

1 **Global Distribution Maps of the Leishmaniases**

2 **David M. Pigott¹, Samir Bhatt¹, Nick Golding¹, Kirsten A. Duda¹, Katherine E. Battle¹,**
3 **Oliver J. Brady¹, Jane P. Messina¹, Yves Balard², Patrick Bastien^{2,3}, Francine**
4 **Pratlong^{2,3}, John S. Brownstein^{4,5}, Clark Freifeld^{4,6}, Sumiko R. Mekaru⁴, Peter W.**
5 **Gething¹, Dylan B. George⁷, Monica F. Myers¹, Richard Reithinger⁸ and Simon I. Hay¹,**
6 ⁷

7 ¹ Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University
8 of Oxford, South Parks Road, Oxford, United Kingdom.

9 ² University Montpellier 1 (UFR Médecine) & CNRS 5290/IRD 224 (UMR “MiVEGEC”),
10 Laboratoire de Parasitologie – Mycologie, Montpellier, France.

11 ³ CHRU de Montpellier, Centre National de Référence des Leishmanioses, Departement de
12 Parasitologie – Mycologie, Montpellier, France.

13 ⁴ Children’s Hospital Informatics Program, Boston Children’s Hospital, Boston, MA, United States of
14 America.

15 ⁵ Department of Pediatrics, Harvard Medical School, Boston, MA, United States of America.

16 ⁶ Department of Biomedical Engineering, Boston University, Boston, MA, United States of America.

17 ⁷ Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, United States of
18 America.

19 ⁸ RTI International, Washington D.C., United States of America.

20

21 **Abstract**

22 The leishmaniasis are vector-borne diseases that have a broad global distribution throughout much of
23 the Americas, Africa and Asia. Despite representing a significant public health burden, our
24 understanding of the global distribution of the leishmaniasis remains vague, reliant upon expert
25 opinion and limited to poor spatial resolution. A global assessment of the consensus of evidence for
26 leishmaniasis was performed at a sub-national level by aggregating information from a variety of
27 sources. A database of records of cutaneous and visceral leishmaniasis occurrence was compiled from
28 published literature, online reports, strain archives and GenBank accessions. These, with a suite of
29 biologically relevant environmental covariates, were used in a boosted regression tree modelling
30 framework to generate global environmental risk maps for the leishmaniasis. These high-resolution
31 evidence-based maps can help direct future surveillance activities, identify areas to target for disease
32 control and inform future burden estimation efforts.

33 **Keywords:** cutaneous leishmaniasis, visceral leishmaniasis, niche based modelling, boosted
34 regression trees, species distribution modelling, disease mapping

35 **Introduction**

36 The leishmaniasis are a group of protozoan diseases transmitted to humans and other
37 mammals by phlebotomine sandflies (*Murray et al., 2005; WHO, 2010*). Considered as one of the
38 neglected tropical diseases (NTD) (*WHO, 2009*), the leishmaniasis can be caused by around 20
39 *Leishmania* species and include a complex life cycle involving multiple arthropod vectors and
40 mammalian reservoir species (*Ashford, 1996; Ready, 2013*). Sandflies belonging to either
41 *Phlebotomus* spp. (Old World) or *Lutzomyia* spp. (New World) are the primary vectors; domestic
42 dogs, rodents, sloths and opossums are amongst a long list of mammals that are either incriminated or
43 suspected reservoir hosts. Non-vector transmission (e.g. by accidental laboratory infection, blood
44 transfusion or organ transplantation) is possible, but rare (*Cardo, 2006*). Transmission of the
45 leishmaniasis can be either anthroponotic or zoonotic. The leishmaniasis rank as the leading NTD in
46 terms of mortality and morbidity with an estimated 50,000 deaths in 2010 (*Lozano et al., 2012*) and
47 3.3 million disability adjusted life years (*Murray et al., 2012*).

48 Symptoms of *Leishmania* infection can take many different and diverse forms (*Banuls et al.,*
49 *2011*), the two main outcomes being cutaneous leishmaniasis (CL) and visceral leishmaniasis (VL).
50 Cutaneous leishmaniasis typically presents as cutaneous nodules or lesions at the site of the sandfly
51 bite (localised cutaneous leishmaniasis). In some cases parasites disseminate through the skin and
52 present as multiple non-ulcerative nodules (diffuse cutaneous leishmaniasis, DCL) or propagate
53 through the lymphatic system resulting in nasobronchial and buccal mucosal tissue destruction
54 (mucosal leishmaniasis, ML) (*Dedet and Pratlong, 2009; Reithinger et al., 2007*). Localised CL may
55 resolve spontaneously and usually responds well to treatment; management of DCL and ML cases is
56 more difficult and cases may take considerably longer to resolve, if at all. Visceral leishmaniasis
57 generally affects the spleen, liver or other lymphoid tissues, and, if left untreated, is fatal; a fraction of
58 successfully treated VL cases may result in maculopapular or nodular rashes (post-kala-azar dermal
59 leishmaniasis) (*Dedet and Pratlong, 2009; Murray et al., 2005*). While the *Leishmania* species
60 determines which of the main two forms of the leishmaniasis will result from infection,
61 establishment, progression and severity of infection as well as treatment regimen and outcome is
62 dependent on a range of other factors, including parasite strain, characteristics of sandfly saliva,
63 parasite infection with *Leishmania* RNA virus, host genetics and immunosuppression, particularly due
64 to HIV co-infection (*Ives et al., 2011; Novais et al., 2013; Reithinger et al., 2007*).

65 Species distribution models provide a robust means of mapping these diseases at a global
66 level. These models define a set of conditions, from a selection of environmental covariates, which
67 best categorise known occurrences. Through this categorisation, areas of unknown pathogen presence
68 can be identified and thus a global evaluation of environmental suitability for presence can be made.
69 A variety of factors can influence the distribution of an organism, including an array of environmental

and other abiotic characteristics as well as biotic factors (*Peterson, 2008*). Whilst many areas may be environmentally suitable for a given species, other factors may prevent the species from being present in all of these locations. This distinction is often referred to as the difference between the fundamental and the realised niche of the species, the former describing a potential distribution based upon specific features of the environment whilst the latter indicates the distribution we observe. Such a framework can be applied just as successfully in the context of pathogens and their vectors as with macroorganisms (*Peterson et al., 2011*) and has already been applied to the mapping of malaria vectors (*Sinka et al., 2011; Sinka et al., 2010; Sinka et al., 2010*) and dengue (*Bhatt et al., 2013*). The relationships between the leishmaniasis and environmental and socioeconomic factors known to influence their distribution at a global scale has not previously been considered in a comprehensive and quantitative manner (*Hay et al., 2013*). This study uses these modelling techniques in order to define the first evidence-based global environmental risk maps of the leishmaniasis.

Results

Evidence of leishmaniasis

For each province or state across the globe (classed as Admin 1 by the Food and Agriculture Organization's Global Administrative Unit Layers (*FAO, 2008*), totalling some 3,450) evidence was collected regarding CL and VL presence or absence. An assessment of the consensus of this evidence ranging from comprehensive agreement on disease presence (+100%) to consensus of disease absence (-100%) was made. Figures 1a – 4a present these evidence consensus maps, with full reasoning for each administrative unit's score outlined in the associated dataset (Dryad dataset doi:10.5061/dryad.05f5h). For Brazil, it was possible to perform this analysis at the district level (classed as Admin 2) totalling some 5,510 units. In total 950 Admin 1 units from 84 countries reported a consensus on CL presence greater than indeterminate (a score of 0), with 310 Admin 1 units from 42 countries reporting a complete consensus on the presence of CL. In Brazil, 2,469 Admin 2 regions recorded CL cases over the period of investigation. Consensus on the presence of VL (score greater than 0) was reported in 793 Admin 1 units from 77 countries, with 88 Admin 1 units from 32 countries reporting complete consensus on VL. In Brazil, 1,320 Admin 2 units recorded VL cases.

Of the ten countries (Afghanistan, Colombia, Brazil, Algeria, Peru, Costa Rica, Iran, Syria, Ethiopia and Sudan) that contribute 75% of the global estimated CL incidence (*Alvar et al., 2012*), only Algeria did not have regions of complete evidence consensus on presence due to incomplete and non-contemporary case data. Similarly, of the six countries (Brazil, Ethiopia, Sudan, South Sudan, India and Bangladesh) that report 90% of all VL cases (*Alvar et al., 2012*), all six had regions of complete consensus on VL.

Figures 1a – 4a also show the spatial distribution of occurrence data, defined as one or more reports of leishmaniasis in a given calendar year, collated from a variety of sources. Overall, there is a relatively broad geographic spread and good correspondence with the evidence consensus maps for each disease. Tunisia, Morocco and Brazil report the highest number of unique CL occurrences in any given year, whilst India reported the largest proportion of the VL occurrence data.

Table 1 reports the sources and types of data within the occurrence database. Whilst the majority of occurrence records contain accurate point data (62%), the remainder were recorded at a provincial or district level. Occurrence records for the two diseases were relatively similar in number with a total of 6,426 records for CL and 6,137 for VL.

Modelled distribution of the leishmaniases

Figures 1b- 4b show the global predicted environmental risk maps for CL and VL. Table 2 identifies the top five predictor variables in each of the four modelled regions (since CL and VL were modelled separately in the Old World and New World) as measured by average contribution to the boosted regression trees (BRT) submodels. Peri-urban and urban land cover is an important predictor of the distribution of CL in the Old World and of VL globally. Abiotic factors such as land surface temperature (LST) were better predictors of CL than of VL. In total, LST variables (annual minimum, maximum and mean) explain 21.99% of CL distribution in the Old World and 43.65% of CL distribution in the New World (with maximum LST having the highest relative contribution). Abiotic factors combined (including LST, normalised difference vegetation index (NDVI) and precipitation) accounted for 29.02% and 48.55% of VL distribution in the Old World and New World, respectively. Validation statistics for all models were high with a mean area under the receiver operator curve (AUC) above 0.97 and mean correlations above 0.85 for all models.

In the New World, CL is predicted to occur primarily within the Amazon basin and other areas of rainforest. By contrast, VL is predicted to occur mainly along the coastline of Brazil, with sporadic foci across the rest of Southern and Central America. Outside of their main foci, both diseases are strongly associated with urban and peri-urban areas, resulting in a focal distribution throughout much of the New World.

In the Old World, both CL and VL are predicted to be present from the Mediterranean Basin across the Near East to Northwest India, with a few foci in Central China as well as in a thin band of predicted risk across West Africa and in the Horn of Africa. The predicted distribution of VL also extends into Northeast India and China with a large predicted focus in the northwest.

