectors. eLife. 6: e22053. doi:10.7554/eLife.22053
23. Killick-Kendrick R. Phlebotomine vectors of the leishmaniases: a review. Medical and Veterinary Entomology. 1990;4: 1–24. doi:10.1111/j.1365-2915.1990.tb00255.x
24. 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 Neglected Tropical Diseases. 2016;10: e0004349. doi:10.1371/journal.pntd.0004349
25. Lewis DJ. Functional morphology of the mouth parts in New World phlebotomine sandflies (Diptera: Psychodidae). Transactions of the Royal Entomological Society of London. 1975;126: 497–532. doi:10.1111/j.1365-2311.1975.tb00859.x
26. Ready PD, Vigoder FM, Rangel EF. Molecular and Biochemical Markers for Investigating the Vectorial Roles of Brazilian Sand Flies. In: Rangel EF, Shaw JJ, editors. Brazilian Sand Flies : Biology, Taxonomy, Medical Importance and Control. Cham: Springer International Publishing; 2018. pp. 213–250. doi:10.1007/978-3-319-75544-1_3
27. Rangel EF, Lainson R, Carvalho BM, Costa SM, Shaw JJ. Sand Fly Vectors of American Cutaneous Leishmaniasis in Brazil. In: Rangel EF, Shaw JJ, editors. Brazilian Sand Flies : Biology, Taxonomy, Medical Importance and Control. Cham: Springer International Publishing; 2018. pp. 341–380. doi:10.1007/978-3-319-75544-1_7
28. Shimabukuro PHF, Andrade AJ de, Galati EAB. Checklist of American sand flies (Diptera, Psychodidae, Phlebotominae): genera, species, and their distribution. ZooKeys. 2017;660: 67–106. doi:10.3897/zookeys.660.10508
29. Brazil RP, Brazil BG. Bionomy: Biology of Neotropical Phlebotomine Sand Flies. In: Rangel EF, Shaw JJ, editors. Brazilian Sand Flies : Biology, Taxonomy, Medical Importance and Control. Cham: Springer International Publishing; 2018. pp. 299–318. doi:10.1007/978-3-319-75544-1_5

30. Young D. G. Duncan M. A.. 1994. Guide to the identification and geographic distribution of Lutzomyia sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae). Mem. Am. Entomol. Inst. 54: 1–881.
31. A Bianchi Galati E, Cáceres AG. Description of Micropygomyia (Micropygomyia) ancashensis sp. nov. and the female of Lutzomyia (Helcocyrtomyia) chavinensis Pérez & Ogusuku (Diptera, Psychodidae, Phlebotominae) from Ancash department, Peru. Mem Inst Oswaldo Cruz. 2007;102: 833–838. doi:10.1590/s0074-02762007005000091
32. Filho J, Galati E, Falcão A. Redescription of Nyssomyia intermedia (Lutz & Neiva, 1912) and Nyssomyia neivai (Pinto, 1926) (Diptera: Psychodidae). Memórias do Instituto Oswaldo Cruz. 2004;98: 1059–65. doi:10.1590/S0074-02762003000800015
33. Da Silva YY, Sales KGDS, Miranda DEDO, Figueredo LA, Brandão-Filho SP, Dantas-Torres F. Detection of Leishmania DNA in Sand Flies (Diptera: Psychodidae) From a Cutaneous Leishmaniasis Outbreak Area in Northeastern Brazil. Journal of Medical Entomology. 2020;57: 529–533. doi:10.1093/jme/tjz189
34. Lozano-Sardaneta YN, Sánchez-Montes S, Sánchez-Cordero V, Becker I, Paternina LE. Molecular detection of Leishmania infantum in sand flies (Diptera: Psychodidae: Phlebotominae) from Veracruz, Mexico. Acta Trop. 2020;207: 105492. doi:10.1016/j.actatropica.2020.105492
35. Diniz MMC de SL, Ovallos FG, de Castro Gomes CM, de Oliveira Lavitschka C, Galati EAB. Host-biting rate and susceptibility of some suspected vectors to Leishmania braziliensis. Parasit Vectors. 2014;7: 139. doi:10.1186/1756-3305-7-139
36. de Oliveira EF, Casaril AE, Mateus NLF, Murat PG, Fernandes WS, Oshiro ET, et al. Leishmania amazonensis DNA in wild females of Lutzomyia cruzi (Diptera: Psychodidae) in the state of Mato Grosso do Sul, Brazil. Mem Inst Oswaldo Cruz. 2015;110: 1051–1057. doi:10.1590/0074-02760150317
37. Moreno M, Ferro C, Rosales-Chilama M, Rubiano L, Delgado M, Cossio A, et al. First report of Warileya rotundipennis (Psychodidae: Phlebotominae) naturally infected with Leishmania (Viannia) in a focus of cutaneous leishmaniasis in Colombia. Acta Tropica. 2015;148: 191–196. doi:10.1016/j.actatropica.2015.04.017
38. Barreto M. P.. 1943. Observações sobre a Biologia em Condições Naturais, dos Flebótomos de São Paulo (Diptera: Psychodidae). Tipografia Rossolino, São Paulo, Brazil.
39. Barrios SPG, Pereira LE, Monaco NZN, Graciolli G, Casaril AE, Infran J de OM, et al. Synanthropy and diversity of Phlebotominae in an area of intense transmission of visceral leishmaniasis in the South Pantanal floodplain, Midwest Brazil. PLOS ONE. 2019;14: e0215741. doi:10.1371/journal.pone.0215741
40. Alves GB, Oshiro ET, Leite M da C, Melão AV, Ribeiro LM, Mateus NLF, et al. Phlebotomine sandflies fauna (Diptera: Psychodidae) at rural settlements in the municipality of Cáceres, State of Mato Grosso, Brazil. Rev Soc Bras Med Trop. 2012;45: 437–443. doi:10.1590/s0037-86822012005000010
41. Rodrigues E de AS, Andrade Filho JD, Limongi JE, Paula MBC de. Sandfly fauna (Diptera: Psychodidae) in Parque do Sabiá complex, Uberlândia, Minas Gerais, Brazil. Rev Inst Med trop S Paulo. 2011;53: 255–258. doi:10.1590/S0036-46652011000500003
42. Dorval MEC, Alves TP, Cristaldo G, Rocha HC da, Alves MA, Oshiro ET, et al. Sand fly captures with Disney traps in area of occurrence of Leishmania (Leishmania) amazonensis in the state of Mato Grosso do Sul, mid-western Brazil. Rev Soc Bras Med Trop. 2010;43: 491–495. doi:10.1590/S0037-86822010000500003

43. Oliveira AG de, Galati EAB, Oliveira O de, Oliveira GR de, Espindola IAC, Dorval MEC, et al. Abundance of Lutzomyia longipalpis (Diptera: Psychodidae: Phlebotominae) and urban transmission of visceral leishmaniasis in Campo Grande, state of Mato Grosso do Sul, Brazil. Mem Inst Oswaldo Cruz. 2006;101: 869–874. doi:10.1590/S0074-02762006000800008
44. Salomón OD, Rossi GC, Cousiño B, Spinelli GR, Rojas de Arias A, López del Puerto DG, et al. Phlebotominae sand flies in Paraguay: abundance distribution in the Southeastern region. Mem Inst Oswaldo Cruz. 2003;98: 185–190. doi:10.1590/S0074-02762003000200004
45. Senne NA, Vilela TS, Sanavria A, Santos HA, Rabello RS, Angelo IC. Ecology and spatial distribution of sand fly species in low endemic areas for American Tegumentary Leishmaniasis in the municipality of Seropédica, Rio de Janeiro, Brazil. Medical and Veterinary Entomology. 2021;35: 371–378. doi:10.1111/mve.12505
46. Sales KG da S, de Oliveira Miranda DE, Costa PL, da Silva FJ, Figueredo LA, Brandão-Filho SP, et al. Home sweet home: sand flies find a refuge in remote indigenous villages in north-eastern Brazil, where leishmaniasis is endemic. Parasites & Vectors. 2019;12: 118. doi:10.1186/s13071-019-3383-1
47. Pinheiro MPG, Silva MM de M, Júnior JBS, da Silva JHT, Alves M de L, Ximenes M de FF de M. Sand flies (Diptera, Psychodidae, Phlebotominae), vectors of Leishmania protozoa, at an Atlantic Forest Conservation Unit in the municipality of Nísia Floresta, Rio Grande do Norte state, Brazil. Parasites & Vectors. 2016;9: 83. doi:10.1186/s13071-016-1352-5
48. Staniek ME, Hamilton JGC. Odour of domestic dogs infected with Leishmania infantum is attractive to female but not male sand flies: Evidence for parasite manipulation. PLOS Pathogens. 2021;17: e1009354. doi:10.1371/journal.ppat.1009354
49. Rêgo FD, Rugani JMN, Shimabukuro PHF, Tonelli GB, Quaresma PF, Gontijo CMF. Molecular Detection of Leishmania in Phlebotomine Sand Flies (Diptera: Psychodidae) from a Cutaneous Leishmaniasis Focus at Xakriabá Indigenous Reserve, Brazil. PLOS ONE. 2015;10: e0122038. doi:10.1371/journal.pone.0122038
50. Rêgo FD, Shimabukuro PHF, Quaresma PF, Coelho IR, Tonelli GB, Silva KMS, et al. Ecological aspects of the Phlebotominae fauna (Diptera: Psychodidae) in the Xakriabá Indigenous Reserve, Brazil. Parasites Vectors. 2014;7: 220. doi:10.1186/1756-3305-7-220
51. Carvalho BM, Maximo M, Costa WA, de Santana ALF, da Costa SM, da Costa Rego TAN, et al. Leishmaniasis transmission in an ecotourism area: potential vectors in Ilha Grande, Rio de Janeiro State, Brazil. Parasites & Vectors. 2013;6: 325. doi:10.1186/1756-3305-6-325
52. Kato H, Gomez EA, Cáceres AG, Vargas F, Mimori T, Yamamoto K, et al. Natural Infections of Man-Biting Sand Flies by Leishmania and Trypanosoma Species in the Northern Peruvian Andes. Vector-Borne and Zoonotic Diseases. 2011;11: 515–521. doi:10.1089/vbz.2010.0138
53. Saraiva L, Reis AS, Rugani JMN, Pereira AAS, Rêgo FD, Lima ACVM da R, et al. Survey of Sand Flies (Diptera: Psychodidae) in an Environmentally Protected Area in Brazil. PLOS ONE. 2015;10: e0134845. doi:10.1371/journal.pone.0134845
54. Zorrilla V, Santos MBDL, Espada L, Santos R del P, Fernandez R, Urquia A, et al. Distribution and identification of sand flies naturally infected with Leishmania from the Southeastern Peruvian Amazon. PLOS Neglected Tropical Diseases. 2017;11: e0006029. doi:10.1371/journal.pntd.0006029