The populations living in areas predicted to be subject to environmental risk of CL and VL are estimated to be 1.71 billion and 1.69 billion, respectively, approximately a quarter of the world's

population. Figure 4-figure supplement 4 compares these national estimates to the annual case incidence data from all countries for which at least one case *per annum* was estimated by *Alvar et al. (2012)*. There is a strong positive association between the two measures of disease occurrence. We provide estimates of the populations at risk in 90 countries for which no human cases of CL or VL were regularly reported (*Alvar et al., 2012*). A full table of this information is presented in the associated Dryad dataset (doi:10.5061/dryad.05f5h). For many of these countries, *Alvar et al. (2012)* reported a handful of sporadic cases over the years indicating very rare occurrence of infection, whilst the remainder were countries with inconclusive evidence of disease presence or absence. It is important to note that the relationship between environmental risk and true incidence of disease remains to be elucidated; however the association between population living in areas of environmental risk and national level estimates of incidence suggests that the modelled occurrence-incidence relationship approach used for by *Bhatt et al. (2013)* for dengue could be applied if the necessary longitudinal cohort study data were available.

Discussion

This work has compiled a large body of qualitative and quantitative information on the global distribution of the leishmaniasis and employed a statistical modelling framework to generate the first published high-resolution global distribution maps of these diseases.

The evidence consensus maps provide a useful assessment of both global and regional knowledge of these diseases. Whilst in many countries consensus on presence or absence of the leishmaniasis exists, in other areas, including large parts of Africa and many states in India, these assessments reveal significant uncertainty in assessing disease presence or absence using currently available evidence. It is in these data-poor countries that increased surveillance efforts should be concentrated to improve our knowledge of the global distribution of the leishmaniasis. In some locations, cases have been reported as locally transmitted without the presence of proven vector species, which could indicate a false positive. However, the overall consensus score will reflect any uncertainty associated with the validity of these reports; if multiple independent sources report autochthonous cases, this increased certainty will be reflected in a higher consensus score. Similarly, whilst the occurrence database contains data from across the globe, this dataset is inevitably subject to spatial bias in reporting, with more data reported from more economically developed countries where we already have good knowledge of the disease (e.g. Spain, France and Italy).

The complexity and diversity of transmission cycles involving not just humans, but also a multitude of vectors and reservoirs, necessitated a modelling approach which can account for highly non-linear effects of covariates on probability of disease presence. The BRT modelling approach employed is able to do this and has previously been shown to produce highly accurate predictions across a wide range of species (*Elith et al., 2006; Elith et al., 2008*). This ecological niche modelling

approach is therefore able to deal with not only the variation in parasites causing infection, but also the various life-histories and habitat preferences associated with the different vector species.

A restriction of the BRT approach (in common with other species distribution modelling approaches) is the need for absence data in addition to occurrence data. Since reliable absence data was not available at this spatial scale, the incorporation of pseudo-data into the modelling framework was necessary. The methodology employed in this study attempted to minimise the problems this can cause, by using a probabilistic approach to generate the pseudo-data which incorporates the evidence consensus and distance from existing occurrence points. Similarly, reporting bias within the occurrence database is an issue with all presence-only species distribution models (*Peterson et al., 2011*). If bias is unaccounted for, there is the potential that the model merely reflects factors that correlate with the probability of reporting disease occurrence rather than the disease itself, such as healthcare expenditure (*Phillips et al., 2009; Syfert et al., 2013*). The pseudo-data selection procedures (which included information from both the occurrence dataset and the less-biased evidence consensus map) coupled with the model ensembling approach aimed to minimise this potential source of bias.

The differences in the most important predictors of disease presence between the two forms of the disease and between the Old and New Worlds highlight the complex and spatially variable epidemiology of the leishmaniasis. Similar to a recent study of the spatial predictors of dengue occurrence (*Bhatt et al., 2013*), environmental and socioeconomic factors were found to be important contributors to the distribution of both CL and VL. For VL, both Old World and New World distributions are driven by peri-urban (and to a lesser degree urban) land cover. This reflects recent trends observed, for instance, in Brazil and Bihar state in India, where areas of highest risk have been found in peridomestic settings (*Bern et al., 2010; Harhay et al., 2011*). This risk factor may well be linked back to aspects of vector bionomics, with many vectors in these regions associating with or near households in general (*Poche et al., 2011; Singh et al., 2008; Uranw et al., 2013*). Furthermore, whilst significant anthroponotic transmission of *L. donovani* occurs across parts of the Old World, zoonotic cycles of VL, primarily tied to canine hosts, dominate *L. infantum* transmission (*Chamaille et al., 2010; Ready, 2013*), with infection in dogs shown to be closely associated with human population density.

Important predictors of CL distribution differed markedly between the Old and New World. Whilst peri-urban land cover was the most important predictor of the disease in the Old World, in the New World temperature was the highest predictor, with abiotic factors predicting 74.18% of CL distribution. This difference in the relative importance of climatic drivers reflects the fact that in the Old World the main endemic CL areas are due to both anthroponotically-transmitted *L. tropica* as well as zoonotic cycles of *L. major*, whereas in the New World the disease is primarily associated

with sylvatic and zoonotic cycles with a variety of different *Leishmania* spp. and wild reservoir hosts implicated (Ashford, 1996; Lima et al., 2013; Ready, 2013; Reithinger et al., 2007; WHO, 2010).

The distribution maps represent a spatially refined assessment of the global environmental risk of leishmaniasis and provide a starting point for various public health activities including targeting areas for control and assessing disease burden. The maps compare favourably to the WHO Expert Committee on the Control of Leishmaniases outputs (WHO, 2010), have high model validation statistics and improve upon the existing body of work by providing a finer resolution of risk at a subnational level. Similarly, the countries indicated by Alvar et al. (2012) as having 90% of all VL and 75% of all CL cases, were all predicted by our maps to have risk for VL and CL, respectively.

There are a number of regions in which our maps do not correspond as closely to these previous findings. Regions such as Northwest China are predicted to have high risk for VL, though the low population densities in this area are likely to lead to very few cases and, given its remoteness, even fewer reported cases. Other regions, such as the Mediterranean coastline of Europe, are predicted to be highly suitable for leishmaniasis, but we see few human cases. This is because the maps presented predict the probability of disease presence in an area, rather than directly infer measures of incidence or burden, which can be influenced by a variety of other factors (e.g. in the Mediterranean coastline of Europe, VL has been associated with immunosuppression). The evidence consensus layer, used to mask out regions with high consensus on leishmaniasis absence, acts as a rough filter on the environmental risk maps. However, in order to model the true relationship between environmental risk and disease incidence, a global dataset of geopositioned disease incidence data would be required; at present this is unavailable.

Estimates of the populations living in areas of environmental risk are therefore supplied as a proxy for the true burden of disease. However, they cannot be directly compared with other global estimates of the leishmaniases' disease burden, such as the WHO estimates of clinical burden of around 350 million (WHO, 2010). Figure 4-figure supplement 4 shows a strong, positive relationship between population at risk estimates and estimated annual incidence from Alvar et al. (2012). The exceptions to this relationship (e.g. Egypt, Nigeria and Côte d'Ivoire) are all countries with indeterminate evidence consensus scores, indicating a genuine lack of knowledge regarding both the distribution and incidence of disease.

Previous estimates of the leishmaniases' global burden have been complicated by poor knowledge of the global distribution of the diseases (Bern et al., 2008; Reithinger, 2008). It is hoped that the maps presented here will help to increase the accuracy of future estimates. Ideally, future improvements to the global distribution maps presented here would distinguish between the different *Leishmania* species and sandfly vectors. Species-specific models at the same level of detail as those presented here are not currently possible due to a lack of suitable data. Developments in the use of

“big data” approaches to disease mapping (such as the incorporation of informal internet resources) may enable the construction of datasets which could be used in these analyses (*Hay et al., 2013*). A further complication with burden estimation is the epidemic nature of the disease, as evidenced by the national case time series in *Alvar et al. (2012)*, leading to significant interannual variation in burden. Therefore, any burden estimation would have to account for this and the temporal spread of data would therefore be critical.

It should be noted that non-environmental drivers of transmission and morbidity, such as HIV immunosuppression and risk of infection via blood transfusions and intravenous drug usage, are not incorporated into our present models. The maps presented here can help inform the wider discussion of these factors and their impact on leishmaniasis (e.g. by identifying regions with greater risk for HIV and leishmaniasis co-infection) (*Desjeux and Alvar, 2003*). Similarly, the niche based models used here could enable a decoupling of environmental from social factors to assess the importance of the latter on leishmaniasis transmission in particular areas. It may indeed be the case that in some specific localities it is these non-environmental risk factors that are the main determinants of disease distribution.

Conclusions

These maps represent evidence-based estimates of the current global distribution of the leishmaniases incorporating a comprehensive occurrence database and a rigorous statistical modelling framework with associated uncertainty statistics. We estimate that 1.71 billion and 1.69 billion individuals live in areas that are suitable for CL and VL transmission, respectively. These figures highlight the need for much greater awareness of this disease at a global scale. These maps provide an important baseline assessment and a strong foundation on which to base future burden estimates, target regions for control efforts and inform public health decisions.

Materials and Methods

A boosted regression tree (BRT) modelling framework was used to generate global predicted environmental risk maps for CL and VL. This framework required four key information components: (i) a map of the consensus of evidence for the global extents of the leishmaniasis; (ii) a comprehensive dataset of geopositioned CL and VL occurrence records; (iii) a suite of global, gridded datasets on environmental correlates of the leishmaniasis; and (iv) pseudo-data to augment the occurrence records. In order to better capture the realised niche of these diseases, prediction by the model is restricted to those areas of known disease transmission, or where transmission is uncertain, as defined by the evidence consensus layer (i). The full procedures used to generate these components and the resulting risk and prevalence maps are outlined below.

Evidence consensus

The methodology used for generating the definitive extents for the leishmaniasis was adapted from work on dengue (*Brady et al., 2012*). Four primary evidence categories were used to determine a consensus on the presence or absence of the leishmaniasis: (i) health reporting organisations; (ii) peer-reviewed evidence of local autochthonous transmission; (iii) case data; (iv) supplementary information. Cutaneous and visceral leishmaniasis were the two symptomatology investigated: other forms of the disease were subset within these two – whilst VL contained cases of post-kala-azar dermal leishmaniasis, CL included diffuse, disseminated and mucosal forms of the disease. Although limited amounts of data were available for some of these forms, their epidemiology is similar, and consequently this categorisation was seen as appropriate. Information was collected at provincial level (termed Admin 1 units by the Food and Agriculture Organization's (FAO) Global Administrative Unit Layers (GAUL) coding (*FAO, 2008*)) to better capture the focal nature of these diseases.

Health Reporting Organisation Evidence (scores between +3 and -3): Two health reporting organisations were referenced, the Global Infectious Diseases and Epidemiology Online Network (GIDEON) (*Edberg, 2005*) and the World Health Organization (WHO) (*WHO, 2010*). The status of disease was recorded for each Admin 1 unit as either present, absent or unspecified. If both reported the disease as present, +3 was scored, if both reported absence, -3 was scored, with +2/-2 scored if one reporting body did not specify the presence or absence of the disease. If the two disagreed, or both were non-specific, 0 was scored reflecting the lack of a consensus on the status of that region.