55. Toro-Cantillo A, Atencia Pineda M, Hoyos R. Flebotomíneos (Diptera: Psychodidae) colectados en área rural de San Bernardo del Viento (Córdoba Colombia). Revista MVZ Córdoba. 2017;22: 6044. doi:10.21897/rmvz.1074
56. Pereira Júnior AM, Souza ABN, Castro TS, da Silva MS, de Paulo PFM, Ferreira GEM, et al. Diversity, natural infection and blood meal sources of phlebotomine sandflies (Diptera, Psychodidae) in the western Brazilian Amazon. Mem Inst Oswaldo Cruz. 2019;114: e190170. doi:10.1590/0074-02760190170
57. de Ávila MM, Brilhante AF, de Souza CF, Bevilacqua PD, Galati EAB, Brazil RP. Ecology, feeding and natural infection by Leishmania spp. of phlebotomine sand flies in an area of high incidence of American tegumentary leishmaniasis in the municipality of Rio Branco, Acre, Brazil. Parasites Vectors. 2018;11: 64. doi:10.1186/s13071-018-2641-y
58. Vasconcelos dos Santos T, Silva F, Barata I, Andrade A, Galati E. A new species of phlebotomine, Trichophoromyia adelsonsouzai (Diptera: Psychodidae) of Brazilian Amazonia. Memorias do Instituto Oswaldo Cruz. 2013;0: 0. doi:10.1590/0074-0276130159
59. Azevedo ACR, Souza NA, Meneses CRV, Costa WA, Costa SM, Lima JB, et al. Ecology of sand flies (Diptera: psychodidae: phlebotominae) in the north of the state of Mato Grosso, Brazil. Mem Inst Oswaldo Cruz. 2002;97: 459–464. doi:10.1590/s0074-02762002000400002
60. Machado TDO, Minuzzi-Souza TTC, Ferreira T de S, Freire LP, Timbó RV, Vital TE, et al. The role of gallery forests in maintaining Phlebotominae populations: potential *Leishmania* spp. vectors in the Brazilian savanna. Mem Inst Oswaldo Cruz. 2017;112: 681–691. doi:10.1590/0074-02760170126
61. Ovallos FG, Silva YRE, Fernandez N, Gutierrez R, Galati EAB, Sandoval CM. The sandfly fauna, anthropophily and the seasonal activities of *Pintomyia spinicrassa* (Diptera: Psychodidae: Phlebotominae) in a focus of cutaneous leishmaniasis in northeastern Colombia. Mem Inst Oswaldo Cruz. 2013;108: 297–302. doi:10.1590/S0074-02762013000300007
62. Brilhante AF, de Ávila MM, de Souza JF, Medeiros-Sousa AR, Sábio PB, de Paula MB, et al. Attractiveness of black and white modified Shannon traps to phlebotomine sandflies (Diptera, Psychodidae) in the Brazilian Amazon Basin, an area of intense transmission of American cutaneous leishmaniasis. Parasite. 24: 20. doi:10.1051/parasite/2017021
63. Dorval MEC, Cristaldo G, Rocha HC da, Alves TP, Alves MA, Oshiro ET, et al. Phlebotomine fauna (Diptera: Psychodidae) of an American cutaneous leishmaniasis endemic area in the state of Mato Grosso do Sul, Brazil. Mem Inst Oswaldo Cruz. 2009;104: 695–702. doi:10.1590/s0074-02762009000500005
64. Baum M, Ribeiro MCV da C, Lorosa ES, Damasio GAC, Castro EA de. Eclectic feeding behavior of Lutzomyia (Nyssomyia) intermedia (Diptera, Psychodidae, Phlebotominae) in the transmission area of American cutaneous leishmaniasis, state of Paraná, Brazil. Rev Soc Bras Med Trop. 2013;46: 560–565. doi:10.1590/0037-8682-0157-2013
65. Alves GB, Oshiro ET, Leite M da C, Melão AV, Ribeiro LM, Mateus NLF, et al. Phlebotomine sandflies fauna (Diptera: Psychodidae) at rural settlements in the municipality of Cáceres, State of Mato Grosso, Brazil. Rev Soc Bras Med Trop. 2012;45: 437–443. doi:10.1590/s0037-86822012005000010
66. Galati EAB, Marassá AM, Gonçalves-Andrade RM, Consales CA, Bueno EFM. Phlebotomines (Diptera, Psychodidae) in the Ribeira Valley Speleological Province - 1. Parque Estadual Intervales, state of São Paulo, Brazil. Rev Bras entomol. 2010;54: 311–321. doi:10.1590/S0085-56262010000200015

67. Balbino VQ, Coutinho-Abreu IV, Sonoda IV, Marques da Silva W, Marcondes CB. Phlebotomine sandflies (Diptera: Psychodidae) of the Atlantic forest in Recife, Pernambuco state, Brazil: the species coming to human bait, and their seasonal and monthly variations over a 2-year period. Ann Trop Med Parasitol. 2005;99: 683–693. doi:10.1179/136485905X65116
68. Pereira Filho AA, Bandeira M da CA, Fonteles RS, Moraes JLP, Lopes CRG, Melo MN, et al. An ecological study of sand flies (Diptera: Psychodidae) in the vicinity of Lençóis Maranhenses National Park, Maranhão, Brazil. Parasites & Vectors. 2015;8: 442. doi:10.1186/s13071-015-1045-5
69. Bray DP, Alves GB, Dorval ME, Brazil RP, Hamilton JG. Synthetic sex pheromone attracts the leishmaniasis vector Lutzomyia longipalpis to experimental chicken sheds treated with insecticide. Parasit Vectors. 2010;3: 16. doi:10.1186/1756-3305-3-16
70. Marassá AM, Galati EAB, Bergamaschi DP, Consales CA. Blood feeding patterns of *Nyssomyia intermedia* and *Nyssomyia neivai* (Diptera, Psychodidae) in a cutaneous leishmaniasis endemic area of the Ribeira Valley, State of São Paulo, Brazil. Rev Soc Bras Med Trop. 2013;46: 547–554. doi:10.1590/0037-8682-0168-2013
71. Alves VR, Freitas RA de, Santos FL, Barrett TV. Diversity of sandflies (Psychodidae: Phlebotominae) captured in sandstone caves from Central Amazonia, Brazil. Mem Inst Oswaldo Cruz. 2011;106: 353–359. doi:10.1590/s0074-02762011000300016
72. Teles CBG, dos Santos AP de A, Freitas RA, de Oliveira AFJ, Ogawa GM, Rodrigues MS, et al. Phlebotomine sandfly (Diptera: Psychodidae) diversity and their Leishmania DNA in a hot spot of American Cutaneous Leishmaniasis human cases along the Brazilian border with Peru and Bolivia. Mem Inst Oswaldo Cruz. 2016;111: 423–432. doi:10.1590/0074-02760160054
73. de Souza AAA, da Rocha Barata I, das Graças Soares Silva M, Lima JAN, Jennings YLL, Ishikawa EAY, et al. Natural Leishmania (Viannia) infections of phlebotomines (Diptera: Psychodidae) indicate classical and alternative transmission cycles of American cutaneous leishmaniasis in the Guiana Shield, Brazil. Parasite. 24: 13. doi:10.1051/parasite/2017016
74. de Souza CF, Brazil RP, Bevilacqua PD, Andrade Filho JD. The phlebotomine sand flies fauna in Parque Estadual do Rio Doce, Minas Gerais, Brazil. Parasites & Vectors. 2015;8: 619. doi:10.1186/s13071-015-1227-1
75. de Aguiar GM, Vieira VR. Regional Distribution and Habitats of Brazilian Phlebotomine Species. In: Rangel EF, Shaw JJ, editors. Brazilian Sand Flies : Biology, Taxonomy, Medical Importance and Control. Cham: Springer International Publishing; 2018. pp. 251–298. doi:10.1007/978-3-319-75544-1_4
76. Lozano-Sardaneta YN, Jiménez-Girón EI, Rodríguez-Rojas JJ, Sánchez-Montes S, Álvarez-Castillo L, Sánchez-Cordero V, et al. Species diversity and blood meal sources of phlebotomine sand flies (Diptera: Psychodidae) from Los Tuxtlas, Veracruz, Mexico. Acta Tropica. 2021;216: 105831. doi:10.1016/j.actatropica.2021.105831
77. Anaguano DF, Ponce P, Baldeón ME, Santander S, Cevallos V. Blood-meal identification in phlebotomine sand flies (Diptera: Psychodidae) from Valle Hermoso, a high prevalence zone for cutaneous leishmaniasis in Ecuador. Acta Tropica. 2015;152: 116–120. doi:10.1016/j.actatropica.2015.09.004