Peer-reviewed evidence (scores between +2 and +6): A review of reported leishmaniasis' cases was performed. Using PubMed and Web of Knowledge with "[admin1 province] leish*" as the search parameters, articles from January 1960 until September 2012 were abstracted. Each abstract was imported into Endnote X4 and assessed for relevance. Papers that included reported cases on either CL or VL were then obtained. Cases were included if there was sufficient evidence to suggest

that local autochthonous transmission had occurred. Where individuals from a non-endemic country had travelled to an endemic country (e.g. tourists and military personnel) and returned with an infection, this was included (as evidence for leishmaniasis in the foreign destination) since these typically represent immunologically naïve individuals who have undergone more rigorous diagnostics in their home country, and thus represent a potentially more informed data source. Each paper was assessed for contemporariness and diagnostic accuracy. Contemporariness was graded in 3 bands: 2005-2012=3, 1997-2004=2 and 1997 and earlier=1, as was diagnostic accuracy where 1 was scored for data that reported “confirmed” cases without detailing methodologies implemented; 2 was scored where evidence of microscopy, serology or the Montenegro skin test had been used; 3 was awarded to those studies that had used PCR or other molecular techniques (*Reithinger and Dujardin, 2007*). Contemporariness bins were based upon the potentially lengthy intrinsic incubation periods present with some *Leishmania* spp. as well as to accommodate the potential for epidemic cycles, where cases may only be detected in peak years and missed in the intervening baseline periods. The most contemporary and diagnostically accurate papers were then subset to maximise the consensus score for any given area.

Case Data (scores between -6 and +6): Case data was derived from reports on the leishmaniases provided by national health officials (*Alvar et al., 2012*). A threshold value of 12 CL cases and 7 VL cases in a given province in a given year was deemed suitable by the authors to distinguish significant disease events from sporadic cases within that region. If cases were reported at or above the threshold and were dated no later than 2005, +6 was scored. If data existed below this threshold, indicating sporadic cases, or data indicated a history of reported cases in the region but with no evidence of time period, scores were assigned stratified by total annual healthcare expenditure (HE) per capita at average US\$ exchange rates (*WHO, 2011*). This was used as a proxy to determine genuine sporadic reporting from inadequate surveillance. Three categories were defined – HE Low (< \$100), HE Medium (\$100 ≤ HE < \$500) and HE High (≥\$500). If sporadic cases were reported in an HE Low country, +4 was scored, whilst in an HE Medium country +2 was scored and in an HE High country, 0 was scored. If there were no reported case data available, HE Low countries scored +2, HE Medium countries scored -2 and HE High countries scored -6 (*Brady et al., 2012*).

Supplementary evidence: Supplementary evidence was provided in cases where a consensus on presence or absence could not be reached using the aforementioned evidence types, typically with areas where the consensus value was close to 0%. For these regions, additional literature searches were undertaken to determine whether known vector species or infected reservoir hosts were reported in the region. The justification for each provincial scenario is outlined in the associated online databases (Dryad dataset doi:10.5061/dryad.05f5h). In total this assessment was required in 24 countries.

An overall consensus score for each administrative region was calculated by the sum of the scores in each category, divided by the maximum possible score, then expressed as a percentage. Consensus was defined as either complete ($\pm 75\%$ to $\pm 100\%$), good ($\pm 50\%$ to $\pm 74\%$), moderate ($\pm 25\%$ to $\pm 49\%$), poor ($\pm 1\%$ to $\pm 24\%$) or indeterminate (0%). Such a classification is intended more as a guide to the quality of evidence for the leishmaniasis in an area, rather than as a strict classification of certainty. The full scores for each country are laid out in the associated online datasets (Dryad dataset doi:10.5061/dryad.05f5h).

Brazil and Peru: The Brazilian Ministry of Health produces, via the Sistema de Informação de Agravos de Notificação (SINAN) reporting network, records of infections at the municipality level. This allowed for a more thorough evidence consensus to be performed at district level (termed Admin 2 (FAO, 2008)) within Brazil. As above, WHO and GIDEON status as well as peer-reviewed literature score were recorded, both aggregated to Admin 1 provincial level. Case data was then defined by the presence of a municipality reporting leishmaniasis between 2008 and 2011 inclusive, with positive reports scoring +6 and absence scoring -6. The overall consensus score was then calculated as above. In addition, provincial level case data for Peru was replaced by Ministry of Health information as it was more contemporary than that listed by *Alvar et al. (2012)*.

Occurrence records

Two separate searches using PubMed and Web of Knowledge were undertaken using the search parameter “leish*,” and including articles up to December 2012, and their respective abstracts were filtered for relevance. From these searches, 4,845 articles were collated, with data recorded at the resolution of either a point or Admin 1 or 2 polygon. These were then geo-positioned using Google Maps (<https://maps.google.co.uk/>). Each entry was evaluated to ensure that non-autochthonous cases and duplicate entries were eliminated. Each occurrence was assigned a start and end date based upon the content of the paper, used to define the time period over which occurrences were reported.

In addition to this resource, reports were taken from the HealthMap database (<http://healthmap.org/en/>). HealthMap is an online based infectious disease surveillance system that compiles data from informal data sources ranging from online news articles to ProMED reports (Freifeld et al., 2008). It parses information from these sources searching for relevant keywords, and then, using crowdsourcing and automated processes, geolocations those relating to the disease of interest. As of December 2012 a total of 690 leishmaniasis relevant articles were archived.

Searches were also performed on GenBank accessions, searching for archived genetic information from *Leishmania* spp. known to infect humans (WHO, 2010). If the host was identified as human, geographic indicators were assigned either as point, Admin 1 or Admin 2, based upon the

information in the location tag. Tags at the national level were filtered out of the dataset. In total, 563 accessions were associated with sub-national location details and added to the database.

Finally, data were provided from the curated strain archives of the Centre National de Référence des Leishmanioses (CNR-L) in Montpellier, France. In total information about 3,465 strains isolated from humans was provided, collected from between 1954 and 2013.

All data were geopositioned as precisely as possible, which resulted in both point-level data (referring to cities, towns or villages) as well as polygon-level data (provinces or districts) with area no greater than one square decimal degree. All data that had been manually geopositioned were checked to ensure coordinates were plausible and then occurrences were standardised annually to remove intra-annual duplicates, so that each individual record used in our model represented an occurrence of leishmaniasis infections in a given 5km x 5km location or administrative unit for one given year. As a result, the occurrence data were independent of burden; a location with 200 cases in one year has equal weighting in the model as a location with just one reported case, since it was only the presence of the disease being modelled.

Environmental correlates

Leishmania spp. are known to have anthroponotic, zoonotic or sylvatic transmission cycles in nature (Ready, 2013; WHO, 2010) which is apparent in the focal nature of the disease; however, there are some key features of the environment that are important in determining the distribution of disease across the globe. Numerous models have been constructed for local transmission scenarios implicating various environmental features from temperature and precipitation to socioeconomic factors relating to standards of living in villages in endemic foci. For the modelling process, a suite of global gridded environmental, biologically plausible, correlates was generated.

Precipitation: Humidity and moisture, whether from rainfall or in the soil, have often been identified as important for the sandfly, with humidity influencing breeding and resting (Ready, 2013). Whilst relatively little is known about these breeding sites, of the few that have been identified, high humidity seems to be a common trait, including moist Amazonian soils, caves, animal burrows and select human dwellings (Felicangeli, 2004; Killick-Kendrick, 1999). Studies have indicated soil type and their moisture profiles as determinants of sandfly distribution (Bhunja et al., 2010; Elnaïem, 2011). Precipitation represents a good global proxy measure for moisture, and has been shown to play a prominent role in shaping disease distribution in previous leishmaniasis modelling efforts (Bhunja et al., 2010; Chamaille et al., 2010; Elnaïem et al., 2003; Elnaïem, 2011; Gonzalez et al., 2011; Gonzalez et al., 2010; Hartemink et al., 2011; Malaviya et al., 2011; Thomson et al., 1999).

399 Estimates of precipitation were obtained from the WorldClim database (www.worldclim.org).
400 This resource, which is freely available online, provides data spanning from 1950 to 2000, describing
401 monthly averages over this time, at a 1km x 1km resolution (*Hijmans et al., 2005*). Using this
402 baseline, interpolated global climate surfaces were produced using ANUSPLIN-SPLINA software
403 (*Hutchinson, 1995*). With the use of temporal Fourier analysis, seasonal and inter-annual variation in
404 precipitation patterns, taken from the interpolated global surface, were used to calculate minimum and
405 maximum monthly precipitation averages (*Rogers et al., 1996; Scharlemann et al., 2008*).

406 *Temperature:* Temperature influences both the development of the infecting *Leishmania*
407 parasite in the sandfly (*Hlavacova et al., 2013*) as well as the life cycle of the sandfly vectors. On one
408 hand, studies have shown that with increasing temperatures, the metabolism of the sandfly increases,
409 influencing oviposition, defecation, hatching and adult emergence rates (*Benkova and Volf, 2007;*
410 *Guzman, 2000; Kasap and Alten, 2005*). On the other hand, higher temperatures have also been
411 shown to increase mortality rates of adults (*Benkova and Volf, 2007; Guzman, 2000*). Studies have
412 integrated the effects of temperature on sandfly biting rates, sandfly mortality and extrinsic incubation
413 periods to produce maps of how the basic reproductive number of canine leishmaniasis varied
414 spatially (*Hartemink et al., 2011*). Multiple studies have also implicated temperature (including
415 maximum, minimum and mean temperatures) as being an important explanatory variable for both
416 sandfly and disease distribution (*Bhunja et al., 2010; Branco et al., 2013; Chamaille et al., 2010;*
417 *Fernandez et al., 2012; Fischer et al., 2010; Galvez et al., 2011; Gebre-Michael et al., 2004;*
418 *Thomson et al., 1999*).

419 Using a similar methodology to generating precipitation surfaces, minimum, maximum and
420 mean monthly temperature values were generated (*Hijmans et al., 2005*).

421 *Normalised difference vegetation index (NDVI) and land cover:* Vegetation provides many
422 roles in sandfly habitat and survival, ranging from maintaining the necessary moisture profile for both
423 immature stages and adults, to a sugar resource for both male and female sandflies (*Feliciangeli,*
424 *2004; Killick-Kendrick, 1999; Ready, 2013*). Moreover, vegetation is an important resource for many
425 mammals that sandflies feed on, and that potentially are *Leishmania* reservoirs. The importance of
426 considering NDVI was demonstrated with respect to the distribution of the reservoir *Psammomys*
427 *obesus* (sand rat) and the distribution of its primary food, chenopods (*Toumi et al., 2012*). NDVI has
428 been implicated as a key explanatory variable in the distribution of leishmaniasis cases in several
429 studies (*Bhunja et al., 2012; Cross et al., 1996; de Oliveira et al., 2012; Elnaiem et al., 2003;*
430 *Elnaiem, 2011; Gebre-Michael et al., 2004; Hartemink et al., 2011; Thomson et al., 1999; Toumi et*
431 *al., 2012*).