78. Ontivero IM, Beranek MD, Rosa JR, Ludueña-Almeida FF, Almirón WR. Seasonal distribution of Phlebotomine sandfly in a vulnerable area for tegumentary leishmaniasis transmission in Córdoba, Argentina. Acta Tropica. 2018;178: 81–85. doi:10.1016/j.actatropica.2017.10.028
79. da Silva Soares Souto A, Dilermando J (2022). Fiocruz/COLFLEB - Coleção de Flebotomíneos. Version 1.54. FIOCRUZ - Oswaldo Cruz Foundation. Occurrence dataset https://doi.org/10.15468/sxcpfp accessed via GBIF.org on 2020-5-20.
80. European Bioinformatics Institute (EMBL-EBI), GBIF Helpdesk (2022). INSDC Sequences. Version 1.8. European Nucleotide Archive (EMBL-EBI). Occurrence dataset https://doi.org/10.15468/sbmztx accessed via GBIF.org on 2022-5-20.
81. Galvis-Ovallos F, Casanova C, Bergamaschi DP, Galati EAB. A field study of the survival and dispersal pattern of Lutzomyia longipalpis in an endemic area of visceral leishmaniasis in Brazil. PLOS Neglected Tropical Diseases. 2018;12: e0006333. doi:10.1371/journal.pntd.0006333
82. Morrison AC, Ferro C, Morales A, Tesh RB, Wilson ML. Dispersal of the sand fly Lutzomyia longipalpis (Diptera: Psychodidae) at an endemic focus of visceral leishmaniasis in Colombia. J Med Entomol. 1993;30: 427–435. doi:10.1093/jmedent/30.2.427
83. Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), (cited 18 April 2022), https://cds.climate.copernicus.eu/cdsapp#!/home
84. Didan, K.. MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006. 2015, distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MOD13A2.006. Accessed 2022-04-21.
85. Buchhorn, M. ; Lesiv, M. ; Tsendbazar, N. - E. ; Herold, M. ; Bertels, L. ; Smets, B. Copernicus Global Land Cover Layers—Collection 2. Remote Sensing 2020, 12Volume 108, 1044. doi:10.3390/rs12061044
86. Amante, C. and B. W. Eakins, ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24, 19 pp, March 2009.
87. Simard, M., Pinto, N., Fisher, J., Baccini, A. 2011. Mapping forest canopy height globally with spaceborne lidar. Journal of Geophysical Research. 116: G04021. doi:10.1029/2011JG001708
88. Bioscience, An Ecoregions-Based Approach to Protecting Half the Terrestrial Realm doi:10.1093/biosci/bix014
89. Kennedy, C.M., J.R. Oakleaf, D.M. Theobald, S. Baurch-Murdo, and J. Kiesecker. 2019. Managing the middle: A shift in conservation priorities based on the global human modification gradient. Global Change Biology 00:1-16. doi:10.1111/gcb.14549
90. Grantham HS, Duncan A, Evans TD, Jones KR, Beyer HL, Schuster R, et al. Anthropogenic modification of forests means only 40% of remaining forests have high ecosystem integrity. Nat Commun. 2020;11: 5978. doi:10.1038/s41467-020-19493-3
91. Fischhoff Ilya R., Castellanos Adrian A., Rodrigues João P. G. L. M., Varsani Arvind and Han Barbara A. 2021. Predicting the zoonotic capacity of mammals to transmit SARS-CoV-2Proc. R. Soc. B.2882021165120211651. http://doi.org/10.1098/rspb.2021.1651
92. Han BA, Schmidt JP, Bowden SE, Drake JM. Rodent reservoirs of future zoonotic diseases. Proc Natl Acad Sci U S A. 2015;112: 7039–7044. doi:10.1073/pnas.1501598112

93. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLOS ONE. 2019;14: e0224365. doi:10.1371/journal.pone.0224365
94. Elder John. Evaluate the Validity of Your Discovery with Target Shuffling. White Paper 2014. Available: https://www.elderresearch.com/wp-content/uploads/2021/01/White-Paper_Evaluate-the-Validity-of-Your-Discovery-with-Target-Shuffling_2021.pdf
95. Kocher A, Cornuault J, Gantier J-C, Manzi S, Chavy A, Girod R, et al. Biodiversity and vector-borne diseases: Host dilution and vector amplification occur simultaneously for Amazonian leishmaniases. Molecular Ecology. n/a. doi:10.1111/mec.16341
96. Ghazanfar M, Malik M. Sandfly and Leishmaniasis: A Review. Journal of Ecosystem & Ecography. 2016;6. doi:10.4172/2157-7625.1000207
97. Sandoval-Ramírez CM, Hernández C, Teherán AA, Gutierrez-Marin R, Martínez-Vega RA, Morales D, et al. Complex ecological interactions across a focus of cutaneous leishmaniasis in Eastern Colombia: novel description of Leishmania species, hosts and phlebotomine fauna. Royal Society Open Science. 7: 200266. doi:10.1098/rsos.200266
98. Remadi L, Chargui N, Jiménez M, Molina R, Haouas N, González E, et al. Molecular detection and identification of Leishmania DNA and blood meal analysis in Phlebotomus (Larroussius) species. PLOS Neglected Tropical Diseases. 2020;14: e0008077. doi:10.1371/journal.pntd.0008077
99. Brazil R, Rodrigues AAF, Filho J. Sand Fly Vectors of Leishmania in the Americas - A Mini Review. Entomol Ornithol Herpetol. 2015;4.
100. Cantanhêde LM, Mattos CB, de Souza Ronconi C, Filgueira CPB, da Silva Júnior CF, Limeira C, et al. First report of Leishmania (Viannia) lindenbergi causing tegumentary leishmaniasis in the Brazilian western Amazon region. Parasite. 26: 30. doi:10.1051/parasite/2019030
101. Silveira FT, Ishikawa E a. Y, De Souza A a. A, Lainson R. An outbreak of cutaneous leishmaniasis among soldiers in Belém, Pará State, Brazil, caused by Leishmania (Viannia) lindenbergi n. sp. A new leishmanial parasite of man in the Amazon region. Parasite. 2002;9: 43–50. doi:10.1051/parasite/200209143
102. Testa JM, Montoya-Lerma J, Cadena H, Oviedo M, Ready PD. Molecular identification of vectors of Leishmania in Colombia: Mitochondrial introgression in the Lutzomyia townsendi series. Acta Tropica. 2002;84: 205–218. doi:10.1016/S0001-706X(02)00187-0
103. Feliciangeli MD, Murillo J. Lutzomyia youngi (Diptera: Psychodidae), a new phlebotomine sand fly previously misidentified as L. townsendi in endemic foci of cutaneous leishmaniasis in Venezuela and Costa Rica. J Med Entomol. 1987;24: 141–146. doi:10.1093/jmedent/24.2.141
104. Lozano-Sardaneta YN, Jacobo-Olvera E, Ruiz-Tovar K, Sánchez-Montes S, Rodríguez-Rojas JJ, Fernández-Figueroa EA, et al. Detection of Wolbachia and Leishmania DNA in sand flies (Diptera: Psychodidae, Phlebotominae) from a focus of cutaneous leishmaniasis in Tabasco, Mexico. Parasitol Res. 2022;121: 513–520. doi:10.1007/s00436-021-07412-4
105. Hashiguchi Y, Hashiguchi K, Zambrano FC, Parraga FD, Martillo VP, Torres EX, et al. Natural Leishmania (Leishmania) mexicana infection and biting activity of anthropophilic sand fly Lutzomyia ayacuchensis in the Ecuadorian Andes. Acta Trop. 2020;203: 105321. doi:10.1016/j.actatropica.2019.105321

Supplementary Legend:

Table S1: Trait table. All traits used in the sandfly vector model, along with definitions and data sources.

Fig S1: A histogram of AUC scores across all 100 BRT models for the primary model. The average AUC score was 0.851, and the median AUC score was 0.863. An AUC = 1.0 means the model is perfectly able to distinguish between the sandflies that are vectors and those that are not.

Table S2: Predicted probabilities for the primary model for sandflies assigned a probability score greater than 0.5.

Fig S2: Primary model partial dependence plots showing the marginal effect of each trait (shown in order of importance) on the probability of being a vector of ACL. The trait value is shown on the x-axis, and the importance is shown on the y-axis. Colored lines represent the marginal dependence of the trait from the 100 BRT models, while the solid black line represents the average dependence. The definition of each variable can be found in Table S1.

Table S3: Variable importance for the top 30 most important variables in the primary model, with categorical variables summed.

Fig S3: Variable importance of the primary model for the top 20 most important variables predicting sandfly vector status. Points represent mean gain value across 100 iterations and error bars represent 95% bootstrapped confidence intervals. Categorial variables are not summed here; each variable is left as it's own.

Fig S4: Occurrence points of predicted sandflies from the primary model, taken from GBIF and plotted in ArcGIS (Esri, USGS | Esri, Garmin, FAO, NOAA, USGS) [79, 80], colored by species to indicate species richness.

Fig S5: Occurrence points for confirmed vectors of each *Leishmania* species causing ACL, colored by sandfly species to indicate species richness. Points taken from GBIF and plotted in ArcGIS (Esri, USGS | Esri, Garmin, FAO, NOAA, USGS) [79, 80]. *Leishmania* species not mapped (e.g. *L. waltoni, L. lindenbergi, L. enrietti*) have no confirmed vectors.

Table S4: **Variable importance for the top 20 most important variables in the citation model**, without categorical variables summed. The trait profile is different compared to the primary model trait profile, ensuring that our primary model is not simply predicting which sandflies are well-studied.

Fig S6: A histogram of AUC scores for the secondary set of 100 BRT models using both potential and confirmed sandfly vectors as positive labels. The average AUC score was 0.867, and the median AUC score was 0.869.

Fig S7: A distribution of predicted probabilities from the secondary model of sandflies separated by vector status, and scaled by percentage. Red bars indicate the proportion of confirmed vectors that were predicted at that probability, while beige bars indicate the proportion of non-vector sandflies that were predicted at that probability.

Table S5: Predicted probabilities for the secondary model, for sandflies that are not confirmed vectors that have been assigned a probability score above the 90th percentile.

Table S6: Variable importance for the top 30 most important variables in the secondary model, with categorical variables summed.

Fig S8: Variable importance for the top 10 most important variables in the secondary model with 95% confidence intervals. Points represent mean gain value across 100 iterations. The importances for binary variables were summed up to obtain a single value for the entire categorical variable.

Fig S9: Variable importance for the top 20 most important variables in the secondary model predicting sandfly vector status. Points represent mean gain value across 100 iterations and error bars represent 95% bootstrapped confidence intervals. Categorial variables are not summed here; each variable is left as its own.

Fig S10: Secondary model partial dependence plots showing the marginal effect (yhat) of each trait (shown in order of importance) on the probability of being a vector of ACL. Variable value is shown on the x-axis, and marginal effect is shown on the y-axis. Partial dependence plots show the dependence of the probability on that trait's value, i.e., how the vector probability changes as the trait value increases.