432 The Advanced Very High Resolution Radiometer (AVHRR) NDVI product uses the spectral
433 reflectance of AVHRR channels 1 and 2 (visible red and near infrared wavelength) to quantitatively

assess the level of photosynthesising vegetation in a region (*Hay et al., 2006*). Using this data, compiled over multiple time intervals, patterns of NDVI were extracted for each gridded 1km x 1km cell.

Poverty: Neglected tropical diseases and poverty are often found to be linked and the use of a purely economic variable was chosen to act as a proxy for a variety of important global risk factors for disease, including malnutrition, housing quality and living with domesticated animals (*Bern et al., 2010; Boelaert et al., 2009; Herrero et al., 2009; Malafaia, 2009; Zeilhofer et al., 2008*).

The G-Econ database (www.gecon.yale.edu) takes economic data, at the smallest administrative division available, and spatially rescales these data to create a 1° x 1° gridded surface of the globe (*Nordhaus, 2008; Nordhaus, 2006*). This rescaling estimates the gross cell product of each grid cell, conceptually similar to gross domestic product, referring to the total market value of all final goods and services produced within one year, and can be considered as an indicator of overall standard of living within that area. Some cells provided multiple data; in these scenarios the best-quality information, as outlined by the quality field associated with the data, was used to select one value. All gross cell product values were then adjusted using purchasing power parity in US\$ for the years 1990, 1995, 2000 and 2005, using national aggregates estimated by the World Bank (*Nordhaus, 2006*) and computed the mean across all years for each gridded cell globally. This adjusted measure was used as the indicator of poverty in the model.

Urbanisation: Over the last few decades, there has been a tendency for the leishmaniasis having a sylvatic/zoontic transmission cycle to transition into the urban and peri-urban environment in response to increasing urbanisation trends (*Harhay et al., 2011*). The increasing overlap in habitat between suitable human and animal hosts and multiple available resting sites for adults can allow for transmission of disease to occur relatively easily (*Poche et al., 2011; Singh et al., 2008; Uranw et al., 2013*).

The Gridded Population of the World version 3 (GPW3) population density database projected for 2010 was used. The core Global Rural-Urban Mapping Project Urban Extents surface used night-time light satellite imagery to differentiate urban areas (*Balk et al., 2006*); GPW3 is a revision which updates the criteria for urban areas to those areas where population density is greater than or equal to 1000 people per km². Using the most up-to-date national censuses available and other demographic data resolved to the smallest available administrative unit, a gridded surface of 5km x 5km cells was generated. Each pixel could then be classified as urban, peri-urban or rural.

Modelling with Boosted Regression Trees

The Boosted Regression Trees (BRT) methodology employed for mapping the leishmaniasis is a variant of the model used in a previous analysis of dengue (Bhatt et al., 2013). Boosted regression tree modelling combines both regression trees, which build a set of decision rules on the predictor variables by portioning the data into successively smaller groups with binary splits (De'ath, 2007; Elith et al., 2008), and boosting, which selects the tree that minimises the loss function, to best capture the variables that define the distribution of the input data. The core BRT setup followed standard protocol already defined elsewhere (Bhatt et al., 2013; Elith et al., 2008).

Pseudo-data generation: As BRT requires both presence and absence data, the latter which is often hard to collate in an unbiased manner, pseudo-data had to be generated (Elith et al., 2008). There is no general consensus on how best to generate pseudo-data (Bhatt et al., 2013); however, several factors of the generation process are known to influence the predicted distribution and thus can be sources of potential bias (Barbet-Massin et al., 2012; Phillips et al., 2009; Phillips and Elith, 2011; Van Der Wal et al., 2009). In order to minimise such effects, pseudo-absence selection was directly related to the evidence consensus layer and restricted to a maximum distance (μ) from any occurrence point. Pseudo-presence data was also incorporated, again informed by the evidence consensus layer, to compensate for poor surveillance capacity in low prevalence regions. As in Bhatt et al. (2013) points were randomly located in regions above an evidence consensus threshold of -25, with regional placement probability weighted by evidence consensus scores, so that regions with higher evidence consensus contained more pseudo-presences than lower scoring areas. Since the occurrence dataset is from a wide range of sources and institutions, this procedure aims to mitigate sampling bias. By referencing the evidence consensus layer for pseudo-data selection, detection bias was also mitigated.

"Ensemble" analysis: There is no definitive procedure for choosing the best number of pseudo-data points to generate the most accurate predictive map. To account for the impact that these parameters might have on the model predictions, an ensemble BRT model was constructed with multiple BRT submodels fitted using pseudo-data points generated using different combinations of parameters n_a , n_p and μ . The numbers of pseudo-absences (n_a) and pseudo-presences (n_p) were defined as a proportion of the total number of actual data occurrence records (6,426 and 6,137 for CL and VL). The proportions used for generating pseudo-absences were 2:1, 4:1, 6:1, 8:1 and 10:1 and pseudo-presences were 0.025:1, 0.05:1 and 0.1:1. The pseudo-data were also generated within a restricted maximum distance (μ) from any actual presence point, and μ was varied through 5 distances: 5, 10, 15, 20 and 25 arc degrees. All combinations of these parameter values resulted in a total of 75 ($5n_a \times 3n_p \times 5\mu$) individual input data sets and BRT submodels (making up the BRT ensemble).

For each disease, the 75 BRT submodels were used to predict a range of different risk maps (each at 5km x 5km resolution) and these were combined to produce a single mean ensemble risk map for each disease, also allowing for computation of the associated range of uncertainty in these predictions for every 5km x 5km pixel as shown in Figures 1-4-figure supplements 1. For both diseases, the New World (the Americas) and Old World (Eurasia and Africa) were modelled separately in order to account for and explore any differences in the epidemiology of the diseases between these regions. This was done to differentiate the potential effect that the different vectors namely *Lutzomyia* spp. in the New World and *Phlebotomus* spp. in the Old World and their varying life histories, might have on the distribution of the diseases within these regions.

Summarising the BRT model: The relative importance of predictor variables was quantified for the final BRT ensemble. Relative importance is defined as the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split and averaged over all trees (Friedman, 2001). These contributions are scaled to sum to 100, with a higher number indicating a greater effect on the response. To evaluate the ensemble's predictive performance, we used the area under the receiver operator curve (AUC) (Fleiss et al., 2003) – the area under a plot of the true positive rate versus false positive rate, reflecting the ability to discriminate between presence and absence. An AUC value of 0.5 indicates no discriminative ability and a value of 1 indicates perfect discrimination.

It is important to note that this distribution modelling technique assesses pixel level risk, rather than population level risk. As such, the ensemble evaluates the likelihood of leishmaniasis presence based upon the covariates supplied. In reality, some other factors, such as national healthcare provisioning and standards of living will influence the true observed burden. Therefore, whilst these two levels of risk are inherently related, additional information, namely incidence data from many different populations, is required in order to assess the link quantitatively (Bhatt et al., 2013).

Estimation of population living in areas of environmental risk

Population living in areas of risk was estimated by using a threshold probability to reclassify the probabilistic risk maps into a binary risk map, then extracting the total human population in the “at risk” areas using a gridded dataset of human population density from 2010 (Balk et al., 2006; CIESIN/IFPRI/WB/CIAT, 2007). The threshold value was set such that 95% of the point occurrence records fell within the at risk area. Five percent of occurrence points were allowed to fall outside the predicted risk area to account for errors which could have arisen either from errors in the occurrence dataset or from inaccuracies in the predicted risk maps.

For external validation, this population at risk information was compared to national reported annual cases (Alvar et al., 2012) to produce Figure 4-figure supplement 4. In these figures the points

534 represent the mean value of the estimated annual incidence reported taking into account the authors
535 estimates of underreporting rates (*Alvar et al., 2012*). The upper and lower limits to these estimates
536 are reflected by the bars around each point. Note that these figures use a log-scale on each axis and
537 that only countries with non-zero estimates by *Alvar et al. (2012)* are included.

538 The threshold probabilities of occurrence used to define “at risk” were as follows: NW CL –
539 0.22, OW CL – 0.19, NW VL – 0.42, OW VL – 0.19.

Competing Interest

The authors declare no competing interests.

Acknowledgments and Funding

DMP is funded by a Sir Richard Southwood Graduate Scholarship from the Department of Zoology at the University of Oxford. SIH is funded by a Senior Research Fellowship from the Wellcome Trust (095066) which also supports KAD and KEB. NG is funded by a grant from the Bill & Melinda Gates Foundation (#OPP1053338). PWG is a Medical Research Council (UK) Career Development Fellow (#K00669X) and receives support from the Bill and Melinda Gates Foundation (#OPP1068048) which also supports SB. OJB is funded by a BBSRC studentship. JPM is funded by, and SIH and OJB acknowledge the support of, the International Research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS, European Commission 7th Framework Programme (#21803) <http://www.idams.eu>). JSB, CF and SRM acknowledge funding from NIH National Library of Medicine (R01LM010812). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conception of experiments: DBG SIH. Evidence consensus generation: DMP OJB. Data contributions: YB PB FP JSB CF SRM. Literature occurrence database collection and curation: DMP KAD KEB JPM MFM SIH. Modelling methodology: SB NG PWG SIH. Analysis of data: DMP SB NG RR SIH. All authors contributed to the writing of the paper.

Abbreviations

AUC – Area under the receiver operator curve
BRT – Boosted regression tree
CL – Cutaneous leishmaniasis
DCL – Diffuse cutaneous leishmaniasis
GIDEON – Global Infectious Disease and Epidemiology Online Network
LST – Land surface temperature
ML – Mucosal leishmaniasis
NDVI – Normalised difference vegetation index
NTD – Neglected tropical disease
VL – Visceral leishmaniasis

Figure 1: **Reported and predicted distribution of cutaneous leishmaniasis in the New World.** (A) evidence consensus for presence of the disease ranging from green (complete consensus on the absence: -100%) to purple (complete consensus on the presence of disease: +100%). The blue spots indicate occurrence points or centroids of occurrences within small polygons. (B) predicted risk of cutaneous leishmaniasis from green (low probability of presence) to purple (high probability of presence).

Figure 2: **Reported and predicted distribution of visceral leishmaniasis in the New World.** (A) evidence consensus for presence of the disease ranging from green (complete consensus on the absence: -100%) to purple (complete consensus on the presence of disease: +100%). The blue spots indicate occurrence points or centroids of occurrences within small polygons. (B) predicted risk of visceral leishmaniasis from green (low probability of presence) to purple (high probability of presence).