Supplementary Materials:
All of the code and data used for this analysis can be found on this github repository:
https://github.com/mudkins/ACL-vector-data-analysis

Table S1: Trait table. All traits used in the sandfly vector model, along with definitions and data sources.

trait (as found in the data)	definition	source	coverage (%)
taxonomy			
tribe	Hertigiini or Phlebotomini (binary)	1	100
subtribe	Brumptomyiina, Hertigiina, Lutzomyiina, Psychodopygina, or Sergentomyiina (binary)	1	100
genus	Bichromomyia, Brumptomyia, Dampfomyia, Deanemyia, Evandromyia, Expapillata, Hertigia, Lutzomyia, Martinsmyia, Micropygomyia, Migonemyia, Nyssomyia, Oligodontomyia, Pintomyia, Pressatia, Psathyromyia, Psychodopygus, Sciopemyia, Trichophoromyia, Trichopygomyia, Viannamyia, Warileya (binary)	1	100
female morphology			
labruml_wingl	ratio labrum length/wing length	2	17.38
A3_wingl	ratio length antennal segment 3/wing length	2, 3	17.38
no.lat.teeth	number of lateral teeth	2, 3	21.48
wing.length	wing length	2, 3	29.88
wingl_wingw	ratio wing length/wing width	2	29.88
wing.width	wing width at widest part of wing	2	13.87
labrum.length	labrum length: part of feeding fascicle, determines depth of skin and penetration	2, 3, 4	22.27
A3	length antennal segment 3	2, 3	6.6
number ventrad teeth	number of ventrad (internal) teeth	3	20.9
number cibarium teeth	number horizontal cibarium teeth	2, 3	32.03
dental.depth	distance from tip of maxilla to most prominent ventral tooth	2, 3	17.38
max.shape	shape of maxillary tip: either spear or sabre	2	22.46

	(binary)		
hypo.teeth	structure of hypopharyngeal teeth: smooth, rough, or spiculate (binary)	2	22.46
biogeography/ecology			
biomes	One or more of Tropical & Subtropical Moist Broadleaf Forests; Tropical & Subtropical Dry Broadleaf Forests; Tropical & Subtropical Grasslands, Savannas & Shrublands; Mangroves; Flooded Grasslands & Savannas; Deserts & Xeric Shrublands; Tropical & Subtropical Coniferous Forests; Temperate Grasslands, Savannas & Shrublands; Temperate Conifer Forests; Temperate Broadleaf & Mixed Forests, Montane Grasslands & Shrublands (binary)	5	41.8
ecoregion.breadth	number of distinct ecoregions the sandfly occurs in	5	41.6
temp	mean monthly temperature of species range from 2009-2019	6	41.6
temp.range	mean of monthly temperature range of species range from 2009-2019	6	41.6
rainfall	mean monthly rainfall of species range from 2009-2019	6	41.6
wind.speed	mean monthly wind speed of species range from 2009-2019	6	41.6
flii	forest land integrity index: measure of forest pressure and connectivity, index of forest integrity determined by degree of anthropogenic modification	7	32.81
elevation	mean elevation of species range from 2009-2019	8	41.6
canopy	global tree heights based on a fusion of spaceborne-lidar data (2005) from the Geoscience Laser Altimeter System (GLAS) and ancillary geospatial data	9	41.6
tree.cover	mean treecover of species range from 2009-2019	10	41.6
ghm	global human modification: the cumulative measure of human modification of terrestrial lands globally at 1 square-kilometer resolution	11	41.6
evi	enhanced vegetation index: 'optimized'	12	41.6

	vegetation index designed to enhance the vegetation signal with improved sensitivity		
crops.cover	percent vegetation cover for cropland land cover	6	41.6
grass.cover	percent vegetation cover for herbaceous vegetation land cover	6	41.6
shrub.cover	percent vegetation cover for shrubland land cover	6	41.6
urban.cover	percent vegetation cover for built-up land cover	6	41.6
water.perm.cover	percent ground cover for permanent water land cover	6	41.6
water.seas.cover	percent ground cover for seasonal water land cover	6	41.6
activity	time of day the sandfly is active: diurnal, noctural, crepuscular (binary)	3	crep (1.56), noct (2.53), diurnal (2.73)
habitat strata	detected in canopy and/or detected on forest floor (binary)	13-54	canopy (13.28), floor (14.06)
main habitat	fallen leaves in forest soil, armadillo burrows, burrows of other wild animals, tree trunks and tabular roots, tree hollows, treetops, crevices in rocks, caves, forest without specific location, marginal areas, annexes of domestic animals, outer and inner walls of human dwellings	51	54.3
proximity to house	captured in extra domicile, peri domicile, and/or intra domicile environments (binary)	13-54	intra (9.37), peri (17.97), extra (21.48)
synanthropy	(1) wild: living in forests or in non-forest regions but only accidentally found associated with humans and domestic animals; (2) semi-domestic: living outside human and domestic animal habitations and only seeking these to obtain blood repast; and (3) domestic: living in association with humans and domestic animals inside or near dwellings	13-54	wild (21.68), semi-domestic (17.38), domestic (10.74)
seasonal activity	whether the sandfly is more active during the winter or summer (binary)	13-54	winter (4.69), summer (4.88)
synanthropy index	synanthropy index, as defined by Barrios et al [40], on a scale from -100 (preference for rural environment) to 100 (preference for	13-54	5.98

	urban environment)		

Primary model: The primary model was trained on confirmed vectors of ACL.

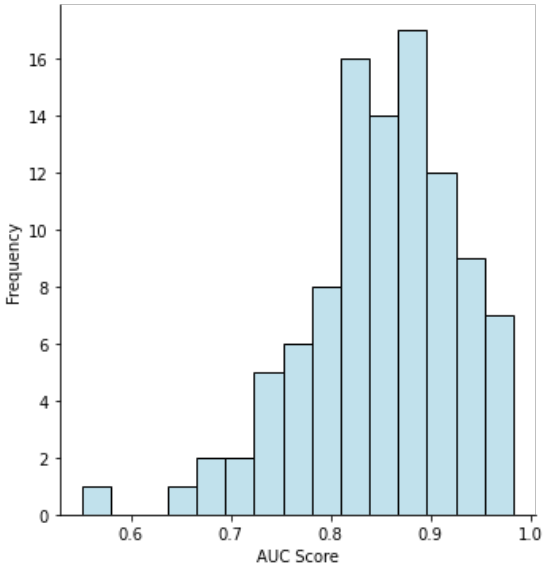


Fig S1: A histogram of AUC scores across all 100 BRT models for the primary model. The average AUC score was 0.851, and the median AUC score was 0.863. An AUC = 1.0 means the model is perfectly able to distinguish between the sandflies that are vectors and those that are not.

Table S2: Predicted probabilities for the primary model for sandflies assigned a probability score greater than 0.5.

species	probability	std	percentile	potential/proven
Bichromomyia flaviscutellata	0.959	0.079	1.0	proven
Nyssomyia whitmani	0.936	0.087	0.998	proven
Psychodopygus davisi	0.932	0.082	0.996	proven
Psychodopygus carrerai	0.932	0.086	0.993	proven
Nyssomyia intermedia	0.932	0.089	0.993	proven
Psathyromyia (Psathyromyia) shannoni	0.923	0.111	0.99	proven
Psychodopygus wellcomei	0.917	0.133	0.988	proven
Lutzomyia (Lutzomyia) longipalpis	0.917	0.1	0.986	proven
Nyssomyia neivai	0.907	0.127	0.984	proven
Psychodopygus ayrozai	0.906	0.087	0.982	proven
Nyssomyia yuilli	0.905	0.096	0.98	proven
Psychodopygus panamensis	0.902	0.112	0.979	proven
Psychodopygus hirsutus	0.901	0.09	0.977	proven
Migonemyia (Migonemyia) migonei	0.89	0.141	0.975	proven
Nyssomyia trapidoi	0.888	0.106	0.973	proven
Psychodopygus complexus	0.87	0.12	0.971	proven
Psychodopygus amazonensis	0.868	0.122	0.969	potential
Trichophoromyia ubiquitalis	0.867	0.163	0.967	proven
Pintomyia (Pintomyia) fischeri	0.867	0.187	0.965	proven
Nyssomyia antunesi	0.852	0.122	0.963	potential
Nyssomyia anduzei	0.84	0.172	0.961	proven
Nyssomyia umbratilis	0.839	0.128	0.959	proven
Psychodopygus claustrei	0.838	0.109	0.957	unknown
Pintomyia (Pifanomyia) verrucarum	0.837	0.209	0.955	proven
Psychodopygus paraensis	0.833	0.218	0.953	proven
Psychodopygus squamiventris maripaensis	0.829	0.149	0.951	proven
Psychodopygus llanosmartinsi	0.817	0.168	0.949	proven
Pintomyia nuneztovari	0.815	0.218	0.947	proven
Lutzomyia (Helcocyrtomyia) hartmanni	0.77	0.189	0.945	proven
Pintomyia (Pifanomyia) spinicrassa	0.766	0.22	0.943	proven
Nyssomyia shawi	0.764	0.208	0.941	proven
Bichromomyia olmeca olmeca	0.75	0.191	0.939	proven
Psychodopygus guyanensis	0.748	0.133	0.938	unknown
Pintomyia (Pintomyia) pessoai	0.739	0.199	0.936	potential
Lutzomyia (Tricholateralis) gomezi	0.735	0.257	0.934	proven
Lutzomyia (Helcocyrtomyia) peruensis	0.73	0.261	0.932	proven
Psathyromyia (Psathyromyia) bigeniculata	0.725	0.15	0.93	unknown
Trichophoromyia auraensis	0.702	0.195	0.928	potential
Bichromomyia reducta	0.676	0.252	0.926	proven
Bichromomyia olmeca nociva	0.668	0.26	0.924	proven
Psychodopygus chagasi	0.592	0.205	0.922	unknown
Trichophoromyia castanheirai	0.584	0.213	0.92	unknown
Sciopemyia sordellii	0.578	0.186	0.918	potential
Psathyromyia (Psathyromyia) lanei	0.556	0.182	0.916	unknown
Evandromyia (Evandromyia) infraspinosa	0.541	0.207	0.914	unknown
Psychodopygus squamiventris squamiventris	0.532	0.2	0.912	proven
Lutzomyia (Helcocyrtomyia) ayacuchensis	0.522	0.246	0.91	proven
Warileya rotundipennis	0.511	0.217	0.908	unknown

Fig S2: Primary model partial dependence plots showing the marginal effect of each trait (shown in order of importance) on the probability of being a vector of ACL. The trait value is shown on the x-axis, and the importance is shown on the y-axis. Colored lines represent the marginal dependence of the trait from the 100 BRT models, while the solid black line represents the average dependence. The definition of each variable can be found in Table S1.

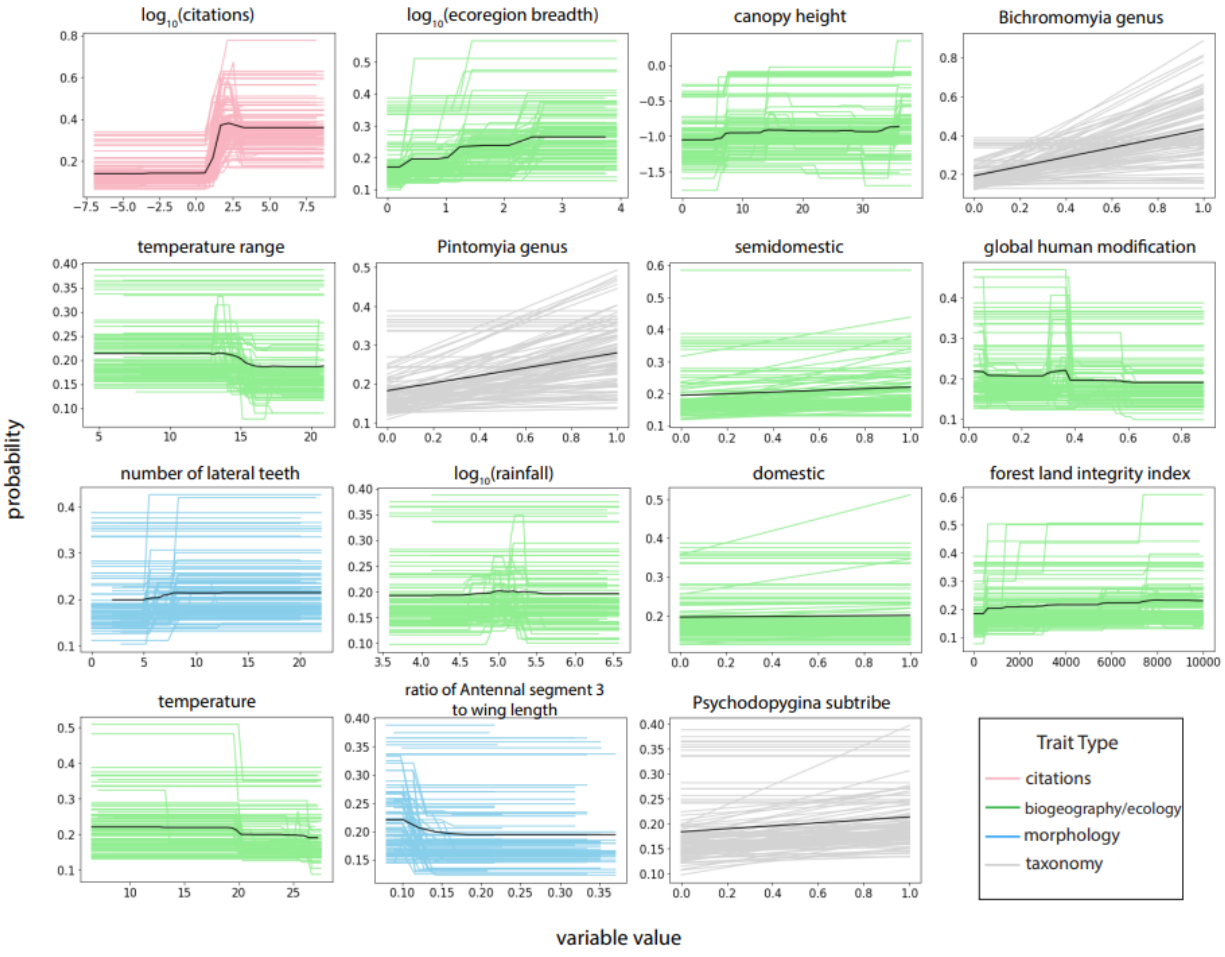


Table S3: Variable importance for the top 30 most important variables in the primary model, with categorical variables summed.

feature	importance
log(citations)	0.218
genus	0.116
log(ecoregion breadth)	0.092
canopy height	0.071
synanthropy	0.068
main habitat	0.061
temperature range	0.045
global human modification	0.036
number of lateral teeth	0.032
log(rainfall)	0.032
biome	0.032
subtribe	0.029
forest land integrity index	0.028
temperature	0.026
ratio of Antennal segment 3 to wing length	0.022
enhanced vegetation index	0.013
log(grass cover)	0.012
log(tree cover)	0.011
log(wind speed)	0.011
log(number of ventrad teeth)	0.008
wing length	0.007
ratio of wing length to wing width	0.007
log(crops cover)	0.006
log(urban cover)	0.005
spear shape of maxillary tip	0.004
log(number of cibarium teeth)	0.003
log(labrum length)	0.003
log(seasonal water cover)	0.002
habitat strata	0.002

Fig S3: Variable importance of the primary model for the top 20 most important variables predicting sandfly vector status. Points represent mean gain value across 100 iterations and error bars represent 95% bootstrapped confidence intervals. Categorial variables are not summed here; each variable is left as it's own.

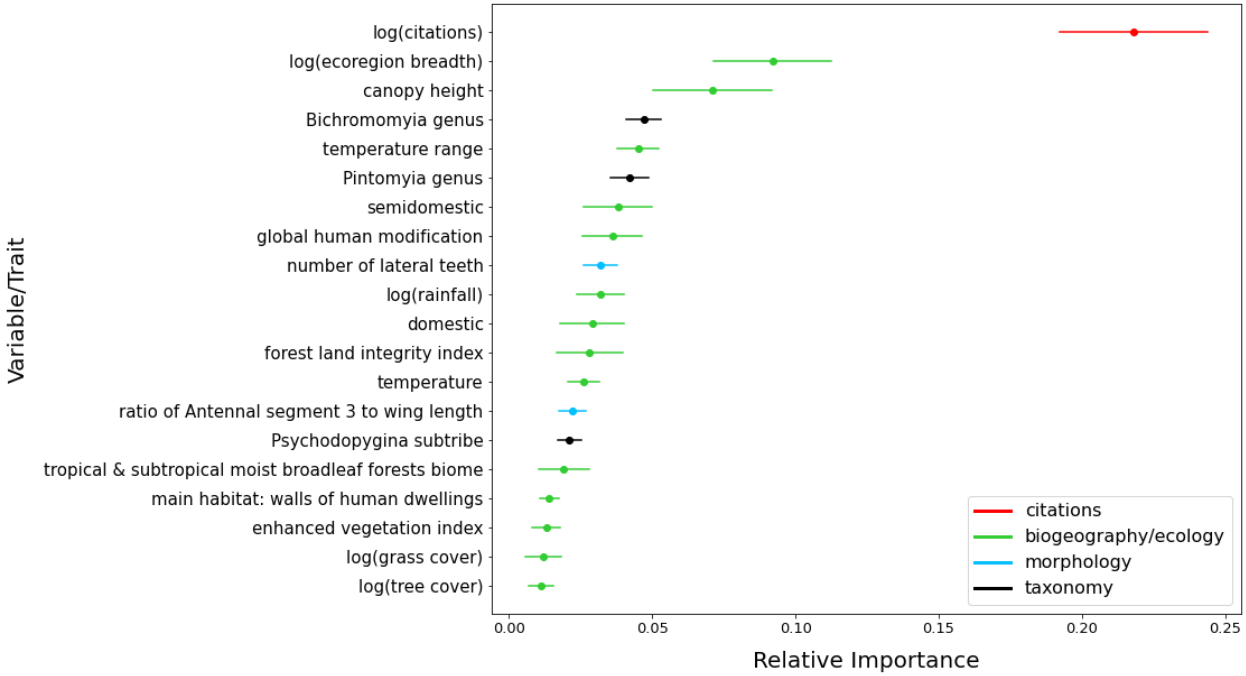


Fig S4: Occurrence points of predicted sandflies from the primary model, taken from GBIF and plotted in ArcGIS (Esri, USGS | Esri, Garmin, FAO, NOAA, USGS) [79, 80], colored by species to indicate species richness.

Fig S5: Occurrence points for confirmed vectors of each *Leishmania* species causing ACL, colored by sandfly species to indicate species richness. Points taken from GBIF and plotted in ArcGIS (Esri, USGS | Esri, Garmin, FAO, NOAA, USGS) [79, 80]. *Leishmania* species not mapped (e.g. *L. waltoni, L. lindenbergi, L. enrietti*) have no confirmed vectors.

L. amazonensis

L. braziliensis

L. guyanensis

L. lainsoni

L. mexicana

L. naiffi

L. panamensis

L. peruviana

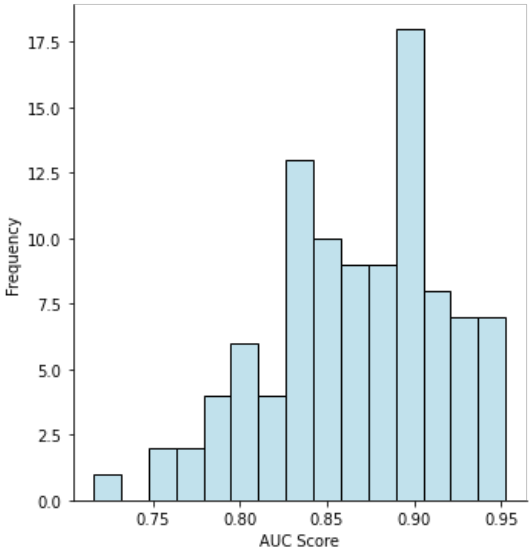
L. shawi

Citation model: The citation model was trained and fit to determine whether sandfly traits were predicting citations or vector status.

Table S4: Variable importance for the top 20 most important variables in the citation model, without categorical variables summed. The trait profile is different compared to the primary model trait profile, ensuring that our primary model is not simply predicting which sandflies are well-studied.

feature	importance
temperature	0.055
wing length	0.051
log(labrum length)	0.042
ratio of wing length to wing width	0.042
log(crops cover)	0.037
log(tree cover)	0.036
number of lateral teeth	0.034
temperature range	0.033
log(wind speed)	0.033
Psathyromyia genus	0.031
forest land integrity index	0.028
log(number of cibarium teeth)	0.028
global human modification	0.026
spiculate structure of hypopharyngeal teeth	0.026
semi domestic	0.025
log(grass cover)	0.023
log(rainfall)	0.023
Trichophoromyia genus	0.022
main habitat: marginal areas	0.02
canopy height	0.02

Secondary Model: The secondary model was trained and fit identically to the primary model in the text but using predicted, rather than confirmed, sandfly vector species as positive labels. We compared predicted sandfly vector species and trait profiles of the primary and secondary models as well. The secondary model was able to generally predict which sandflies have the potential to carry *Leishmania* spp., while the primary model more specifically predicts which sandflies can transmit *Leishmania spp.* to humans.



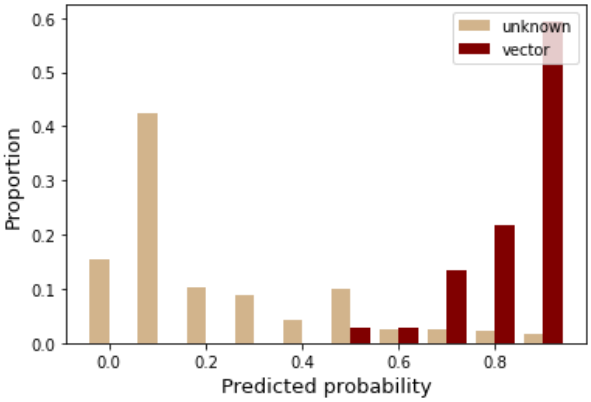


Fig S6 (left): A histogram of AUC scores for the secondary set of 100 BRT models using both potential and confirmed sandfly vectors as positive labels. The average AUC score was 0.867, and the median AUC score was 0.869.