Figure 3: **Reported and predicted distribution of cutaneous leishmaniasis in the Old World.** (A) evidence consensus for presence of the disease ranging from green (complete consensus on the absence: -100%) to purple (complete consensus on the presence of disease: +100%). The blue spots indicate occurrence points or centroids of occurrences within small polygons. (B) predicted risk of cutaneous leishmaniasis from green (low probability of presence) to purple (high probability of presence).

Figure 4: **Reported and predicted distribution of visceral leishmaniasis in the Old World.** (A) evidence consensus for presence of the disease ranging from green (complete consensus on the absence: -100%) to purple (complete consensus on the presence of disease: +100%). The blue spots indicate occurrence points or centroids of occurrences within small polygons. (B) predicted risk of visceral leishmaniasis from green (low probability of presence) to purple (high probability of presence).

Figure 1- figure supplement 1: **Uncertainty associated with predictions in figure 1(B).** Uncertainty was calculated as the range of the 95% confidence interval in predicted probability of occurrence for each pixel. Regions of highest uncertainty are in dark brown, with blue representing low uncertainty.

Figure 2- figure supplement 1: **Uncertainty associated with predictions in figure 2(B).** Uncertainty was calculated as the range of the 95% confidence interval in predicted probability of occurrence for each pixel. Regions of highest uncertainty are in dark brown, with blue representing low uncertainty.

Figure 3- figure supplement 1: **Uncertainty associated with predictions in figure 3(B).** Uncertainty was calculated as the range of the 95% confidence interval in predicted probability of occurrence for each pixel. Regions of highest uncertainty are in dark brown, with blue representing low uncertainty.

Figure 3- figure supplement 2: **Reported and predicted distribution of cutaneous leishmaniasis in northeast Africa.** (A) evidence consensus for presence of the disease ranging from green (complete consensus on the absence: -100%) to purple (complete consensus on the presence of disease: +100%). The blue spots indicate occurrence points or centroids of occurrences within small polygons. (B) predicted risk of cutaneous leishmaniasis from green (low probability of presence) to purple (high probability of presence).

Figure 3- figure supplement 3: **Reported and predicted distribution of cutaneous leishmaniasis across the Near East, including Syria, Iran and Afghanistan.** (A) evidence consensus for presence of the disease ranging from green (complete consensus on the absence: -100%) to purple (complete consensus on the presence of disease: +100%). The blue spots indicate occurrence points or centroids of occurrences within small polygons. (B) predicted risk of cutaneous leishmaniasis from green (low probability of presence) to purple (high probability of presence).

Figure 4- figure supplement 1: **Uncertainty associated with predictions in figure 4(B).** Uncertainty was calculated as the range of the 95% confidence interval in predicted probability of occurrence for each pixel. Regions of highest uncertainty are in dark brown, with blue representing low uncertainty.

Figure 4- figure supplement 2: **Reported and predicted distribution of visceral leishmaniasis in northeast Africa.** (A) evidence consensus for presence of the disease ranging from green (complete consensus on the absence: -100%) to purple (complete consensus on the presence of disease: +100%). The blue spots indicate occurrence points or centroids of occurrences within small polygons. (B)

631 predicted risk of visceral leishmaniasis from green (low probability of presence) to purple (high
632 probability of presence).

633
634 Figure 4- figure supplement 3: **Reported and predicted distribution of visceral leishmaniasis in**
635 **the Indian subcontinent.** (A) evidence consensus for presence of the disease ranging from green
636 (complete consensus on the absence: -100%) to purple (complete consensus on the presence of
637 disease: +100%). The blue spots indicate occurrence points or centroids of occurrences within small
638 polygons. (B) predicted risk of visceral leishmaniasis from green (low probability of presence) to
639 purple (high probability of presence).

640
641 Figure 4- figure supplement 4: **Population at risk estimates for leishmaniasis.** Four scatterplots
642 showing the relationship between non-zero estimated mean annual incidence (*Alvar et al., 2012*) and
643 estimated population at risk derived from the cartographic approach for (A) New World cutaneous
644 leishmaniasis, (B) New World visceral leishmaniasis, (C) Old World cutaneous leishmaniasis and (D)
645 Old World visceral leishmaniasis. For each country the bars represent the annual incidence estimate
646 range.

649 **Table 1:** Origin and spatial resolution of leishmaniasis occurrence data.

Origin and resolution of occurrence data

Cutaneous leishmaniasis	<i>Point data</i>	<i>Province level data</i>	<i>District level data</i>	<i>TOTAL</i>
<i>Literature</i>	3,680	879	1,220	5,779
<i>CNR-L</i>	531	47	31	609
<i>HealthMap</i>	31	-	-	31
<i>GenBank</i>	6	-	1	7
<i>TOTAL</i>	4,248	926	1,252	6,426
Visceral leishmaniasis				
<i>Literature</i>	3,050	1,500	1,068	5,618
<i>CNR-L</i>	429	24	29	482
<i>HealthMap</i>	32	1	-	33
<i>GenBank</i>	3	-	1	4
<i>TOTAL</i>	3,514	1,525	1,098	6,137

650 Each cell gives the number of occurrence records added to the dataset by considering each additional
651 datasource after removing duplicate records. Occurrence records are separated by spatial resolution –
652 whether they are recorded as points (typically representing settlements) or as province level (admin 1)
653 or district level (admin 2) data.

654 **Table 2:** Mean relative contribution of predictor variables to the ensemble BRT models of CL and VL
655 in both the Old and New World.

Predictor covariates of cutaneous and visceral leishmaniasis

Old World

<i>Top predictors of CL</i>	<i>Relative contribution</i>	<i>Top predictors of VL</i>	<i>Relative contribution</i>
Peri-urban extents	47.34	Peri-urban extents	51.50
Minimum LST	18.36	Urban extents	17.38
Urban extents	9.01	Maximum NDVI	7.87
G-Econ	7.33	Minimum LST	5.87
Minimum Precipitation	4.95	Maximum Precipitation	4.00

New World

<i>Top predictors of CL</i>	<i>Relative contribution</i>	<i>Top predictors of VL</i>	<i>Relative contribution</i>
Maximum LST	36.91	Peri-urban extents	25.90
Peri-urban extents	18.61	Urban extents	21.24
Maximum precipitation	12.06	Mean LST	9.18
Minimum precipitation	6.21	Mean NDVI	7.83
Minimum LST	4.39	Maximum LST	6.40

656 LST = Land Surface Temperature, G-Econ = Geographically based Economic data, NDVI =
657 Normalised Difference Vegetation Index.

658 **References**

- 659 Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, et al. 2012. Leishmaniasis worldwide and global
660 estimates of its incidence. *PLoS One* **7**: e35671. doi:10.1371/journal.pone.0035671.
- 661 Ashford RW. 1996. Leishmaniasis reservoirs and their significance in control. *Clin Dermatol* **14**: 523-
662 532. doi:10.1016/0738-081x(96)00041-7.
- 663 Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, et al. 2006. Determining global population
664 distribution: methods, applications and data. *Adv Parasitol* **62**: 119-156. doi:10.1016/S0065-
665 308X(05)62004-0.
- 666 Banuls AL, Bastien P, Pomares C, Arevalo J, Fisa R, et al. 2011. Clinical pleiomorphism in human
667 leishmaniasis, with special mention of asymptomatic infection. *Clin Microbiol Infect* **17**:
668 1451-1461. doi:10.1111/j.1469-0691.2011.03640.x.
- 669 Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. 2012. Selecting pseudo-absences for species
670 distribution models: how, where and how many? *Methods Ecol Evol* **3**: 327-338.
671 doi:10.1111/j.2041-210X.2011.00172.x.
- 672 Benkova I, Volf P. 2007. Effect of temperature on metabolism of *Phlebotomus papatasi* (Diptera :
673 Psychodidae). *J Med Entomol* **44**: 150-154. doi:10.1603/0022-
674 2585(2007)44[150:Eotomo]2.0.Co;2.
- 675 Bern C, Courtenay O, Alvar J. 2010. Of cattle, sand flies and men: a systematic review of risk factor
676 analyses for South Asian visceral leishmaniasis and implications for elimination. *PLoS Negl Trop Dis* **4**: e599. doi:10.1371/journal.pntd.0000599.
- 677 Bern C, Maguire JH, Alvar J. 2008. Complexities of assessing the disease burden attributable to
678 leishmaniasis. *PLoS Negl Trop Dis* **2**: e313. doi:10.1371/journal.pntd.0000313.
- 680 Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, et al. 2013. The global distribution and
681 burden of dengue. *Nature* **496**: 504-507. doi:10.1038/Nature12060.
- 682 Bhunia GS, Kesari S, Chatterjee N, Mandal R, Kumar V, et al. 2012. Seasonal relationship between
683 normalized difference vegetation index and abundance of the *Phlebotomus kala-azar* vector
684 in an endemic focus in Bihar, India. *Geospat Health* **7**: 51-62.
- 685 Bhunia GS, Kumar V, Kumar AJ, Das P, Kesari S. 2010. The use of remote sensing in the identification
686 of the eco-environmental factors associated with the risk of human visceral leishmaniasis
687 (kala-azar) on the Gangetic plain, in north-eastern India. *Ann Trop Med Parasitol* **104**: 35-53.
688 doi:10.1179/136485910x12607012373678.
- 689 Boelaert M, Meheus F, Sanchez A, Singh SP, Vanlerberghe V, et al. 2009. The poorest of the poor: a
690 poverty appraisal of households affected by visceral leishmaniasis in Bihar, India. *Trop Med*
691 *Int Health* **14**: 639-644. doi:10.1111/j.1365-3156.2009.02279.x.
- 692 Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, et al. 2012. Refining the global spatial
693 limits of dengue virus transmission by evidence-based consensus. *PLoS Negl Trop Dis* **6**:
694 e1760. doi:10.1371/Journal.Pntd.0001760.
- 695 Branco S, Alves-Pires C, Maia C, Cortes S, Cristovao JM, et al. 2013. Entomological and ecological
696 studies in a new potential zoonotic leishmaniasis focus in Torres Novas municipality, Central
697 region, Portugal. *Acta Trop* **125**: 339-348. doi:10.1016/j.actatropica.2012.12.008.
- 698 Cardo LJ. 2006. *Leishmania*: risk to the blood supply. *Transfusion* **46**: 1641-1645. doi:10.1111/j.1537-
699 2995.2006.00941.x.
- 700 Chamaille L, Tran A, Meunier A, Bourdoiseau G, Ready P, et al. 2010. Environmental risk mapping of
701 canine leishmaniasis in France. *Parasit Vectors* **3**: 31. doi:10.1186/1756-3305-3-31.
- 702 CIESIN/IFPRI/WB/CIAT. Global Rural Urban Mapping Project (GRUMP) : Gridded Population of the
703 World, version 3. Available: <http://sedac.ciesin.columbia.edu/gpw>. Accessed: December
704 2013
- 705 Cross ER, Newcomb WW, Tucker CJ. 1996. Use of weather data and remote sensing to predict the
706 geographic and seasonal distribution of *Phlebotomus papatasi* in southwest Asia. *Am J Trop*
707 *Med Hyg* **54**: 530-536.

708 De'ath G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* **88**: 243-251.
709 doi:10.1890/0012-9658(2007)88[243:Btfema]2.0.Co;2.

710 de Oliveira EF, Silva EAE, Fernandes CED, Paranhos AC, Gamarra RM, et al. 2012. Biotic factors and
711 occurrence of *Lutzomyia longipalpis* in endemic area of visceral leishmaniasis, Mato Grosso
712 do Sul, Brazil. *Mem Inst Oswaldo Cruz* **107**: 396-401. doi:10.1590/S0074-
713 02762012000300015.

714 Dedet JP, Pratlong F. 2009. Leishmaniasis. In: Cook GC, Zumla AI, editors. Manson's Tropical
715 Diseases, 22nd Edition. London: Saunders Elsevier. pp. 1341-1365.

716 Desjeux P, Alvar J. 2003. Leishmania/HIV co-infections: epidemiology in Europe. *Ann Trop Med*
717 *Parasitol* **97**: 3-15. doi:10.1179/000349803225002499.

718 Edberg SC. 2005. Global infectious diseases and epidemiology network (GIDEON): A world wide web-
719 based program for diagnosis and informatics in infectious diseases. *Clin Infect Dis* **40**: 123-
720 126. doi:10.1086/426549.

721 Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, et al. 2006. Novel methods improve prediction
722 of species' distributions from occurrence data. *Ecography* **29**: 129-151.
723 doi:10.1111/j.2006.0906-7590.04596.x.

724 Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. *J Anim Ecol* **77**:
725 802-813. doi:10.1111/j.1365-2656.2008.01390.x.

726 Elnaïem DA, Schorscher J, Bendall A, Obsomer V, Osman ME, et al. 2003. Risk mapping of visceral
727 leishmaniasis: the role of local variation in rainfall and altitude on the presence and
728 incidence of kala-azar in eastern Sudan. *Am J Trop Med Hyg* **68**: 10-17.

729 Elnaïem DEA. 2011. Ecology and control of the sand fly vectors of *Leishmania donovani* in East Africa,
730 with special emphasis on *Phlebotomus orientalis*. *J Vector Ecol* **36**: S23-S31.
731 doi:10.1111/j.1948-7134.2011.00109.x.

732 FAO. 2008. The Global Administrative Unit Layers (GAUL): Technical Aspects. Rome: Food and
733 Agriculture Organization of the United Nations, EC-FAO Food Security Programme (ESTG).

734 Feliciangeli MD. 2004. Natural breeding places of phlebotomine sandflies. *Med Vet Entomol* **18**: 71-
735 80. doi:10.1111/j.0269-283X.2004.0487.x.

736 Fernandez MS, Lestani EA, Cavia R, Salomon OD. 2012. Phlebotominae fauna in a recent deforested
737 area with american tegumentary leishmaniasis transmission (Puerto Iguazu, Misiones,
738 Argentina): seasonal distribution in domestic and peridomestic environments. *Acta Trop*
739 **122**: 16-23. doi:10.1016/j.actatropica.2011.11.006.

740 Fischer D, Thomas SM, Beierkuhnlein C. 2010. Temperature-derived potential for the establishment
741 of phlebotomine sandflies and visceral leishmaniasis in Germany. *Geospat Health* **5**: 59-69.

742 Fleiss JL, Levin B, Paik MC. 2003. Statistical Methods for Rates and Proportions. Hoboken, New
743 Jersey: John Wiley & Sons. 800 p.

744 Freifeld CC, Mandl KD, Ras BY, Bronwnstein JS. 2008. HealthMap: global infectious disease
745 monitoring through automated classification and visualization of internet media reports. *J*
746 *Am Med Inform Assoc* **15**: 150-157. doi:10.1197/Jamia.M2544.

747 Friedman JH. 2001. Greedy function approximation: A gradient boosting machine. *Ann Stat* **29**: 1189-
748 1232. doi:10.1214/aos/1013203451.

749 Galvez R, Descalzo MA, Guerrero I, Miro G, Molina R. 2011. Mapping the current distribution and
750 predicted spread of the leishmaniosis sand fly vector in the Madrid region (Spain) based on
751 environmental variables and expected climate change. *Vector Borne Zoonotic Dis* **11**: 799-
752 806. doi:10.1089/vbz.2010.0109.

753 Gebre-Michael T, Malone JB, Balkew M, Ali A, Berhe N, et al. 2004. Mapping the potential
754 distribution of *Phlebotomus martini* and *P. orientalis* (Diptera : Psychodidae), vectors of kala-
755 azar in East Africa by use of geographic information systems. *Acta Trop* **90**: 73-86.
756 doi:10.1016/j.actatropica.2003.09.021.

757 Gonzalez C, Rebollar-Tellez EA, Ibanez-Bernal S, Becker-Fauser I, Martinez-Meyer E, et al. 2011.
758 Current knowledge of *Leishmania* vectors in Mexico: how geographic distributions of species

759 relate to transmission areas. *Am J Trop Med Hyg* **85**: 839-846. doi:10.4269/ajtmh.2011.10-
760 0452.

761 Gonzalez C, Wang O, Strutz SE, Gonzalez-Salazar C, Sanchez-Cordero V, et al. 2010. Climate change
762 and risk of leishmaniasis in North America: predictions from ecological niche models of
763 vector and reservoir species. *PLoS Negl Trop Dis* **4**: e585. doi:10.1371/journal.pntd.0000585.

764 Guzman H, Tesh, R. 2000. Effects of temperature and diet on growth and longevity of phlebotomine
765 sand flies (Diptera: Psychodidae). *Biomedica* **20**: 190-199.

766 Harhay MO, Olliaro PL, Costa DL, Costa CHN. 2011. Urban parasitology: visceral leishmaniasis in
767 Brazil. *Trends Parasitol* **27**: 403-409. doi:10.1016/J.Pt.2011.04.001.

768 Hartemink N, Vanwambeke SO, Heesterbeek H, Rogers D, Morley D, et al. 2011. Integrated mapping
769 of establishment risk for merging vector-borne infections: a case study of canine
770 leishmaniasis in southwest France. *PLoS One* **6**: e20817. doi:10.1371/journal.pone.0020817.

771 Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, et al. 2013. Global mapping of infectious disease.
772 *Philos Trans R Soc Lond B Biol Sci* **368**: 20120250. doi:10.1098/Rstb.2012.0250.

773 Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ. 2006. Global environmental data for mapping
774 infectious disease distribution. *Adv Parasitol* **62**: 37-77. doi:10.1016/S0065-308x(05)62002-7.

775 Herrero M, Orfanos G, Argaw D, Mulugeta A, Aparicio P, et al. 2009. Natural history of a visceral
776 leishmaniasis outbreak in highland Ethiopia. *Am J Trop Med Hyg* **81**: 373-377.

777 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate
778 surfaces for global land areas. *Int J Climatol* **25**: 1965-1978. doi:10.1002/Joc.1276.

779 Hlavacova J, Votypka J, Volf P. 2013. The effect of temperature on *Leishmania* (Kinetoplastida:
780 Trypanosomatidae) development in sand flies. *J Med Entomol* **50**: 955-958.
781 doi:10.1603/Me13053.

782 Hutchinson MF. 1995. Interpolating mean rainfall using thin-plate smoothing splines. *Int J Geogr Inf*
783 *Syst* **9**: 385-403. doi:10.1080/02693799508902045.

784 Ives A, Ronet C, Prevel F, Ruzzante G, Fuertes-Marraco S, et al. 2011. *Leishmania* RNA virus controls
785 the severity of mucocutaneous leishmaniasis. *Science* **331**: 775-778.
786 doi:10.1126/science.1199326.

787 Kasap OE, Alten B. 2005. Laboratory estimation of degree-day developmental requirements of
788 *Phlebotomus papatasi* (Diptera : Psychodidae). *J Vector Ecol* **30**: 328-333.

789 Killick-Kendrick R. 1999. The biology and control of phlebotomine sand flies. *Clin Dermatol* **17**: 279-
790 289. doi:10.1016/S0738-081x(99)00046-2.

791 Lima BS, Dantas-Torres F, de Carvalho MR, Marinho JF, de Almeida EL, et al. 2013. Small mammals as
792 hosts of *Leishmania* spp. in a highly endemic area for zoonotic leishmaniasis in north-eastern
793 Brazil. *Trans R Soc Trop Med Hyg* **107**: 592-597. doi:10.1093/trstmh/trt062.

794 Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, et al. 2012. Global and regional mortality from
795 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global
796 Burden of Disease Study 2010. *Lancet* **380**: 2095-2128. doi:10.1016/S0140-6736(12)61728-0.

797 Malafaia G. 2009. Protein-energy malnutrition as a risk factor for visceral leishmaniasis: a review.
798 *Parasite Immunol* **31**: 587-596. doi:10.1111/j.1365-3024.2009.01117.x.

799 Malaviya P, Picado A, Singh P, Hasker E, Singh RP, et al. 2011. Visceral leishmaniasis in Muzaffarpur
800 district, Bihar, India from 1990 to 2008. *PLoS One* **6**: e14751.
801 doi:10.1371/journal.pone.0014751.

802 Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, et al. 2012. Disability-adjusted life years
803 (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the
804 Global Burden of Disease Study 2010. *Lancet* **380**: 2197-2223. doi:10.1016/S0140-
805 6736(12)61689-4.

806 Murray HW, Berman JD, Davies CR, Saravia NG. 2005. Advances in leishmaniasis. *Lancet* **366**: 1561-
807 1577. doi:10.1016/S0140-6736(05)67629-5.

808 Nordhaus W. 2008. New metrics for environmental economics: gridded economic data. *Integr Assess*
809 **8**: 73-84.

810 Nordhaus WD. 2006. Geography and macroeconomics: new data and new findings. *Proc Natl Acad*
811 *Sci U S A* **103**: 3510-3517. doi:10.1073/pnas.0509842103.

812 Novais FO, Carvalho LP, Graff JW, Beiting DP, Ruthel G, et al. 2013. Cytotoxic T cells mediate
813 pathology and metastasis in cutaneous leishmaniasis. *PLoS Pathog* **9**: e1003504.
814 doi:10.1371/journal.ppat.1003504.

815 Peterson AT. 2008. Biogeography of diseases: a framework for analysis. *Naturwissenschaften* **95**:
816 483-491. doi:10.1007/s00114-008-0352-5.

817 Peterson AT, Soberon J, Pearson RG, Anderson RP, Martinez-Meyer E, et al. 2011. Ecological niches
818 and geographic distributions. Princeton: Princeton University Press. 314 p.