Fig S7 (right): A distribution of predicted probabilities from the secondary model of sandflies separated by vector status, and scaled by percentage. Red bars indicate the proportion of confirmed vectors that were predicted at that probability, while beige bars indicate the proportion of non-vector sandflies that were predicted at that probability.

Table S5: Predicted probabilities for the secondary model, for sandflies that are not confirmed vectors that have been assigned a probability score above the 90th percentile.

species	probability	std	percentile	infection status
Nyssomyia antunesi	0.963	0.106	0.99	potential
Pintomyia (Pintomyia) pessoai	0.963	0.096	0.988	potential
Trichophoromyia auraensis	0.94	0.123	0.975	potential
Psychodopygus amazonensis	0.936	0.105	0.971	potential
Sciopemyia sordellii	0.934	0.096	0.969	potential
Lutzomyia (Lutzomyia) cruzi	0.923	0.112	0.955	potential
Evandromyia (Aldamyia) lenti	0.919	0.136	0.953	potential
Micropygomyia (Sauromyia) trinidadensis	0.903	0.118	0.945	potential
Lutzomyia (Lutzomyia) lichyi	0.9	0.149	0.941	potential
Lutzomyia (Tricholateralis) cruciata	0.899	0.157	0.939	potential
Micropygomyia (Sauromyia) villelai	0.894	0.129	0.936	potential
Pintomyia (Pifanomyia) monticola	0.879	0.191	0.932	potential
Evandromyia edwardsi	0.878	0.161	0.93	potential
Lutzomyia (Helcocyrtomyia) noguchii	0.874	0.159	0.926	potential
Martinsmyia minasensis	0.846	0.208	0.92	potential
Micropygomyia (Sauromyia) peresi	0.839	0.19	0.918	potential
Psychodopygus lloydi	0.835	0.15	0.916	potential
Lutzomyia (Tricholateralis) diabolica	0.822	0.229	0.914	potential
Micropygomyia (Sauromyia) capixaba	0.789	0.22	0.902	potential

Table S6: Variable importance for the top 30 most important variables in the secondary model, with categorical variables summed.

feature	importance
genus	0.145
main habitat	0.109
synanthropy	0.09
log(ecoregion breadth)	0.089
log(crops cover)	0.054
spear shape of maxillary tip	0.05
log(citations)	0.047
temperature	0.035
global human modification	0.032
log(grass cover)	0.03
log(number of cibarium teeth)	0.029
number of lateral teeth	0.025
forest land integrity index	0.024
ratio of Antennal segment 3 to wing length	0.022
subtribe	0.02
biome	0.019
log(tree cover)	0.018
log(labrum length)	0.017
log(urban cover)	0.017
wingl_wingw	0.016
log.no.ven.teeth	0.015
evi	0.014
temp.range	0.013
log.wind.speed	0.013
canopy	0.012
log.rainfall	0.012
wing.length	0.01
bites.mammals	0.005
log.water.seas.cover	0.004
hypopharyngeal teeth structure	0.004
log.water.perm.cover	0.003
habitat strata	0.002

Fig S8: Variable importance for the top 10 most important variables with 95% confidence intervals. Points represent mean gain value across 100 iterations. The importances for binary variables were summed up to obtain a single value for the entire categorical variable.

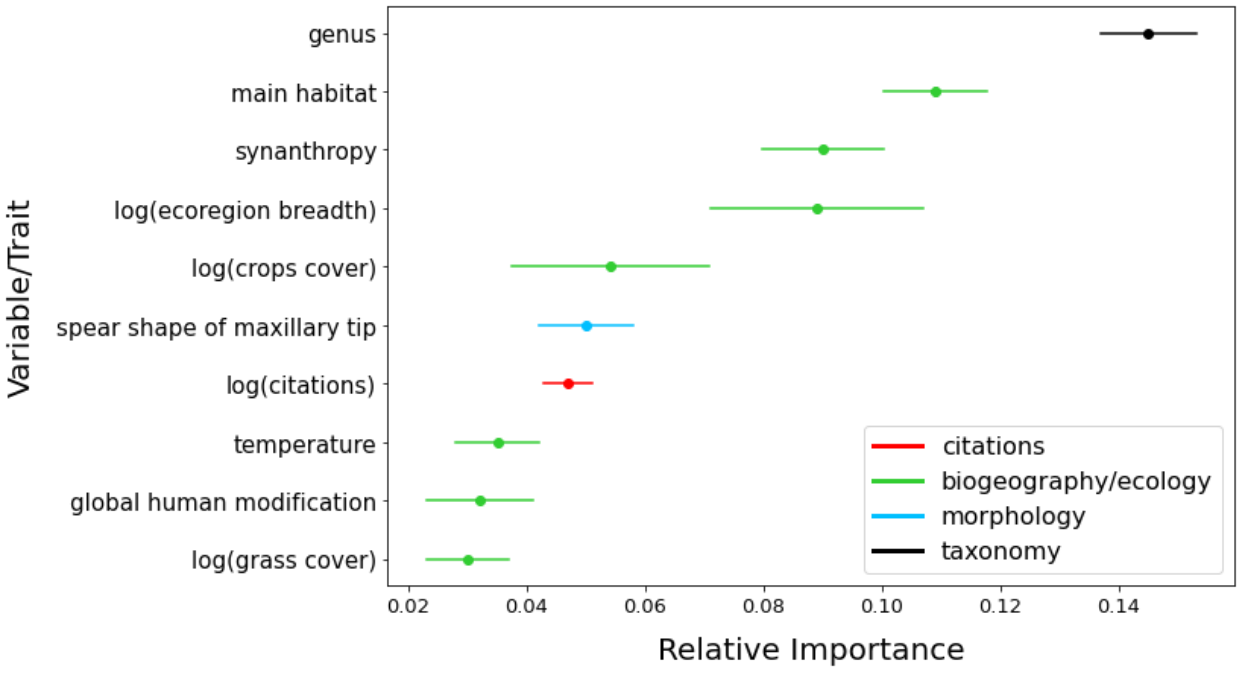


Fig S9: Variable importance for the top 20 most important variables predicting sandfly vector status. Points represent mean gain value across 100 iterations and error bars represent 95% bootstrapped confidence intervals. Categorial variables are not summed here; each variable is left as it's own.

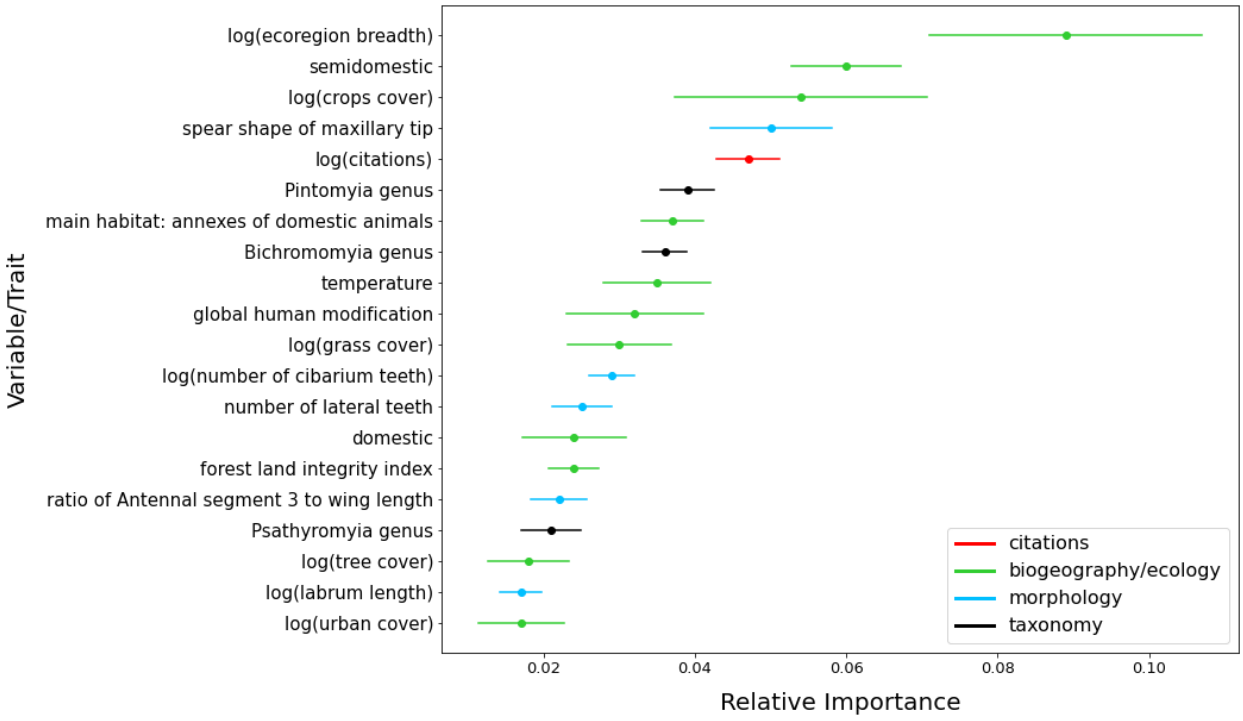
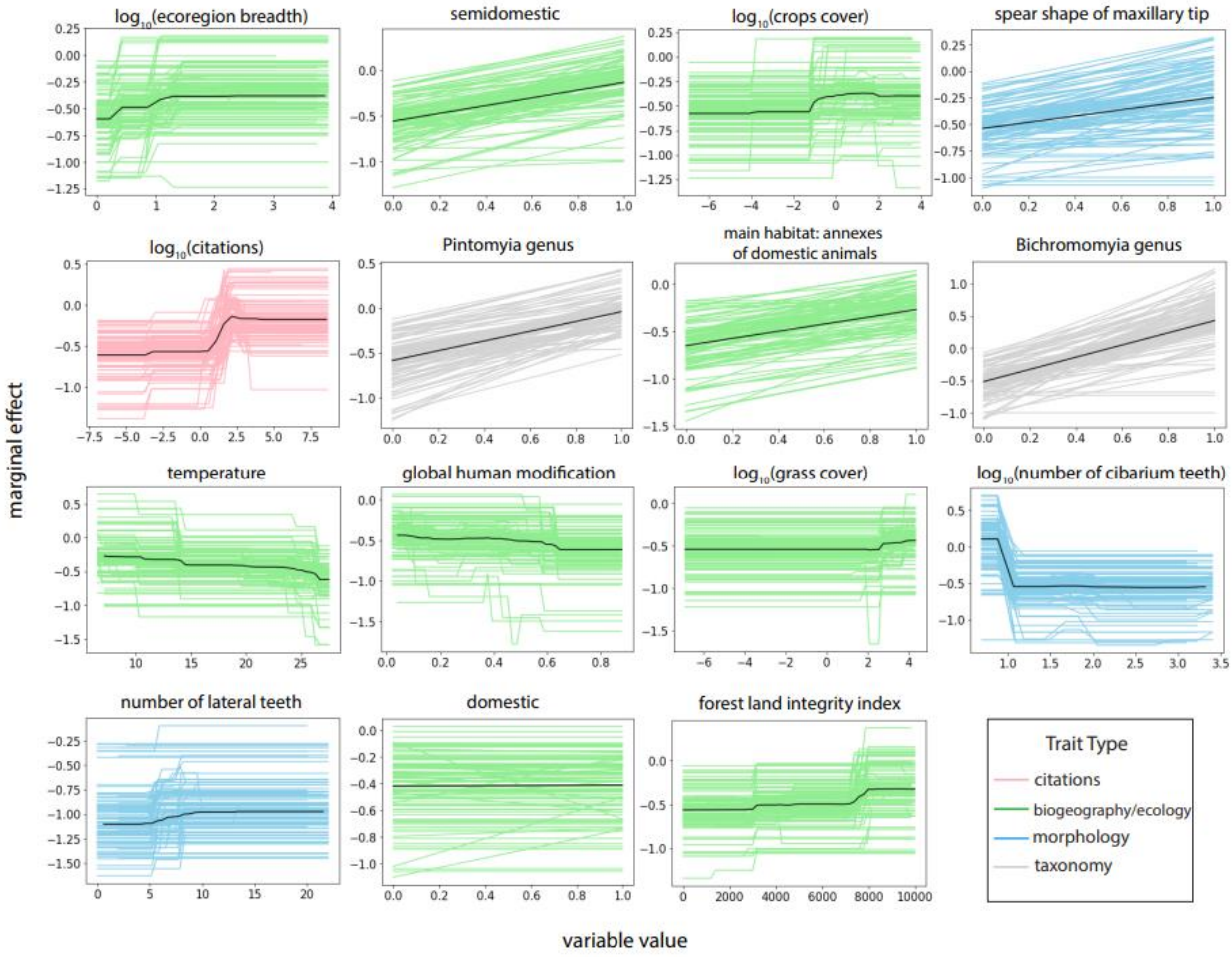