819 Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, et al. 2009. Sample selection bias and presence-
820 only distribution models: implications for background and pseudo-absence data. *Ecol Appl*
821 **19**: 181-197. doi:10.1890/07-2153.1.

822 Phillips SJ, Elith J. 2011. Logistic methods for resource selection functions and presence-only species
823 distribution models. In: AAAI, editor. Proceedings of the Twenty-Fifth AAAI Conference on
824 Artificial Intelligence. San Francisco: AAAI (Association for the Advancement of Artificial
825 Intelligence). pp. 1384-1389.

826 Poche D, Garlapati R, Ingenloff K, Remmers J, Poche R. 2011. Bionomics of phlebotomine sand flies
827 from three villages in Bihar, India. *J Vector Ecol* **36**: S106-S117. doi:10.1111/j.1948-
828 7134.2011.00119.x.

829 Ready PD. 2013. Biology of phlebotomine sand flies as vectors of disease agents. *Annu Rev Entomol*
830 **58**: 227-250. doi:10.1146/annurev-ento-120811-153557.

831 Reithinger R. 2008. Leishmaniasis' burden of disease: ways forward for getting from speculation to
832 reality. *PLoS Negl Trop Dis* **2**: e285. doi:10.1371/journal.pntd.0000285.

833 Reithinger R, Dujardin JC. 2007. Molecular diagnosis of leishmaniasis: current status and future
834 applications. *J Clin Microbiol* **45**: 21-25. doi:10.1128/Jcm.02029-06.

835 Reithinger R, Dujardin JC, Louzir H, Pirmez C, Alexander B, et al. 2007. Cutaneous leishmaniasis.
836 *Lancet Infect Dis* **7**: 581-596. doi:10.1016/S1473-3099(07)70209-8.

837 Rogers DJ, Hay SI, Packer MJ. 1996. Predicting the distribution of tsetse flies in west Africa using
838 temporal Fourier processed meteorological satellite data. *Ann Trop Med Parasitol* **90**: 225-
839 241.

840 Scharlemann JPW, Benz D, Hay SI, Purse BV, Tatem AJ, et al. 2008. Global data for ecology and
841 epidemiology: a novel algorithm for temporal Fourier processing MODIS data. *PLoS One* **3**:
842 e1408. doi:10.1371/Journal.Pone.0001408.

843 SINAN. Sistema de Informação de Agravos de Notificação. Available:
844 <http://dtr2004.saude.gov.br/sinanweb/index.php>. Accessed: March 2013

845 Singh R, Lal S, Saxena VK. 2008. Breeding ecology of visceral leishmaniasis vector sandfly in Bihar
846 state of India. *Acta Trop* **107**: 117-120. doi:10.1016/j.actatropica.2008.04.025.

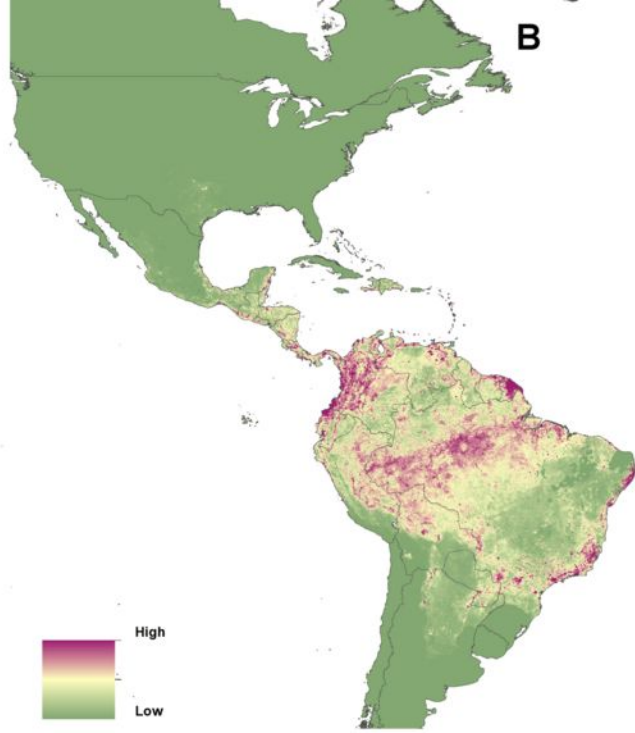
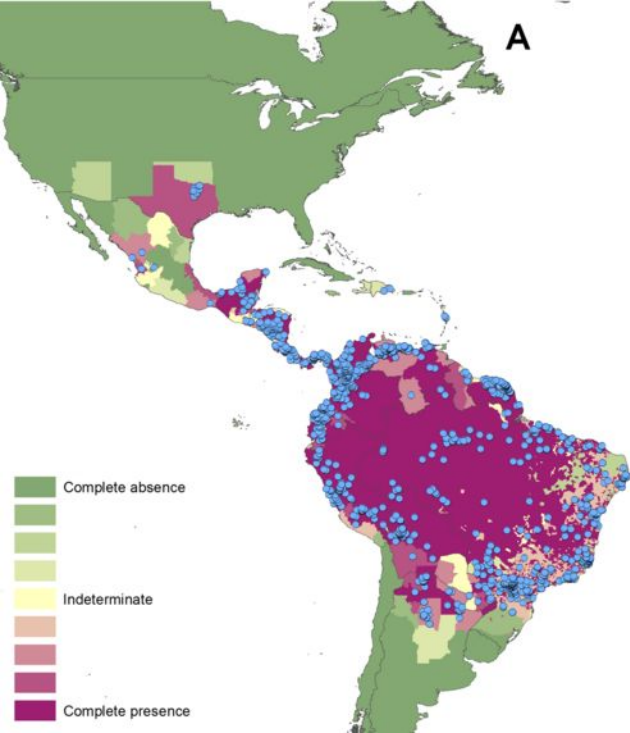
847 Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, et al. 2011. The dominant *Anopheles*
848 vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and
849 bionomic precis. *Parasit Vectors* **4**: 89. doi:10.1186/1756-3305-4-89.

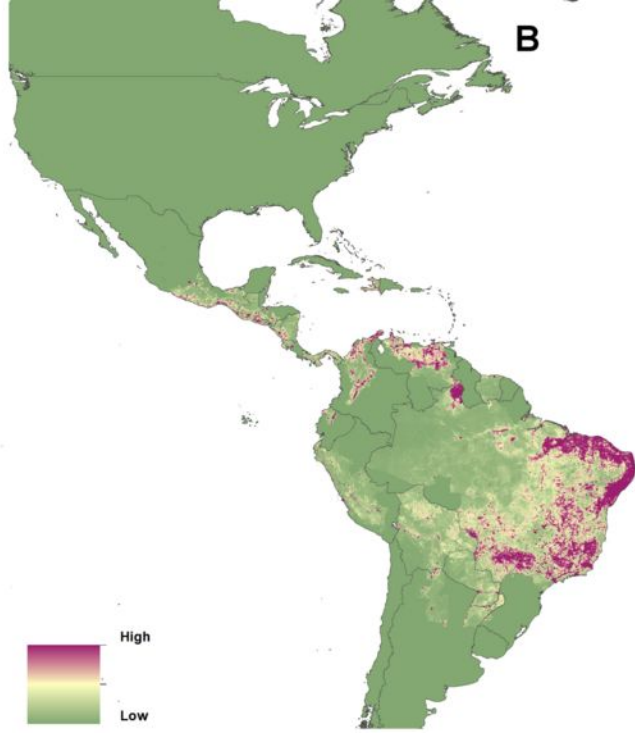
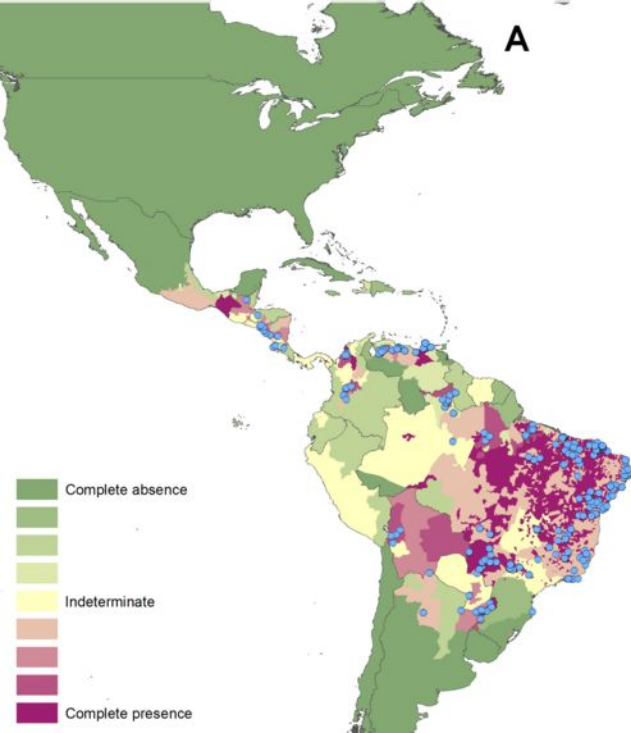
850 Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, et al. 2010. The dominant *Anopheles*
851 vectors of human malaria in Africa, Europe and the Middle East: occurrence data,
852 distribution maps and bionomic precis. *Parasit Vectors* **3**: 117. doi:10.1186/1756-3305-3-
853 117.

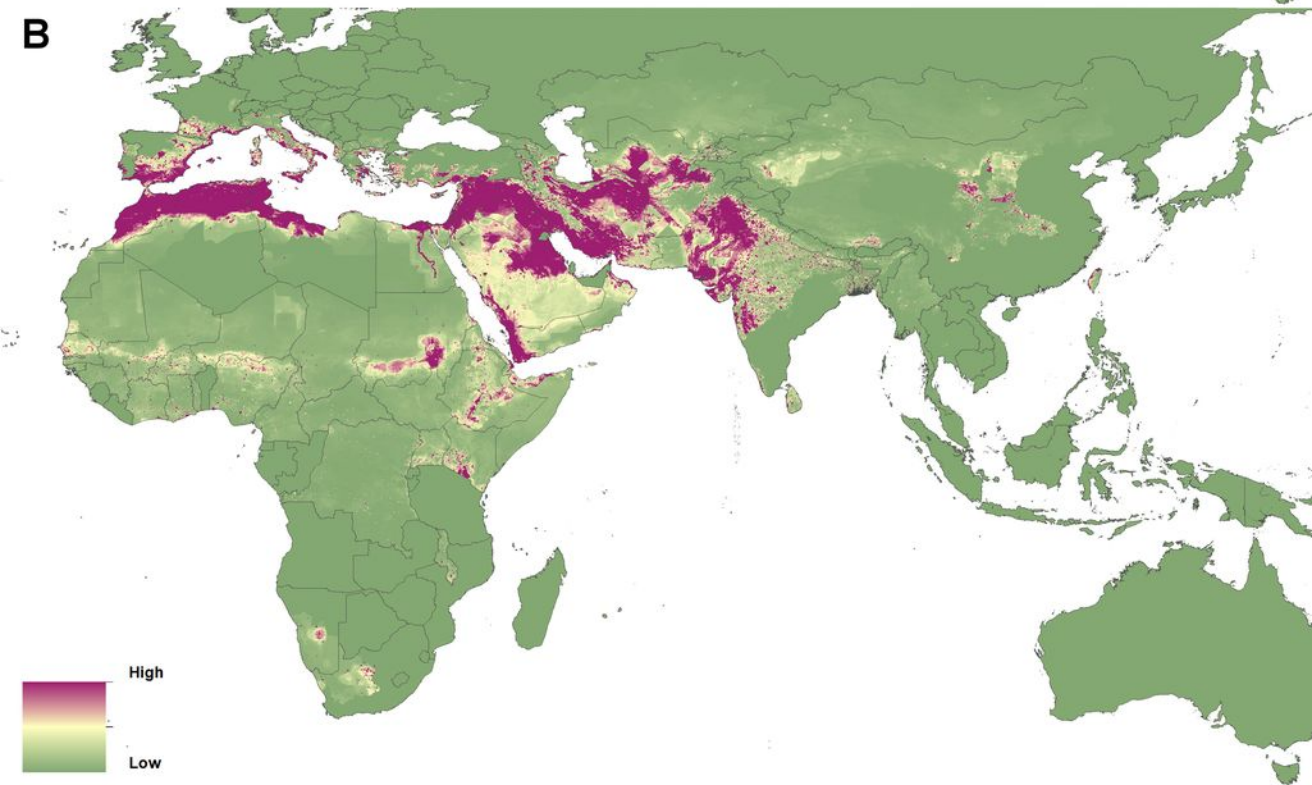
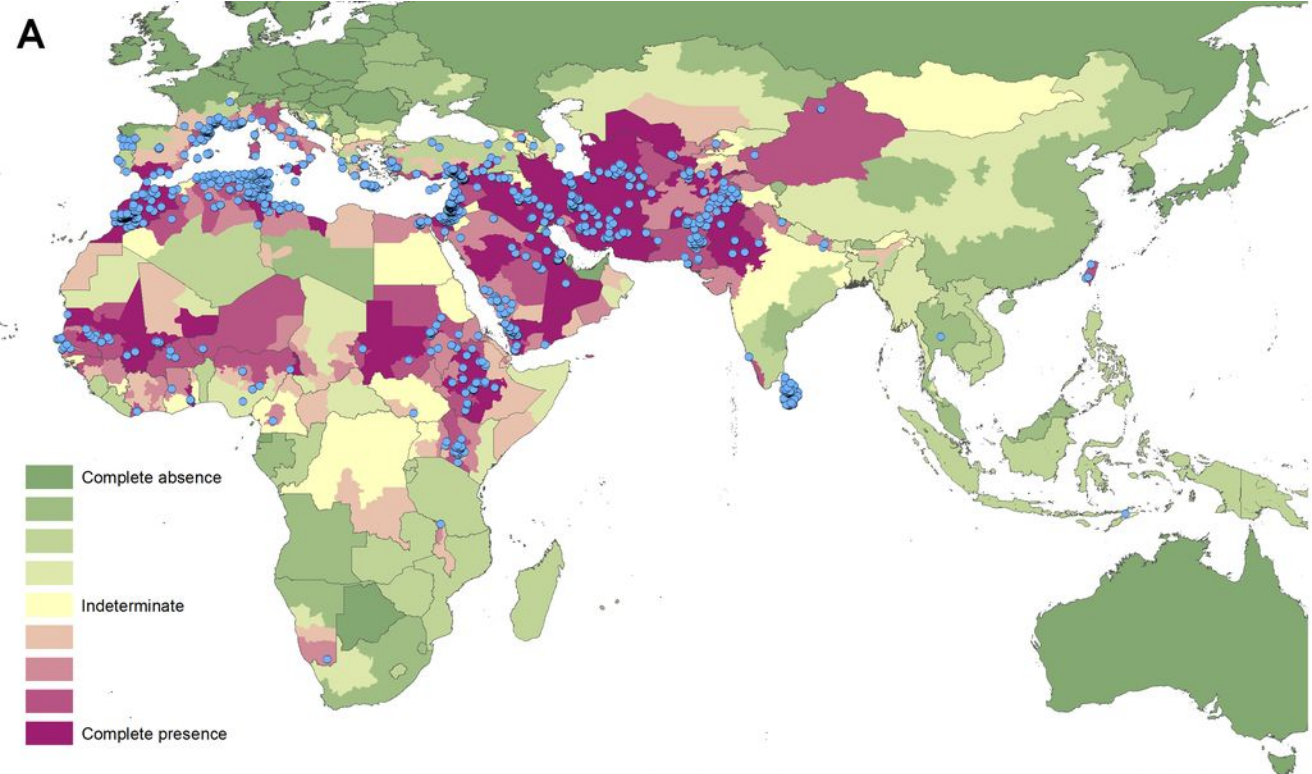
854 Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, et al. 2010. The dominant *Anopheles*
855 vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic
856 precis. *Parasit Vectors* **3**: 72. doi:10.1186/1756-3305-3-72.

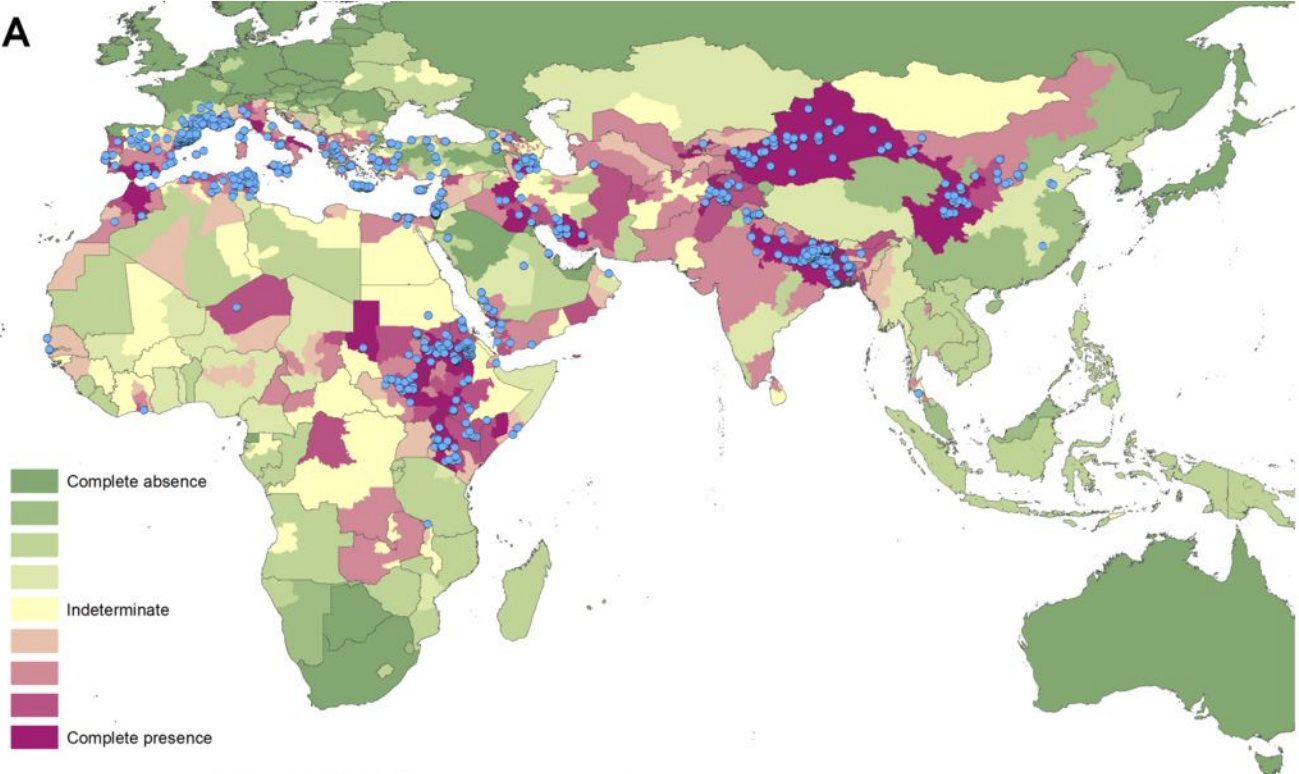
857 Syfert MM, Smith MJ, Coomes DA. 2013. The effects of sampling bias and model complexity on the
858 predictive performance of MaxEnt species distribution models. *PLoS One* **8**: e55158.
859 doi:10.1371/journal.pone.0055158.

- Thomson MC, Elnaiem DA, Ashford RW, Connor SJ. 1999. Towards a kala azar risk map for Sudan: mapping the potential distribution of *Phlebotomus orientalis* using digital data of environmental variables. *Trop Med Int Health* **4**: 105-113. doi:10.1046/j.1365-3156.1999.00368.x.
- Toumi A, Chlif S, Bettaieb J, Ben Alaya N, Boukthir A, et al. 2012. Temporal dynamics and impact of climate factors on the incidence of zoonotic cutaneous leishmaniasis in central Tunisia. *PLoS Negl Trop Dis* **6**: e1633. doi:10.1371/Journal.Pntd.0001633.
- Uranw S, Hasker E, Roy L, Meheus F, Das ML, et al. 2013. An outbreak investigation of visceral leishmaniasis among residents of Dharan town, eastern Nepal, evidence for urban transmission of *Leishmania donovani*. *BMC Infect Dis* **13**: 21. doi:10.1186/1471-2334-13-21.
- Van Der Wal J, Shoo LP, Graham C, William SE. 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecol Modell* **220**: 589-594. doi:10.1016/j.ecolmodel.2008.11.010.
- WHO. 2009. Neglected tropical diseases, hidden successes, emerging opportunities. Geneva: World Health Organization. 59 p.
- WHO. 2010. Control of the Leishmaniasis. Report of a Meeting of the WHO Expert Committee on the Control of Leishmaniasis, Geneva, 22-26 March 2010. Geneva: World Health Organization. 186 p.
- WHO. 2011. World Health Statistics 2011. Geneva: World Health Organization. 170 p.
- Zeilhofer P, Kummer OP, dos Santos ES, Ribeiro ALM, Missawa NA. 2008. Spatial modelling of *Lutzomyia (Nyssomyia) whitmani sensu lato* (Antunes & Coutinho, 1939) (Diptera: Psychodidae: Phlebotominae) habitat suitability in the state of Mato Grosso, Brazil. *Mem Inst Oswaldo Cruz* **103**: 653-660. doi:10.1590/S0074-02762008000700005.







A**B**