Fig S10: Secondary model partial dependence plots showing the marginal effect (yhat) of each trait (shown in order of importance) on the probability of being a vector of ACL. Variable value is shown on the x-axis, and marginal effect is shown on the y-axis. Partial dependence plots show the dependence of the probability on that trait's value, i.e. how the vector probability changes as the trait value increases.


log10(ecoregion breadth)
semidomestic
log10(crops cover)
spear shape of maxillary tip
log10(citations)
Pintomyia genus
main habitat: annexes of domestic animals
Bichromomyia genus
temperature
global human modification
log10(grass cover)
log10(number of cibarium teeth)
number of lateral teeth
domestic
forest land integrity index
Trait Type
citations
biogeography/ecology
morphology
taxonomy
marginal effect
variable value

Supplementary Materials References:

1. Shimabukuro PHF, Andrade AJ de, Galati EAB. Checklist of American sand flies (Diptera, Psychodidae, Phlebotominae): genera, species, and their distribution. ZooKeys. 2017;660: 67–106. doi:10.3897/zookeys.660.10508
2. Lewis DJ. Functional morphology of the mouth parts in New World phlebotomine sandflies (Diptera: Psychodidae). Transactions of the Royal Entomological Society of London. 1975;126: 497–532. doi:10.1111/j.1365-2311.1975.tb00859.x
3. Young D. G. Duncan M. A.. 1994. Guide to the identification and geographic distribution of Lutzomyia sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae). Mem. Am. Entomol. Inst. 54: 1–881.
4. Filho J, Galati E, Falcão A. Redescription of Nyssomyia intermedia (Lutz & Neiva, 1912) and Nyssomyia neivai (Pinto, 1926) (Diptera: Psychodidae). Memórias do Instituto Oswaldo Cruz. 2004;98: 1059–65. doi:10.1590/S0074-02762003000800015
5. Bioscience, An Ecoregions-Based Approach to Protecting Half the Terrestrial Realm doi:10.1093/biosci/bix014
6. Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), (cited 18 April 2022), https://cds.climate.copernicus.eu/cdsapp#!/home
7. Grantham HS, Duncan A, Evans TD, Jones KR, Beyer HL, Schuster R, et al. Anthropogenic modification of forests means only 40% of remaining forests have high ecosystem integrity. Nat Commun. 2020;11: 5978. doi:10.1038/s41467-020-19493-3
8. Amante, C. and B. W. Eakins, ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24, 19 pp, March 2009.
9. Simard, M., Pinto, N., Fisher, J., Baccini, A. 2011. Mapping forest canopy height globally with spaceborne lidar. Journal of Geophysical Research. 116: G04021. doi:10.1029/2011JG001708
10. Buchhorn, M. ; Lesiv, M. ; Tsendbazar, N. - E. ; Herold, M. ; Bertels, L. ; Smets, B. Copernicus Global Land Cover Layers—Collection 2. Remote Sensing 2020, 12Volume 108, 1044. doi:10.3390/rs12061044
11. Kennedy, C.M., J.R. Oakleaf, D.M. Theobald, S. Baurch-Murdo, and J. Kiesecker. 2019. Managing the middle: A shift in conservation priorities based on the global human modification gradient. Global Change Biology 00:1-16. doi:10.1111/gcb.14549
12. Didan, K.. MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006. 2015, distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MOD13A2.006. Accessed 2022-04-21.
13. Moreno M, Ferro C, Rosales-Chilama M, Rubiano L, Delgado M, Cossio A, et al. First report of Warileya rotundipennis (Psychodidae: Phlebotominae) naturally infected with Leishmania (Viannia) in a focus of cutaneous leishmaniasis in Colombia. Acta Tropica. 2015;148: 191–196. doi:10.1016/j.actatropica.2015.04.017
14. Barreto M. P.. 1943. Observações sobre a Biologia em Condições Naturais, dos Flebótomos de São Paulo (Diptera: Psychodidae). Tipografia Rossolino, São Paulo, Brazil.

15. Barrios SPG, Pereira LE, Monaco NZN, Graciolli G, Casaril AE, Infran J de OM, et al. Synanthropy and diversity of Phlebotominae in an area of intense transmission of visceral leishmaniasis in the South Pantanal floodplain, Midwest Brazil. PLOS ONE. 2019;14: e0215741. doi:10.1371/journal.pone.0215741
16. Alves GB, Oshiro ET, Leite M da C, Melão AV, Ribeiro LM, Mateus NLF, et al. Phlebotomine sandflies fauna (Diptera: Psychodidae) at rural settlements in the municipality of Cáceres, State of Mato Grosso, Brazil. Rev Soc Bras Med Trop. 2012;45: 437–443. doi:10.1590/s0037-86822012005000010
17. Rodrigues E de AS, Andrade Filho JD, Limongi JE, Paula MBC de. Sandfly fauna (Diptera: Psychodidae) in Parque do Sabiá complex, Uberlândia, Minas Gerais, Brazil. Rev Inst Med trop S Paulo. 2011;53: 255–258. doi:10.1590/S0036-46652011000500003
18. Dorval MEC, Alves TP, Cristaldo G, Rocha HC da, Alves MA, Oshiro ET, et al. Sand fly captures with Disney traps in area of occurrence of Leishmania (Leishmania) amazonensis in the state of Mato Grosso do Sul, mid-western Brazil. Rev Soc Bras Med Trop. 2010;43: 491–495. doi:10.1590/S0037-86822010000500003
19. Oliveira AG de, Galati EAB, Oliveira O de, Oliveira GR de, Espindola IAC, Dorval MEC, et al. Abundance of Lutzomyia longipalpis (Diptera: Psychodidae: Phlebotominae) and urban transmission of visceral leishmaniasis in Campo Grande, state of Mato Grosso do Sul, Brazil. Mem Inst Oswaldo Cruz. 2006;101: 869–874. doi:10.1590/S0074-02762006000800008
20. Salomón OD, Rossi GC, Cousiño B, Spinelli GR, Rojas de Arias A, López del Puerto DG, et al. Phlebotominae sand flies in Paraguay: abundance distribution in the Southeastern region. Mem Inst Oswaldo Cruz. 2003;98: 185–190. doi:10.1590/S0074-02762003000200004
21. Senne NA, Vilela TS, Sanavria A, Santos HA, Rabello RS, Angelo IC. Ecology and spatial distribution of sand fly species in low endemic areas for American Tegumentary Leishmaniasis in the municipality of Seropédica, Rio de Janeiro, Brazil. Medical and Veterinary Entomology. 2021;35: 371–378. doi:10.1111/mve.12505
22. Sales KG da S, de Oliveira Miranda DE, Costa PL, da Silva FJ, Figueredo LA, Brandão-Filho SP, et al. Home sweet home: sand flies find a refuge in remote indigenous villages in north-eastern Brazil, where leishmaniasis is endemic. Parasites & Vectors. 2019;12: 118. doi:10.1186/s13071-019-3383-1
23. Pinheiro MPG, Silva MM de M, Júnior JBS, da Silva JHT, Alves M de L, Ximenes M de FF de M. Sand flies (Diptera, Psychodidae, Phlebotominae), vectors of Leishmania protozoa, at an Atlantic Forest Conservation Unit in the municipality of Nísia Floresta, Rio Grande do Norte state, Brazil. Parasites & Vectors. 2016;9: 83. doi:10.1186/s13071-016-1352-5
24. Staniek ME, Hamilton JGC. Odour of domestic dogs infected with Leishmania infantum is attractive to female but not male sand flies: Evidence for parasite manipulation. PLOS Pathogens. 2021;17: e1009354. doi:10.1371/journal.ppat.1009354
25. Rêgo FD, Rugani JMN, Shimabukuro PHF, Tonelli GB, Quaresma PF, Gontijo CMF. Molecular Detection of Leishmania in Phlebotomine Sand Flies (Diptera: Psychodidae) from a Cutaneous Leishmaniasis Focus at Xakriabá Indigenous Reserve, Brazil. PLOS ONE. 2015;10: e0122038. doi:10.1371/journal.pone.0122038
26. Rêgo FD, Shimabukuro PHF, Quaresma PF, Coelho IR, Tonelli GB, Silva KMS, et al. Ecological aspects of the Phlebotominae fauna (Diptera: Psychodidae) in the Xakriabá Indigenous Reserve, Brazil. Parasites Vectors. 2014;7: 220. doi:10.1186/1756-3305-7-220

27. Carvalho BM, Maximo M, Costa WA, de Santana ALF, da Costa SM, da Costa Rego TAN, et al. Leishmaniasis transmission in an ecotourism area: potential vectors in Ilha Grande, Rio de Janeiro State, Brazil. Parasites & Vectors. 2013;6: 325. doi:10.1186/1756-3305-6-325
28. Kato H, Gomez EA, Cáceres AG, Vargas F, Mimori T, Yamamoto K, et al. Natural Infections of Man-Biting Sand Flies by Leishmania and Trypanosoma Species in the Northern Peruvian Andes. Vector-Borne and Zoonotic Diseases. 2011;11: 515–521. doi:10.1089/vbz.2010.0138
29. Saraiva L, Reis AS, Rugani JMN, Pereira AAS, Rêgo FD, Lima ACVM da R, et al. Survey of Sand Flies (Diptera: Psychodidae) in an Environmentally Protected Area in Brazil. PLOS ONE. 2015;10: e0134845. doi:10.1371/journal.pone.0134845
30. Zorrilla V, Santos MBDL, Espada L, Santos R del P, Fernandez R, Urquia A, et al. Distribution and identification of sand flies naturally infected with Leishmania from the Southeastern Peruvian Amazon. PLOS Neglected Tropical Diseases. 2017;11: e0006029. doi:10.1371/journal.pntd.0006029
31. Toro-Cantillo A, Atencia Pineda M, Hoyos R. Flebotomíneos (Diptera: Psychodidae) colectados en área rural de San Bernardo del Viento (Córdoba Colombia). Revista MVZ Córdoba. 2017;22: 6044. doi:10.21897/rmvz.1074
32. Pereira Júnior AM, Souza ABN, Castro TS, da Silva MS, de Paulo PFM, Ferreira GEM, et al. Diversity, natural infection and blood meal sources of phlebotomine sandflies (Diptera, Psychodidae) in the western Brazilian Amazon. Mem Inst Oswaldo Cruz. 2019;114: e190170. doi:10.1590/0074-02760190170
33. de Ávila MM, Brilhante AF, de Souza CF, Bevilacqua PD, Galati EAB, Brazil RP. Ecology, feeding and natural infection by Leishmania spp. of phlebotomine sand flies in an area of high incidence of American tegumentary leishmaniasis in the municipality of Rio Branco, Acre, Brazil. Parasites Vectors. 2018;11: 64. doi:10.1186/s13071-018-2641-y
34. Vasconcelos dos Santos T, Silva F, Barata I, Andrade A, Galati E. A new species of phlebotomine, Trichophoromyia adelsonsouzai (Diptera: Psychodidae) of Brazilian Amazonia. Memorias do Instituto Oswaldo Cruz. 2013;0: 0. doi:10.1590/0074-0276130159
35. Azevedo ACR, Souza NA, Meneses CRV, Costa WA, Costa SM, Lima JB, et al. Ecology of sand flies (Diptera: psychodidae: phlebotominae) in the north of the state of Mato Grosso, Brazil. Mem Inst Oswaldo Cruz. 2002;97: 459–464. doi:10.1590/s0074-02762002000400002
36. Machado TDO, Minuzzi-Souza TTC, Ferreira T de S, Freire LP, Timbó RV, Vital TE, et al. The role of gallery forests in maintaining Phlebotominae populations: potential *Leishmania* spp. vectors in the Brazilian savanna. Mem Inst Oswaldo Cruz. 2017;112: 681–691. doi:10.1590/0074-02760170126
37. Ovallos FG, Silva YRE, Fernandez N, Gutierrez R, Galati EAB, Sandoval CM. The sandfly fauna, anthropophily and the seasonal activities of *Pintomyia spinicrassa* (Diptera: Psychodidae: Phlebotominae) in a focus of cutaneous leishmaniasis in northeastern Colombia. Mem Inst Oswaldo Cruz. 2013;108: 297–302. doi:10.1590/S0074-02762013000300007
38. Brilhante AF, de Ávila MM, de Souza JF, Medeiros-Sousa AR, Sábio PB, de Paula MB, et al. Attractiveness of black and white modified Shannon traps to phlebotomine sandflies (Diptera, Psychodidae) in the Brazilian Amazon Basin, an area of intense transmission of American cutaneous leishmaniasis. Parasite. 24: 20. doi:10.1051/parasite/2017021
39. Dorval MEC, Cristaldo G, Rocha HC da, Alves TP, Alves MA, Oshiro ET, et al. Phlebotomine fauna (Diptera: Psychodidae) of an American cutaneous leishmaniasis endemic area in the state of Mato

Grosso do Sul, Brazil. Mem Inst Oswaldo Cruz. 2009;104: 695–702. doi:10.1590/s0074-02762009000500005

40. Baum M, Ribeiro MCV da C, Lorosa ES, Damasio GAC, Castro EA de. Eclectic feeding behavior of Lutzomyia (Nyssomyia) intermedia (Diptera, Psychodidae, Phlebotominae) in the transmission area of American cutaneous leishmaniasis, state of Paraná, Brazil. Rev Soc Bras Med Trop. 2013;46: 560–565. doi:10.1590/0037-8682-0157-2013
41. Alves GB, Oshiro ET, Leite M da C, Melão AV, Ribeiro LM, Mateus NLF, et al. Phlebotomine sandflies fauna (Diptera: Psychodidae) at rural settlements in the municipality of Cáceres, State of Mato Grosso, Brazil. Rev Soc Bras Med Trop. 2012;45: 437–443. doi:10.1590/s0037-86822012005000010
42. Galati EAB, Marassá AM, Gonçalves-Andrade RM, Consales CA, Bueno EFM. Phlebotomines (Diptera, Psychodidae) in the Ribeira Valley Speleological Province - 1. Parque Estadual Intervales, state of São Paulo, Brazil. Rev Bras entomol. 2010;54: 311–321. doi:10.1590/S0085-56262010000200015
43. Balbino VQ, Coutinho-Abreu IV, Sonoda IV, Marques da Silva W, Marcondes CB. Phlebotomine sandflies (Diptera: Psychodidae) of the Atlantic forest in Recife, Pernambuco state, Brazil: the species coming to human bait, and their seasonal and monthly variations over a 2-year period. Ann Trop Med Parasitol. 2005;99: 683–693. doi:10.1179/136485905X65116
44. Pereira Filho AA, Bandeira M da CA, Fonteles RS, Moraes JLP, Lopes CRG, Melo MN, et al. An ecological study of sand flies (Diptera: Psychodidae) in the vicinity of Lençóis Maranhenses National Park, Maranhão, Brazil. Parasites & Vectors. 2015;8: 442. doi:10.1186/s13071-015-1045-5
45. Bray DP, Alves GB, Dorval ME, Brazil RP, Hamilton JG. Synthetic sex pheromone attracts the leishmaniasis vector Lutzomyia longipalpis to experimental chicken sheds treated with insecticide. Parasit Vectors. 2010;3: 16. doi:10.1186/1756-3305-3-16
46. Marassá AM, Galati EAB, Bergamaschi DP, Consales CA. Blood feeding patterns of *Nyssomyia intermedia* and *Nyssomyia neivai* (Diptera, Psychodidae) in a cutaneous leishmaniasis endemic area of the Ribeira Valley, State of São Paulo, Brazil. Rev Soc Bras Med Trop. 2013;46: 547–554. doi:10.1590/0037-8682-0168-2013
47. Alves VR, Freitas RA de, Santos FL, Barrett TV. Diversity of sandflies (Psychodidae: Phlebotominae) captured in sandstone caves from Central Amazonia, Brazil. Mem Inst Oswaldo Cruz. 2011;106: 353–359. doi:10.1590/s0074-02762011000300016
48. Teles CBG, dos Santos AP de A, Freitas RA, de Oliveira AFJ, Ogawa GM, Rodrigues MS, et al. Phlebotomine sandfly (Diptera: Psychodidae) diversity and their Leishmania DNA in a hot spot of American Cutaneous Leishmaniasis human cases along the Brazilian border with Peru and Bolivia. Mem Inst Oswaldo Cruz. 2016;111: 423–432. doi:10.1590/0074-02760160054
49. de Souza AAA, da Rocha Barata I, das Graças Soares Silva M, Lima JAN, Jennings YLL, Ishikawa EAY, et al. Natural Leishmania (Viannia) infections of phlebotomines (Diptera: Psychodidae) indicate classical and alternative transmission cycles of American cutaneous leishmaniasis in the Guiana Shield, Brazil. Parasite. 24: 13. doi:10.1051/parasite/2017016
50. de Souza CF, Brazil RP, Bevilacqua PD, Andrade Filho JD. The phlebotomine sand flies fauna in Parque Estadual do Rio Doce, Minas Gerais, Brazil. Parasites & Vectors. 2015;8: 619. doi:10.1186/s13071-015-1227-1
51. de Aguiar GM, Vieira VR. Regional Distribution and Habitats of Brazilian Phlebotomine Species. In: Rangel EF, Shaw JJ, editors. Brazilian Sand Flies : Biology, Taxonomy, Medical Importance and

Control. Cham: Springer International Publishing; 2018. pp. 251–298. doi:10.1007/978-3-319-75544-1_4

52. Lozano-Sardaneta YN, Jiménez-Girón EI, Rodríguez-Rojas JJ, Sánchez-Montes S, Álvarez-Castillo L, Sánchez-Cordero V, et al. Species diversity and blood meal sources of phlebotomine sand flies (Diptera: Psychodidae) from Los Tuxtlas, Veracruz, Mexico. Acta Tropica. 2021;216: 105831. doi:10.1016/j.actatropica.2021.105831
53. Anaguano DF, Ponce P, Baldeón ME, Santander S, Cevallos V. Blood-meal identification in phlebotomine sand flies (Diptera: Psychodidae) from Valle Hermoso, a high prevalence zone for cutaneous leishmaniasis in Ecuador. Acta Tropica. 2015;152: 116–120. doi:10.1016/j.actatropica.2015.09.004
54. Ontivero IM, Beranek MD, Rosa JR, Ludueña-Almeida FF, Almirón WR. Seasonal distribution of Phlebotomine sandfly in a vulnerable area for tegumentary leishmaniasis transmission in Córdoba, Argentina. Acta Tropica. 2018;178: 81–85. doi:10.1016/j.actatropica.2017.10.028